

Р. С. Гиляревский

ОСНОВЫ ИНФОРМАТИКИ

Курс лекций

Лекции рассчитаны на студентов высших учебных заведений. Цель курса лекций – дать студентам представление о современной информатике как научной дисциплине и теоретической основе информационной технологии, познакомить их с основными понятиями и закономерностями интеллектуальной коммуникации, информационной деятельности, научить методам информационного поиска.

В лекциях раскрывается сущность информации и информационной технологии. Студенты знакомятся с возможностями электронной обработки текста от проверки орфографии и грамматики до машинного перевода с одного языка на другой, с электронными видами коммуникации. Излагается современное понимание информационной инфраструктуры, даются сведения о сетевой технологии и начальных шагах работы в Интернете.

Курс "Основы информатики" читался на общеуниверситетской кафедре научной информации Московского государственного университета им. М. В. Ломоносова в 1964–1985 гг. для студентов биологического, физического, химического, юридического факультетов, Института стран Азии и слушателей факультета повышения квалификации. С 1986 г. он читается студентам факультета журналистики в переработанном виде с учетом новой информационной технологии и особой специфики гуманитарного труда.

Содержание

ВВОДНАЯ ЛЕКЦИЯ	5
Информатика как научная дисциплина	5
Становление информатики	7
Предмет и объекты исследования	16
Информатика и другие науки и научные дисциплины	17
ИНФОРМАЦИЯ – ЗНАНИЕ – НАУКА	21
Информация и данные	22
Свойства информации	24
Структура информации	25
Особенности информации	26
Информация научная и техническая	28
Наука как социальное явление	32
Перспективы развития науки	42
ИНТЕЛЛЕКТУАЛЬНАЯ КОММУНИКАЦИЯ	46
Основные понятия	46
Система научной коммуникации	47
Библиотечная и информационная деятельность	53
Научно-информационная деятельность	57
Этапы и задачи коммуникации	58
Информационное обслуживание	61
Перспективы развития	63
ЧЕЛОВЕК В ПРОЦЕССЕ КОММУНИКАЦИИ	64
Потребители информации	64
Эгалитарность информационного обслуживания	68
Информационные потребности в развитии	73
ЛИТЕРАТУРА КАК ИСТОЧНИК ИНФОРМАЦИИ	77
Основные понятия, эволюция и типология	77
Закономерности роста и старения	83
Закон рассеяния	88
Перспективы развития	92
ИНФОРМАЦИОННЫЕ ИЗДАНИЯ И УСЛУГИ	95
Основные виды	95
Реферирование и библиографирование	96
Реферативный журнал ВИНТИ	103
Зарубежные реферативные журналы	105
Электронная информация и базы данных	112
Сети передачи и средства хранения и обработки данных	114
Информационные услуги	116
Информационные структуры и инфраструктура	119
ИНФОРМАЦИОННЫЙ ПОИСК	126
Предыстория и сущность	126
Процедуры и понятия	129
Координатное индексирование	133
Цитирование, библиографическое сочетание, социитирование	136

Иерархические и фасетные классификации	140
Рубрикаторы информационных изданий	147
Разные типы информационно-поисковых языков	149
Базы и банки данных	152
ИНФОРМАЦИОННЫЕ СИСТЕМЫ	156
Информационно-поисковые системы	156
Интеллектуальные информационные системы	157
Гипертекстовые системы	159
Системы машинного перевода	169
ИНФОРМАЦИОННАЯ ТЕХНОЛОГИЯ	174
О понятии информационной технологии	174
Тенденции развития информационных технологий	177
Влияние информационных технологий на развитие науки	180
Социальные последствия новой технологии	183
КОМПЬЮТЕРНЫЕ СРЕДСТВА КОММУНИКАЦИИ	187
Электронные вычислительные машины	187
Новые поколения компьютеров	190
Персональный компьютер и персональные вычисления	194
РАБОТА С ТЕКСТОМ НА КОМПЬЮТЕРЕ	201
О языках программирования	201
Прикладные программы подготовки текстов	206
"Понимание" текста на естественном языке	215
ЭЛЕКТРОННАЯ КНИГА	223
Новая концепция книги	223
Сущность, особенности и разновидности электронной книги	225
Электронный журнал: проблемы распространения и хранения	226
Организационные и юридические проблемы	231
Электронная библиотека реальная и виртуальная	233
Лучше ли электронная книга традиционной печатной?	237
СВЕДЕНИЯ ОБ ИНТЕРНЕТ	245
Интернет как глобальная компьютерная сеть	245
Организация доступа к первоисточникам	251
ЗАКЛЮЧИТЕЛЬНАЯ ЛЕКЦИЯ	263
Идеи и методы информатики	263
Поиски фундаментального закона	267
Определение и предметная область информатики	272
Перспективы информатики	276
СЛОВАРЬ ТЕРМИНОВ	280

Вводная лекция

Информатика как научная дисциплина

В современном обществе все большее место занимают различные виды деятельности по передаче и распространению результатов умственной деятельности. Журналисты, редакторы, референты, документалисты, библиотекари и библиографы, информационные и архивные работники только по традиции считают, что их профессии относятся к разным видам деятельности. Теперь уже многим понятно, что эти виды являются этапами одного и того же процесса интеллектуальной коммуникации. Коммуникация – связь, общение, сообщение (процесс и путь сообщения) – может происходить непосредственно, на физическом уровне, но интеллектуальная коммуникация, т. е. относящаяся к мыслительным способностям человека, всегда идеальна и осуществляется информационным путем. Поэтому ее часто называют также информационной коммуникацией.

В последние десятилетия в информационной технологии, т. е. методах и технических средствах передачи информации, произошла компьютерная, а вернее электронная революция. Появление персональных компьютеров, их объединение в сети между собой и с крупными вычислительными центрами, внедрение в привычные процессы информационного обмена настольных издательских комплексов, копировальных и множительных аппаратов, относительно дешевых, но очень емких средств хранения информации (компактных оптических дисков) приводит к существенным переменам в сфере коммуникации. Многие тысячелетиями существовавшие профессии сливаются воедино, появляются новые, для которых у нас еще нет хороших русских названий, например, *специалист по паблик рилейшнз (общественным связям)* или *администратор баз данных*. Разумеется, пройдет много времени, пока специалисты всех этих профессий полностью осознают свое единство. Но новые поколения специалистов должны вырабаты-

вать этические принципы и профессиональное мировоззрение сообщая в рамках общего информационного направления.

Новая информационная технология – компьютерная подготовка текстов, информационно-поисковые, гипертекстовые и интеллектуальные системы, передача машиночитаемых файлов по вычислительным сетям и телефонным каналам, настольные издательские системы, копировальная и множительная техника – позволяет по-новому трактовать всю эту сферу деятельности от журналистики до библиотечного и архивного дела. Специалисты этих профессий в первую очередь должны понять, что в основе всех информационных процессов лежат общие закономерности, коренящиеся в структуре и свойствах самой информации. Но владеть навыками самостоятельного использования информационных технологий и технических средств необходимо всем работникам умственного труда, выпускникам высших учебных заведений.

Студенты, пришедшие в вуз из средней школы, убеждены, что информатика сводится к процессу овладения этими навыками, а попросту, к умению работать на компьютере, поскольку именно этому учат в школе по предмету, который называют "информатикой". Но это односторонняя и потому неправильная трактовка информатики, которая по самому своему названию является дисциплиной, изучающей информацию. Компьютер и связанные с ним электронные аппараты действительно служат наиболее мощными в наше время средствами переработки, хранения и передачи информации, а точнее – данных, в которых эта информация может содержаться. Но эти процессы происходили в обществе до появления компьютера и будут происходить, когда для их осуществления будут созданы иные средства.

Информатика – научная дисциплина, изучающая структуру и общие свойства семантической информации, закономерности процессов ее функционирования в обществе, является теоретической базой информационной технологии, которую часто отождествляют с информатикой.

В дальнейшем смысл и отдельные термины этого определения будут подробно разъяснены, и вы поймете, почему мы говорим о *семантической* информации и *информационной* технологии, а также узнаете, что *функционирование* информации в обществе осуществляется в процессах *интеллектуальной* (или *информационной*) коммуникации.

Становление информатики

Информатика возникла не сразу. Она развивалась с начала нашего столетия, и ее общие контуры были четко очерчены ее создателем Полем Отле (1868–1944). В числе ученых и специалистов, внесших существенный вклад в развитие информатики, следует упомянуть такие имена, как С. Бредфорд, В. Буш, К. Муэрс, М. Таубе, А. Кент, Б. Виккери, Х.-П. Лун, С. Клевердон, О. Вейнберг, Д. Прайс, Ю. Гарфилд, А. И. Михайлов, В. А. Успенский, Г. Э. Вледуц, Ю. А. Шрейдер, А. И. Черный, В. К. Финн. Работы и достижения этих авторов (которые приведены в табл. 1) могут служить вехами в развитии информатики.

Представляют интерес высказывания некоторых американских ученых, которые с разных позиций оценивали развитие информатики. Проф. Т. Сарацевич подчеркивал значение для ее становления исследований в области научной коммуникации: "Несмотря на то, что информационная наука никак не ограничена одной лишь научной коммуникацией (большинство теоретических и практических работ в этой области имеют универсальный характер), проблемы научной коммуникации в том виде, в каком они понимались в 30-х и 40-х годах, послужили основным импульсом для появления этой науки, причиной ее особенного, философского характера и структуры принимаемых в ней решений. Таким образом, понимание природы проблем научной коммуникации и их толкования помогает понять суть информационной науки и ее взаимоотношения с другими науками"¹.

На протяжении 50-х и 60-х годов информатика исходила из общественной потребности упорядочить обмен информацией, главным образом, внутри самой науки. Внешними факторами, обуславливавшими такую потребность, являлись: быстрый рост научной литературы, трудности ее тематического отбора в связи с ее рассеянием и ограниченностью поисковых средств, процессы специализации и интеграции в науке, размывающие традиционные границы между научными дисциплинами. В течение этого периода наши представления о способах и методах решения проблемы информационного поиска претерпевали существенные изменения. Первоначально надежды возлагались на одни только технические средства.

¹ Saracevic T. An essay of the past and future (?) of information science education // Information Processing and Management. – 1979. – Vol. 15, N 1. – P. 1–15.

Т а б л и ц а 1

Основные достижения в развитии информатики

Авторы	Публикации	Достижения
П. Отле (1868–1944)	<i>Универсальная библиографическая классификация. Трактат о документации</i> .-Брюссель, 1905, 1934	Создание классификации для библиографического репертуара. Первый труд по информатике и прогноз ее развития.
С. Бредфорд (1878–1948)	<i>Источники информации по определенному предмету</i> .- Лондон, 1934	Открытие закона рассеяния научных публикаций.
В. Буш (1890–1974)	<i>Возможный способ нашего мышления</i> . - Нью-Йорк, 1945	Прогноз развития информатики на полстолетия.
К. Муэрс (1919–1994)	<i>Кодирование для механической организации знаний</i> . - Кембридж (Массачусетс) 1951	Формулирование основных понятий информационного поиска.
М. Таубе (1910–1965)	<i>Исследования по координатному индексированию. Вычислительные машины и здравый смысл</i> . - Вашингтон, 1953–1959	Создание теории информационно-поисковых систем дескрипторного типа.
А. Кент	<i>Машинный поиск литературы. Энциклопедия библиотековедения и информатики</i> . – Нью-Йорк, 1956,1970–1990	Применение технических средств для информационного поиска. Создание многотомной энциклопедии по информатике.
Б. Виккери	<i>Классификация и индексирование в науке</i> .- Лондон, 1957	Развитие информационно-поисковых языков.
С. Клевердон (1914–1997)	<i>Крэнфилдский проект</i> .- Крэнфилд, 1957	Изучение эффективности информационно-поисковых систем.
Х.-П. Лун (1896–1964)	<i>Избирательное распространение информации</i> . - Нью-Йорк, 1958	Применение компьютеров в информационной деятельности.
О. Вейнберг	<i>Наука, правительство и информация</i> .- Вашингтон, 1963	Доклад комиссии ученых президенту США о значении информации для науки.
Д. Прайс (1922–1983)	<i>Малая наука, большая наука</i> .- Нью Хейвен, 1963	Анализ неформальных коммуникаций ученых и распределения их публикаций.
Ю. Гарфилд	<i>Указатель цитированной литературы, Атласы науки, Текущее содержание журналов</i> . - Филадельфия, 1964	Новый принцип поиска и анализа литературы и информации.
Дж. Солтон (1927–1995)	<i>Автоматическая обработка, хранение и поиск информации. Динамические библиотечно-информационные системы</i> . –	Современная теория информационных систем. Создание «Совершенной системы поиска текстов Солтона» (SMART)

	<i>Нью Джерси, 1975-79</i>	
Б. Брукс (1910–1991)	<i>Основания информационной науки. – Лондон, 1960</i>	Создание математического аппарата к закону Бредфорда и попытка открыть общие закономерности информатики
А.И Михайлов (1905–1987)	<i>Основы научной информации. Основы информатики. Научные коммуникации и информатика. - Москва, 1965–1976</i>	Создание обобщающих монографий по информатике и научных основ информационной деятельности в СССР.
В.А.Успенский	<i>К проблеме построения машинного языка. К проблематике теории научной информации. - Москва, 1959-1963</i>	Определение принципов формализации интеллектуальных процессов.
Г.Э. Влэдудц (1928– 1993)	<i>Автоматизированные информационные системы для химии. - Москва, 1974</i>	Формулирование основ химической информатики
Ю.А.Шрейдер (1927–1998)	<i>Равенство, сходство, порядок. Системность и эволюция. Москва, 1980-1983</i>	Создание теории бинарных отношений (сходства), теории ранговых распределений и общей концепции системы.
А.И. Черный	<i>Введение в теорию информационного поиска. - Москва, 1975. Соавторство в монографиях А.И.Михайлова</i>	Обобщение теории информационного поиска. Создание крупнейшей информационной системы АССИСТЕНТ (ВИНИТИ)
В.К. Финн	<i>Некоторые логические проблемы информационного поиска. Интеллектуальные системы и общество. - Москва, 1975, 2001</i>	Создание теории интеллектуальных информационных систем типа “Джон Стюарт Миль”.
Д.Г. Лахути	<i>Вопросы теории поисковых систем. Автоматизированные документально-фактографические информационно-поисковые системы. – Москва, 1963, 1988</i>	Разработка теории информационного поиска. Создание информационных систем Пусто-Непусто, Скобки, Реферат.

Довольно скоро пришло понимание необходимости разработать семантические средства анализа и синтеза научной информации. Не случайно теория информационного поиска, методы координатного индексирования, понятия релевантности (соответствия запросу) и pertinентности (соответствия потребности в информации) стали стержневыми для информатики.

В 70-е годы произошло дальнейшее расширение сферы приложения информатики. Научная информация стала играть существенную роль в вопросах управления народным хозяйством, принятия политических решений, глобального моделирования развития общества. Системы научно-технической информации стали рассматриваться как элемент автоматизированных систем управления, что было правомерно лишь до известной степени. При этом выяснилось, что как бы широко ни трактовалась научная информация, ее одной недостаточно для информационных систем, обеспечивающих решение столь общих социальных проблем. С другой стороны, оказалось, что в наиболее существенных аспектах методы построения информационно-поисковых систем научной информации применимы и для многих других видов информации (экономической, статистической, политической, производственной и т. п.). Особую значимость приобрела, наряду с научной, информация фактографическая, а информационное обслуживание, сохранив дисциплинарную ориентацию, стало развиваться и в проблемных аспектах.

С нынешних позиций хорошо видно и то влияние, которое на протяжении всего этого периода оказывал на развитие информатики быстрый прогресс технических средств обработки информации, прежде всего, вычислительных машин. История техники еще не знала таких стремительных темпов совершенствования новой технологии, ее внедрения во все сферы жизни общества. Это влияние мы учитываем по реальному проникновению компьютеров в информационную деятельность. Но не меньшим является и влияние компьютерной науки, ее системы понятий, терминологии, разработанных ею принципов обработки и хранения информации.

Крупный американский экономист в области информатики Ф. Махлуп (1903–1983) перечислил следующие объекты исследования информатики: "процессы коммуникации в науке, методы классификации, рост и рассеяние литературы, моделирование информационных систем и сетей, методы обмена информацией, управление доступом к

ней, поведение абонентов информационных систем и служб". Вслед за тем он добавил: "Трудно сказать, составляют ли исследования этого типа самостоятельную дисциплину или являются развитием компьютерной науки или библиотековедения"¹. Однако то, в чем он сомневался два десятилетия назад, теперь представляется очевидным. Информатика (information science – в англо-американской терминологии) является самостоятельной научной дисциплиной, имеющая собственный предмет и специфические методы исследования.

Информатика зародилась в недрах библиотековедения, из которого она начала выделяться сравнительно недавно – в конце XIX в. Это было обусловлено огромными достижениями в науке и технике, которые сопровождались быстрым ростом количества выходящей в мире научно-технической литературы и расширением круга ее потенциальных читателей. Первым шагом в этом направлении было создание в 1895 г. Международного института библиографии, ставшего колыбелью Международной федерации по информации и документации – главной профессиональной организации специалистов по информатике, просуществовавшей до 2001 г.

До появления специальных информационных органов потребности ученых в научной информации удовлетворялись библиотеками, ибо в течение многих веков их деятельность была ориентирована на решение этой задачи. Поэтому библиотековедение и было той отраслью знания, в которой накапливался, анализировался и обобщался опыт информационного обслуживания ученых, вырабатывались наиболее эффективные формы и методы работы библиотек в этом направлении.

До изобретения книгопечатания подвижными литерами в середине XV в. широкие слои населения были неграмотны и поэтому не нуждались в библиотеках. Лишь к концу XIX в. распространение грамотности среди населения развитых стран возросло до такой степени, что массовых читателей в библиотеках стало значительно больше, чем ученых и специалистов. Поэтому потребовалась специализация фондов и способов обслуживания в библиотеках этой новой категории читателей, что привело к разделению библиотек на *специальные* (научные и

¹ Machlup F., Mansfield U. Cultural diversity in studies of information. // The study of information: Interdisciplinary messages. – New York: Wiley, 1983. – P. 22.

научно-технические) и *массовые* и положило начало возникновению информатики.

Для массовых библиотек – а таковыми является подавляющее их большинство (80 % в России) – главными стали культурно-просветительные задачи, требующие от библиотекарей выполнения определенных педагогических функций, постоянной работы с читателями, руководства их чтением. Что же касается научных и научно-технических библиотек, то они стали специализироваться на информационном обслуживании ученых и специалистов, все более превращаясь в органы научной и технической информации. В настоящее время многие специальные библиотеки называются библиотеками лишь по традиции или из-за ведомственной принадлежности, а фактически по выполняемым функциям давно являются органами научной и технической информации. И хотя в подготовке кадров, их профессиональном мировоззрении, этических принципах библиотечного и информационного дела сохраняются существенные различия, нередки случаи официального преобразования библиотек в информационные центры. Российским примером такого рода может служить преобразование Фундаментальной библиотеки общественных наук Академии наук СССР в Институт научной информации по общественным наукам.

Итак, между библиотековедением и информатикой существуют тесные родственные связи, которые сохраняются до сих пор. Но это не означает, что информатика – лишь часть библиотековедения, как считают некоторые наиболее радикальные представители последнего. Сейчас библиотековедение и информатика – уже отдельные научные дисциплины с разными предметами исследований. Они изучают разные объекты или разные стороны одних и тех же объектов и решают разные научно-познавательные и практические задачи.

Для информатики, например, очень важно изучение неформальных каналов научной коммуникации, а также мотивов написания и структуры первичных научных документов, что никогда не интересовало библиотековедение. А за пределами интересов информатики остаются проблемы непрерывного образования, библиографического учета всех выходящих в мире произведений печати или долговременного хранения оригиналов книг, все более волнующие специалистов библиотеко- и библиографоведения.

Возрастание роли научно-информационной деятельности и информатики в развитии науки и техники стало ясно осознаваться с нача-

лом научно-технической революции, которое может быть отнесено на конец 30-х – начало 40-х годов нашего столетия. Именно в это время появляются специальные журналы по информатике, которые выходят до настоящего времени. В 1948 г. Королевское общество Великобритании провело первую Международную конференцию по научной информации, в 1949 г. состоялась Международная конференция по научному реферированию, организованная ЮНЕСКО. В 1952 г. в Академии наук СССР был создан Институт научной информации, преобразованный в 1955 г. во Всесоюзный (ныне Всероссийский) институт научной и технической информации (ВИНИТИ). Этим было положено начало создания в нашей стране единой государственной системы научной и технической информации. С 1961 г. начал выходить периодический сборник *Научно-техническая информация*, посвященный проблемам информатики и научно-информационной деятельности.

До начала 60-х годов специалисты по информатике еще не осознавали огромной роли неформальных каналов научной коммуникации – личных контактов между учеными и специалистами, научных совещаний, так называемых «незримых коллективов» (*invisible colleges*). Информатика не включала их в круг объектов своего изучения и ограничивалась рассмотрением лишь формальных каналов, основанных на использовании различных научных документов и изданий. Поэтому информатика в ряде западноевропейских языков называлась *документацией* или *документалистикой*. По традиции эти термины до сих пор сохраняются в названиях некоторых специальных журналов, выходящих в ряде зарубежных стран. Но в названиях многих профессиональных ассоциаций и специальных журналов термин «документация» был заменен термином «информатика» или «информационная наука» (*information science*). Сейчас «документацию» или «документалистику» можно рассматривать лишь как дисциплину, изучающую документальные формы коммуникации.

С середины 60-х годов начался новый этап в развитии информатики. Все большее осознание ограниченности прежнего, документалистского подхода нашло свое выражение в поисках нового названия для данной научной дисциплины, которое правильнее раскрывало бы сущность ее главной проблематики и было бы удобным для использования, прежде всего – однозначным и однословным. В качестве такого названия был предложен термин «информатика» (*англ.* *information science, informatics; франц.* *science de l'information, informatique; нем.*

Informationswissenschaft, Informatik), который быстро получил довольно широкое распространение и признание. Об этом свидетельствует, в частности, переименование многих специальных научных журналов по информатике, а также появление новых журналов, содержащих этот термин в своих названиях.

В настоящее время информатику можно считать уже вполне признанной и сложившейся в самостоятельную дисциплину. Важными показателями этого служат следующие факты:

- имеется довольно обширная специальная литература, посвященная проблемам информатики: в мире регулярно выходит несколько сотен журналов по информатике, ежегодно издаются тысячи новых книг;

- действует сеть специальных учебных заведений, в которых ведется подготовка специалистов по информатике – как исследователей, так и практиков;

- во многих странах созданы профессиональные объединения информатиков, входящих также в различные международные ассоциации;

- по проблемам информатики ежегодно защищаются тысячи кандидатских и докторских диссертаций.

Всего в мире ежегодно появляются десятки тысяч публикаций по различным вопросам информатики. Для сравнения следует указать, что по таким давно сложившимся наукам, как астрономия и биофизика, и по таким сравнительно новым дисциплинам, как вычислительная техника и техническая кибернетика, в год выходит приблизительно по 10 тыс. публикаций.

Параллельно с выделением и становлением информатики как новой научной дисциплины происходили и другие важные события, которые оказали большое влияние на ее развитие. Расширяющееся внедрение компьютеров во все сферы интеллектуальной деятельности требовало решения разнообразных проблем, связанных с использованием этих совершенно новых и непривычных технических средств. Это привело к возникновению специальной научной дисциплины, которая в США получила название «computer science» (наука о вычислительных машинах), во Франции – «informatique» (1962 г.), в ФРГ – «Informatik» (1968 г.), т. е. название, совпадающее с русским термином «информатика». В дальнейшем этот термин укоренился в романских языках – испанском, итальянском, португальском, румынском.

В сегодняшней трактовке французских и германских специалистов «информатика» охватывает следующие основные вопросы: теория формальных языков; теория автоматов; теория алгоритмов; математическое программирование (языки программирования, программные системы для ЭВМ, информационные системы и системы обработки данных, автоматизированное проектирование); технические средства (элементы и схемы – арифметические, логические, кодирующие и декодирующие, преобразующие и т. п., системы хранения данных; устройства ввода, вывода и обработки данных, системное проектирование и архитектура ЭВМ, надежность); проблемы искусственного интеллекта (распознавание образов, распознавание речи, автоматическая поддержка решения проблем, кластеризация, адаптивные системы); применение ЭВМ.

Имеются явные свидетельства того, что в последние годы в информатике, как она понимается во Франции и Германии, происходит переоценка стоящих перед ней проблем: на передний план все более выступают проблемы логической переработки информации, которая невозможна без обращения к смысловой стороне данных, а технические и технологические проблемы – из-за того, что они оказались более простыми для решения – отступают на второй план. Этот процесс ускорился с началом разработки компьютеров пятого поколения, которые ориентированы на решение проблем искусственного интеллекта и в которых главная роль отводится не базам данных, а базам знаний. Таким образом, «вычислительная» информатика все более сближается с нашей, «семантической» информатикой, которая тоже расширяет границы, включая в круг своего внимания данные и информацию, циркулирующие не только в сфере интеллектуальной коммуникации, но и в производстве и управлении.

Автоматизация различных информационных процессов становится возможной лишь тогда, когда мы способны моделировать или имитировать – хотя бы грубо – механизмы выполнения этих процессов людьми. Изучением и обобщением такого рода механизмов – главным образом, ради их воспроизведения при помощи тех или иных технических средств, включая и компьютеры – с самого начала занимается информатика, которая при этом широко использует результаты, полученные в других науках и научных дисциплинах: библиотковедении, языкознании, психологии, математической логике, семиотике. Из сказанного следует, что основы информатики, имеющей дело преимуще-

ственно с наиболее сложными видами логической переработки информации (например, такими, как перевод текстов с одного естественного языка на другой, реферирование, аннотирование и индексирование научных документов, составление обзоров и т. п.), образуют теоретическую основу информационной технологии и практической деятельности по автоматизированной переработке информации.

Предмет и объекты исследования

Предметом изучения информатики как научной дисциплины являются структура, свойства и закономерности *семантической* информации, процессы ее сбора, переработки, хранения, поиска, распространения (передачи) и использования.

Определение «семантическая», которым мы уточняем содержание понятия «информация», необходимо, поскольку в русском языке информация трактуется весьма широко. Семантическая информация, будучи социальной, т. е. циркулирующей в обществе, существенно отличается от вероятностной, или *шенноновской* информации, в которой определяющую роль играет ее знаковое представление, а смысл игнорируется. Во всем многообразии социальной информации определение «семантическая» позволяет выделить информацию, выраженную на естественном языке и имеющую смысл, доступный для логического восприятия. Это определение было бы излишним, если бы в языковой практике четко различали понятия «информация» и «данные», которые часто употребляются как синонимы в литературе по вычислительной технике и программированию.

Данные – это представление фактов и понятий в форме, пригодной для их передачи, обработки и интерпретации (т. е. толкования, объяснения, раскрытия смысла), а *информация* – это смысл, который человек приписывает данным на основании известных ему правил их представления. Среди видов семантической информации научная информация играет особенно важную роль, так как она неразрывно связана с наукой, в повышении эффективности которой все в большей степени заинтересовано общество. Поскольку научная информация обладает всеми свойствами семантической информации и сверх того своими собственными, изучение свойств и закономерностей семантической информации осуществлялось в основном на материале информации научной. Эффективность развития науки в значительной мере зависит от того, насколько быстро и полно используется в науке и про-

изводстве получаемая научная информация. Именно общественная потребность в этом и обусловила выделение научно-информационной деятельности в самостоятельный вид научного труда, а информатики – в самостоятельную научную дисциплину.

Объектами исследований информатики служат методы и средства, используемые для сбора, переработки, хранения, поиска и распространения (передачи) семантической информации, а также особенности ее использования различными категориями потребителей. Одним из главных объектов служит система научной коммуникации, от которой в значительной степени зависят темпы развития науки и техники, а также эффективность всей системы интеллектуальной (информационной) коммуникации в обществе.

Под системой интеллектуальной (информационной) коммуникации здесь понимается совокупность всех методов и средств, которые используются для распространения (передачи) семантической информации, получаемой и используемой во всех сферах деятельности. Информатика призвана значительно повысить эффективность функционирования этой системы.

Итак, информатика как научная дисциплина изучает предметы и явления, присущие лишь человеческому обществу. Это означает, что информатика является общественной, а не естественной или технической научной дисциплиной.

Информатика и другие науки и научные дисциплины

Информатика сложилась на стыке ряда наук и научных дисциплин, из которых основными являются: библиотеко-, библиографо-, книго-, архиво- и науковедение, документалистика, структурная лингвистика, семиотика, социальная психология, математическая логика, теория информации, кибернетика, математическое программирование и вычислительная техника. Достижения и методы этих наук и научных дисциплин в разной степени используются в информатике. Сами они тоже все в большей мере испытывают влияние идей и методов информатики.

В свое время острый полемический интерес представляли взаимосвязи информатики с дисциплинами коммуникативного цикла, в первую очередь, с библиотековедением. Теперь острота этих споров ослабла, а подробного обсуждения требуют взаимосвязи информатики с программированием и вычислительной техникой. Быстрое внедрение

компьютеров во все сферы деятельности создают ложное представление об их всемогуществе, о превращении информационной техники в определяющую силу развития современного общества. Возникают и распространяются вредные технократические иллюзии, что массовое применение все более совершенных компьютеров и других средств информационной техники само по себе решит острейшие экономические и социальные проблемы, стоящие сегодня перед человечеством.

При рассмотрении взаимосвязей между информатикой, программированием и вычислительной техникой необходимо помнить, что для решения на компьютере любой задачи помимо исходной информации обязательны три компонента: 1) общая схема решения задачи; 2) программа, представляющая собой определенную последовательность команд, выполнение которых приводит к решению этой задачи; 3) компьютер и другие связанные с ним технические средства, при помощи которых реализуется программа решения.

Каждый из этих компонентов требует участия людей, обладающих специальными знаниями и навыками. Задачу обычно ставит и формулирует в содержательных терминах специалист в той области, к которой относится задача. Общую схему решения предлагают или разрабатывают информатики, которые обычно пользуются помощью специалиста, поставившего задачу. Программу решения задачи на компьютере на основании предложенного информатиками способа составляют, отлаживают или выбирают программисты. Если же программисты пытаются сами решить задачу пользователя, результат нередко бывает похож на карикатуру, представленную на рис. 1. Ясно, что каждый участник процесса решения информационных задач на компьютере в своих исследованиях стремится отвлечься от конкретного содержания решаемых задач и за внешним многообразием распознать сходство или даже совпадение многих из них, чтобы разработать и применить типовые методы, алгоритмы и программы решения.

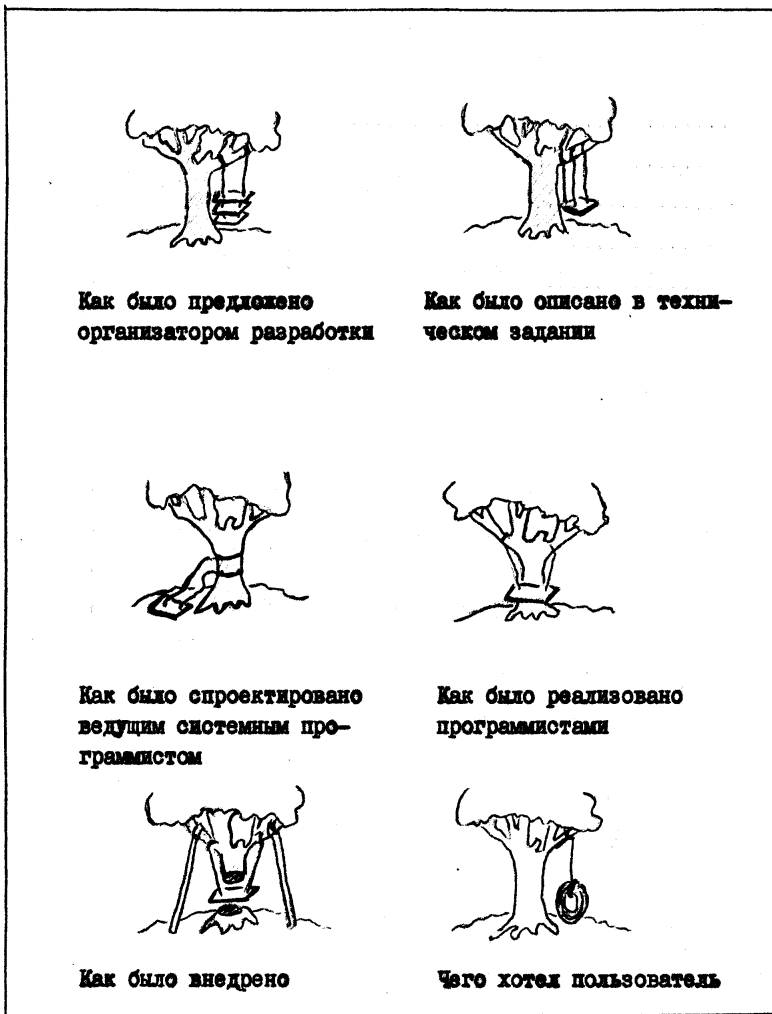


Рис. 1. Барьер непонимания между пользователем и программистами

Информатики стараются найти общие закономерности семантической информации, не зависящие от конкретных отраслей ее получения и/или использования, и разработать методы алгоритмического решения типовых информационных задач.

Программисты отвлекаются от семантического аспекта обрабатываемой информации и рассматривают лишь ее знаково-структурный аспект, т. е. воспринимают информацию как цепочки тех или иных символов, над которыми в компьютере должны производиться определенные действия для получения требуемого результата. Специалисты по вычислительной технике трудятся над созданием таких компьютеров, которые были бы в состоянии эффективно работать под управлением любых программ и с любыми массивами данных.

Таким образом, информатика, программирование и вычислительная техника имеют разные предметы, изучают разные объекты или разные стороны одних и тех же объектов и решают разные задачи. Но достижения этих дисциплин используются совместно для решения практических задач при помощи вычислительных машин. Это означает, что информатика, программирование и вычислительная техника тесно соприкасаются, но не пересекаются и тем более не поглощают друг друга. Поэтому нет никаких оснований считать, что какая-либо из этих научных дисциплин менее важна, чем другие. Можно лишь говорить о том, что уровень развития той или иной из них отстает от требований сегодняшнего дня.

Необходимо также отметить, что в настоящее время ощущается возрастающая потребность в более тесном взаимодействии информатики с вычислительной техникой и программированием, ибо при разработке вычислительных машин и их программного обеспечения мало учитывались достижения и требования, относящиеся к решению информационно-логических задач в информатике. Эта сфера непосредственно примыкает к направлению комплексных исследований, получившему название «искусственного интеллекта». Они ведутся с целью создания автоматизированных систем, способных выполнять некоторые функции человеческого мозга. В них используются методы и результаты многих научных дисциплин, включая информатику.

Информация – знание – наука

"Информация есть информация, а не материя и не энергия", – писал Н. Винер, который ввел это понятие в науку, отказавшись от его определения, считая его сродни таким общим категориям, как движение, жизнь, сознание и т. п.¹ Ему вторит известный американский специалист в области информатики Т. Сарацевич: "Все недовольны тем, что информационная наука не хочет потрудиться над определением информации... На самом же деле ни одна современная наука не имеет определений своих основных феноменов. В биологических науках нет определения *жизни*, в медицинских – *здоровья*, в физике – *энергии*, в электротехнике – *электричества*, а в ньютоновских законах – *противодействия*. Это просто основные явления, и эта их первичность и служит им определением"².

В последние годы в информатике обсуждается возможность разных подходов к понятию информации, которая может трактоваться как управляющий сигнал в автоматической системе, ряд символов, переданных по каналам связи, или же как знание, отчужденное от индивидуального его носителя и обобществленное в процессе коммуникации. В первом случае мы имеем дело с материальными объектами – электрическими сигналами, изображениями, буквами, а во втором – со знаниями, которые относятся к сфере идеального, поскольку отражают человеческое представление о мире. Это различие важно, так как в данном курсе лекций речь идет о семантической информации, т. е. об информации имеющей определенный смысл и циркулирующей в обществе.

¹ Винер Н. Кибернетика, или Управление и связь в животном и машине. – Изд. 2-е. – М.: Советское радио, 1968. – С. 201.

² Saracevic T. Report at conference on "Education for Information Science – Strategies for Change in Library School Programs", Albany (N. Y.) 1977 // Journal for Information Science. – 1979. – Vol. 1, N 2. – P. 76.

Информация и данные

Термин «информация» прочно вошел в самые различные отрасли современной науки и техники, в повседневную жизнь. Однако точное и исчерпывающее его определение остается одной из труднейших научных задач.

В *повседневной жизни* информация означает сообщение, осведомление о положении дел, сведения о чем-то. Для *философов*, склонных рассматривать информацию в одном ряду с такими категориальными понятиями, как материя и энергия, информация – это передача, отражение разнообразия в любых объектах и процессах живой и даже неживой природы. *Математики, физики и специалисты по системам связи* рассматривают информацию как фактор и меру уменьшения, снятия неопределенности в результате получения сообщения, а *кибернетики* – как сообщение, неразрывно связанное с управлением в единстве синтаксических, семантических и прагматических характеристик. *Биологи*, как и философы, довольствуются трактовкой информации как того, что отражает, ограничивает многообразие, но в отличие от философов относят это понятие только к живой природе. Для *социологов* важны аксиологические (т. е. связанные с ценностью, полезностью) свойства информации, а для специалистов по *программированию и вычислительной технике* наиболее существенным является знаковое представление информации, и т. д.¹

Для специалистов в области *информатики*, изучающей структуру и общие свойства и закономерности информации, а также процессы ее сбора, обработки, хранения, поиска, распространения и использования, такие довольно односторонние трактовки явно недостаточны. Особенно неприемлемо игнорирование самого главного в информации – ее смыслового, семантического содержания. Требуется также уточнение значения и взаимосвязей терминов «сведения»², «данные», «информация» и «знания», которые очень часто считаются синонимами и определяются или трактуются друг через друга, в результате чего образуются порочные логические круги. Приведем еще раз наше тол-

¹ Информация // Философский энциклопедический словарь. – М.: Советская энциклопедия, 1983.

² Толкование термина «сведения» см.: Словарь русского языка: В 4-х т. – 2-е изд., испр. и доп. – М.: Русский язык, 1981. – 1984. – Т. 4. – С. 38.

кование терминов «данные» и «информация», их смысловых различий, так как они очень важны для правильного понимания рассматриваемой темы.

Данные суть факты, идеи, сведения, представленные в знаковой (символьной) форме, позволяющей производить их передачу, обработку и интерпретацию (т. е. толкование, объяснение, раскрытие смысла), а *информация* – это смысл, который человек приписывает данным на основании известных ему правил представления в них фактов, идей, сообщений. Такое понимание информации соответствует и этимологии обозначающего его слова (от лат. *information* – разъяснение, изложение).

Структурированная информация, т. е. связанная причинно-следственными и иными отношениями и образующая систему, составляет *знания*. Из этих толкований следует, что если данные воспринимаются и интерпретируются человеком, то они становятся для него информацией. Данные в определенной степени подобны письменному сообщению, передающему какие-то сведения грамотному человеку и остающиеся набором непонятных знаков для неграмотного.

Таким образом, информация – это потенциальное свойство данных, которое может быть реализовано одним воспринявшим их человеком и не реализовано другим. Объектом машинной обработки являются данные, а не информация, так как никакая машина не в состоянии интерпретировать данные, т. е. преобразовать их в информацию, ибо машина не обладает, как человек, необходимым запасом знаний о мире и не может мыслить. Такова диалектическая связь между понятиями «данные» и «информация». Однако в устной и письменной речи часто слово «информация» употребляют вместо слова «данные» как метафору, т. е. в переносном смысле, по аналогии. Такое встречается даже и в этом курсе лекций. Это надо учитывать, когда встречаются высказывания о том, что *компьютер обрабатывает информацию*. На самом деле, речь в данном контексте может идти только об обработке данных.

В обществе циркулируют разные виды информации. Но из них особенно важную роль играет научная информация, так как она неразрывно связана с наукой. Определение «научный» в термине «научная информация» означает, что эта информация удовлетворяет общепринятым в данное время критериям научности (т. е. она объективна, истинна, проверяема и т. п.), но не обязательно получена или использует-

ся только в сфере науки. Если попытаться дать более четкое и полное определение понятию «научная информация», то оно может иметь следующий вид. *Научная информация – это логическая информация, получаемая методами опытно-рационального познания объективного мира в любой сфере деятельности людей, не противоречащая господствующей системе научных представлений и используемая в общественно-исторической практике.*

Свойства информации

Ниже кратко характеризуются важнейшие свойства информации, которые можно условно разделить на собственные и потребительские (аксиологические, прагматические).

Собственные свойства информации:

- неаддитивность: прибавление информации к уже имеющейся не увеличивает ее суммарное количество на величину прибавленной;
- некоммутативность (неперестановочность): суммарное количество полученной информации зависит от последовательности поступления (получения) информационных сообщений $A+B \neq B+A$, где A и B – разные информационные сообщения);
- неассоциативность: количество полученной информации зависит от конкретных сочетаний поступивших информационных сообщений $((A+B)+C \neq A+(B+C))$;
- независимость содержания информации от формы и способа ее записи (фиксации) и представления;
- устаревание во времени.

Потребительские свойства информации:

- неэквивалентность количества и качества (ценности, полезности) информации: ценность полученной человеком информации определяется не количеством снимаемой ею неопределенности, а потребностью этого человека в данной информации, подготовленностью последнего к ее восприятию и использованию;
- неисклезаемость информации при ее использовании (потреблении);
- независимость ценности (полезности) информации для потребителя от количества затрат на ее получение.

С понятием ценности (полезности) информации тесно связаны и такие ее потребительские свойства, как точность, полнота, глубина,

достоверность, надежность, доказательность, новизна, оперативность, эффективность и другие.

Структура информации

Общим свойством данных, содержащих семантическую информацию, является их *дискретность*. Этим они отличается, например, от данных, получаемых в процессе чувственного познания, которые, как и показания приборов, могут быть непрерывными. Если пользоваться лингвистическими терминами, то можно сказать, что в плане выражения и в плане содержания эта дискретность различна. В плане языкового выражения единицами информации являются слова, предложения, отрывки текста, в плане содержания – понятия, высказывания, описания фактов, гипотезы, концепции, теории, законы, учения.

Поскольку отчуждение информации от ее создателей происходит не непрерывно, а дискретно, в виде относительно законченных творческих актов, получающих материальное воплощение, постольку основной единицей информации, ее квантом можно считать произведение. *Произведение* это результат целенаправленной познавательной деятельности, обладающий логической завершенностью и материальным оформлением. Будучи изложено в письменном виде оно становится *документом*.

С дискретностью семантической информации тесно связана и ее структура, которая имеет явно выраженный иерархический характер. В этой структуре также можно различать содержательный и формальный аспекты. При рассмотрении обоих аспектов видно, что чем выше уровень иерархии, тем большую специфику имеет класс информации. В содержательной структуре это связано с ростом кумулятивности информации. Такими классами служат: а) факты; б) гипотезы, теории и концепции; в) основы областей знания; г) мировоззрение.

Формальная структура *данных / семантической информации* также иерархична, как и содержательная. Низшие уровни этой иерархии являются общими для всех *данных / семантической информации*, в которой можно выделить звуки или буквы, слова, фразы, смысловые фрагменты, произведения. На высших уровнях иерархии, при обращении к научным документам и, особенно, к их потокам мы сталкиваемся со спецификой, присущей одной лишь научной информации. Мотивы создания научных документов, пути и способы их распространения и использования находятся в сфере действия особых закономерностей.

Научная литература является высшим уровнем организации в формальной структуре научной информации. Она хотя и обладает всеми внешними атрибутами литературы художественной, представляет собой совершенно особое социальное явление, целиком находящееся в сфере научной коммуникации. Содержательный и формальный аспекты структуры научной информации находятся в известной взаимосвязи. Каждый из названных содержательных классов тяготеет к определенным видам научных документов и отражается в формальной структуре по некоторым не вполне строгим правилам, знание которых одинаково важно для всех работников умственного труда.

Особенности информации

Информация является продуктом человеческого труда. Однако в отличие от других продуктов труда она обладает рядом существенных свойств. Информация не расходуется при ее потреблении. В этом она похожа на орудия труда, которыми можно пользоваться многократно для создания и обработки разных изделий. Но информация не тождественна своему материальному носителю. Поэтому она (в отличие от орудий труда) может быть потреблена неограниченное число раз, ее можно также размножить и одновременно использовать в разных местах для решения всевозможных задач. Единственное ограничение на полезное использование информации устанавливается сроком ее устаревания.

Часто информацию, которая играет все более важную экономическую роль, называют товаром (*commodity*), наивно полагая, что товаром является любой объект купли-продажи. Если согласиться с тем, что информация – товар, то о таком товаре может только мечтать любой торговец: продавая одну и ту же информацию много раз, он продолжает оставаться ее владельцем.

Информация – категория нематериальная, и продается не она сама, а право на ее использование. Информация изначально не предназначается для обмена путем купли-продажи, т. е. не обладает этим определяющим признаком товара. Поэтому на информацию, как и на другие духовные ценности, нельзя безоговорочно распространять законы товарного производства, ибо это породило бы трудно преодолить

мые противоречия с действительностью¹. Например, условия товарного производства требуют массового выпуска объектов купли-продажи, т. е. изготовления их в большом количестве экземпляров. Информация же уникальна по своей природе: одно и то же научное открытие или изобретение, порождающее новую информацию, не нужно делать дважды, а тем более много раз. При товарном производстве одно и то же изделие может выпускаться в одно и то же время разными производителями, которые конкурируют друг с другом на рынке, что отражается в ценах, устанавливаемых ими на это изделие. А цена на изделие (товар), будучи денежным выражением его стоимости, определяется количеством труда, общественно необходимого для изготовления такого изделия при средних общественно нормальных условиях производства и среднем в данном обществе уровне профессиональной подготовленности работников и интенсивности их труда.

Однако на любое научное открытие или изобретение из-за его уникальности (неповторяемости) требуется затратить именно столько труда, сколько на него было затрачено. При этом стоимость информации хотя и зависит от количества труда, затраченного на ее получение, но явно не определяется им. В этом информация схожа с произведениями искусства и литературы. Товарное производство требует определенной стандартизации однотипных объектов купли-продажи. А информация теснейшим образом связана с понятием разнообразия.

Н. Винер подчеркивал, что необоснованный подход к информации как к товару приводит к неправильному пониманию информации и связанных с ней понятий: «Что делает вещь доброкачественным товаром? В сущности, то обстоятельство, что она может переходить из рук в руки, прочно сохраняя свою стоимость, а также тот факт, что отдельные образцы этого товара должны арифметически складываться точно таким же образом, как и уплаченные за них деньги. Способность сохранять себя представляет собой очень удобное свойство доброкачественного товара <...>². С другой стороны, информацию нельзя сохранить столь просто, ибо <...> объем переданной информации относится к неаддитивной величине, называемой энтропией, и отличается от нее своим алгебраическим знаком и возможным цифровым коэффициентом. Подобно тому, как в замкнутой системе энтропия стихийно стре-

¹ Hawkins D. T. The commodity nature of information // Online. –1987. – Vol. 11, N 1. – P. 67–70.

² Винер Н. Кибернетика и общество. – Изд-во иностр. лит., 1958. – С. 120.

мится к увеличению, точно так же информация стремится к уменьшению; подобно тому, как энтропия есть мера дезорганизации, информация есть мера организации. Информация и энтропия не сохраняются и в равной мере непригодны для того, чтобы быть товарами»¹.

Информацию не следует смешивать с информационными изданиями (или продукцией) – книгами, периодическими изданиями, описаниями изобретений, научно-техническими отчетами и другими видами документов, с разнообразными базами данных, копиями документов и т. п., которые выступают в качестве материальных носителей, содержащих информацию. Покупку информационного изделия можно уподобить покупке хлеба: мы покупаем и потребляем хлеб вместе с содержащимися в нем питательными веществами, а не отдельно эти питательные вещества. Таким образом, информационные издания и услуги имеют явно выраженную товарную природу и в силу этого подчиняются законам товарного производства.

Информация научная и техническая

Быстрый рост количества информации, характеризуемый метафорой "информационный взрыв", не замедляется, как это предсказывалось некоторыми учеными на основании общих соображений, а, возможно, даже ускоряется. Это происходит в результате постоянного расширения масштабов познавательной деятельности людей – главным образом за счет информации (данных), получаемой при проведении инструментальных исследований поверхности, недр и атмосферы Земли, океанов, космического пространства, микромира и т. д. Такая информация обычно записывается на машиночитаемые носители и не публикуется в традиционном, печатном виде. Эти данные оцениваются и интерпретируются учеными и специалистами, что обуславливает нарастающий поток публикаций и непубликуемых документов, содержащих научную и техническую информацию.

Из сказанного следует, что обеспечение сбора, обработки, хранения, поиска, распространения и использования возрастающих объемов цифровой информации, получаемой в результате комплексного инструментального исследования Земли и космоса, становится все более важной и трудной задачей системы научной и технической инфор-

¹ Там же. – С. 122–123.

мации. Удовлетворительное решение этой задачи, имеющей глобальный характер, может быть найдено только в результате международного сотрудничества.

Кроме того, возрастанию потоков информации в большой степени способствует информатизация, начавшаяся в 70-х годах прошлого столетия. *Информатизация, ее развитие и углубление проявляются в быстром увеличении потоков технико-экономической, финансовой, технологической и иной информации, циркулирующей внутри предприятий и организаций, между ними и между разными странами.* Эту информацию также необходимо собирать, обрабатывать, хранить, искать и передавать, для чего требуются соответствующие методы и системы. Таким образом, у современного общества нет никаких оснований ослаблять внимание к проблеме научной и технической информации, которая становится все более трудной для решения.

Под *научной информацией* мы понимаем любую достоверную информацию, относящуюся к любой области естественных, технических и общественных наук, полученную научными методами познания и оцененную учеными или их коллективами (учеными и редакционными советами, редакционными коллегами журналов и т. п.). Под *технической информацией* понимается любая достоверная информация, относящаяся к конструированию, производству и эксплуатации технических объектов, оцененная специалистами соответствующей отрасли (разработчиками, их руководителями, редакторами) или их коллективами.

Научной и технической информацией (НТИ) в дальнейшем обозначается совокупность достоверных сведений, относящихся к развитию науки и техники и используемых в любой сфере жизни общества. Этот термин отражает лишь одну сторону многогранного понятия информации, а именно – способ ее получения для отграничения НТИ от бытовой, массовой, коммерческой, учетно-финансовой и другой социальной информации. Таким образом, НТИ – это информация о результатах исследований природы, общества, мышления, технических систем, сведения о машинах, материалах, технологических процессах и т. п.

Научную и техническую информацию нельзя отождествлять со *знаниями*, которые являются результатом логической переработки информации, имеют структурированный характер и соответствуют обще-

признанным научным представлениям (*научной парадигме*) на данном этапе развития общества. Важными для понимания социальных механизмов распространения научной информации являются такие ее и только ей присущие специфические свойства, как кумулятивность, старение и рассеяние.

Кумулятивность научной информации связана с преемственностью и интернациональным характером науки. Если бы каждый ученый, ученые каждой страны и эпохи должны были самостоятельно накапливать необходимые знания, заново открывать законы природы, общества и мышления, то вряд ли наука могла бы развиваться нынешними быстрыми темпами. Достижения ученых всего мира и всех предшествующих поколений являются фундаментом, на котором строят свою работу современные ученые. Поэтому каждое поколение ученых занято не только получением новых научных данных и знаний, но и специальной работой по систематизации, оценке и обобщению научной информации, созданной их коллегами и предшественниками, с тем чтобы сделать эту информацию более доступной не только для современников, но и для новых поколений ученых.

Это становится возможным благодаря тому, что научная информация обладает свойством кумулятивности, т. е. допускает более краткое обобщенное изложение. Со временем все второстепенное, частное отсеивается, а главное, основное получает простое и краткое выражение. Часто то, на изложение чего сто лет назад требовался целый трактат, теперь можно объяснить в нескольких словах с помощью двух-трех формул. Процесс генерализации, или концентрации научной информации во времени обеспечивает переход научного знания ко все более высоким уровням абстракции.

Важно учитывать, что свойством кумулятивности научная информация разных областей знания обладает в разной степени. В точных и естественных науках, особенно таких, как математика и физика, располагающих развитым формализованным языком и строгими правилами вывода, кумулятивность проявляется сильнее. В большинстве общественных и гуманитарных наук это свойство выражено в меньшей степени.

Старение научной информации, столь очевидное на первый взгляд, в действительности оказывается сложным для понимания. Мы привыкли оперировать с понятием старения документов, измерять его

периодами, в течение которых какая-то часть содержащейся в нем информации перестает использоваться. Тот факт, что на произведения Архимеда, Ньютона или Ломоносова сейчас уже почти не ссылаются, вовсе не означает, что содержащаяся в них информация совершенно устарела. В строгом понимании этого свойства полностью устаревает лишь та информация, которая с появлением новой оказывается неверной, т. е. перестает быть научной информацией. В наибольшей мере старению подвержены гипотезы, концепции, теории, и в меньшей – научные факты и основные положения областей знания. Однако, когда говорят о старении научной информации, чаще всего имеют в виду не полное ее старение, а уточнение, более строгое, сжатое и обобщенное изложение в процессе создания новой информации. Такое понимание старения связано со свойством кумулятивности научной информации.

Свойство старения информации находится в сложной зависимости с устареванием произведений, в которых она впервые была изложена, и доступностью их изданий для специалистов. Примерами, подтверждающими наличие такой связи, могут служить переиздания произведений и их переводы с одного языка на другой. Они способствуют не только более широкому распространению произведений, но и являются важной частью механизма, задерживающего процесс их старения. Этот механизм обеспечивает долголетие наиболее ценной информации в ее первоначальной (не "переупакованной") форме.

Рассеяние научной информации проявляется в том, что одни и те же ее содержательные единицы – понятия, высказывания, факты, гипотезы, теории, учения – по-разному и в разных контекстах используются в различных произведениях. Будучи первоначально сгруппированы их авторами в соответствии с внутренней логикой тех произведений, в которых они первоначально были опубликованы, они впоследствии получают новую жизнь в произведениях других авторов. Там они располагаются среди других содержательных единиц информации (вводимых в текст этими авторами или почерпнутых из других источников) и несут иную смысловую нагрузку. Это свойство хорошо изучено на макроуровне – сформулированы закономерности рассеяния научных публикаций по периодическим изданиям (С. Бредфорд, Б. Веккери, Б. Брукс). Дальнейшее изучение этого свойства научной информации (на уровне идей и фактов, слов и словосочетаний, имен и ссылок) открывает пути к постижению внутренних механизмов релевант-

ности и пертинентности – центральных понятий информационного поиска.

Наука как социальное явление

Когда говорят и пишут "наука", чаще всего имеют в виду результаты научных исследований или совокупность научных знаний. Но наука это еще и люди, получающие эти результаты и создающие это знание – математики, физики, химики, биологи, психологи, философы, социологи, историки, филологи, специалисты технических и других наук, которые объединены в национальные отряды, составляющие российскую, американскую, японскую науку. Если в первом аспекте наука изучается логикой науки, то во втором – ее социологией. В совокупности они составляют дисциплину, которая у нас называется науковедением, а на Западе – наукой о науке. Многие ее проблемы пересекаются с информатикой. Наука как социальное явление претерпевает большие изменения, но еще быстрее меняются наши представления о науке, точнее, о совокупной области науки и техники.

В изучении социологических тенденций развития науки в последние десятилетия наблюдается значительный прогресс. Большинство отечественных и зарубежных исследователей в этой области отмечает несоответствие между традиционными концепциями науки и современными ее социологическими моделями. Интересные взгляды на этот счет принадлежат видному английскому физика Джону Займану, известному своими оригинальными работами по социологии науки. По его мнению, все многообразие прежних представлений о науке как социальном явлении можно свести к двум традиционным концепциям – прикладной и теоретической.

Первая из них, называемая им "инструментальной", характерна для представителей политических и экономических кругов. В ней наука и техника, обычно обозначаемые как исследования и разработки, рассматриваются в качестве инструмента для решения социальных и экономических проблем. При достаточных организации и финансировании этот "черный ящик" позволяет получать необходимые консультации, решения, изобретения. Ей противостоит теоретическая концепция науки, трактуемая как модель "открытия" и бытующая в сферах образования, культуры и науки. В соответствии с этой моделью ученый, исследуя законы природы, приходит к новым открытиям, попол-

няющим общечеловеческое знание, которое служит объектом применения в технике.

Надо сказать, что эти концепции в какой-то мере соответствуют нашим представлениям о прикладной и фундаментальной науке, первая из которых служит для непосредственного внедрения ее результатов в народное хозяйство, а вторая меняет наши воззрения на окружающий нас или находящийся внутри нас мир. Обе модели, имеющие достаточно глубокие исторические корни, неадекватны современному состоянию науки и техники. В них отсутствуют существенные элементы для понимания истинного положения этой совокупной области в обществе.

Анализ своей новой модели науки Дж. Займан начинает с рассмотрения ее внутренних параметров, к которым наряду с традиционными элементами "ученый" и "знание" он относит "научное сообщество" (рис. 2). На пересечении этих элементов обозначены такие понятия, как "незримый коллектив", "научная парадигма", "научная коммуникация" и ряд других, которые были введены в научный оборот в последние десятилетия и существенно продвинули наше понимание социального механизма науки.

Эти понятия он связывает с именами известных западных социологов науки Майкла Поланьи, Роберта Мертона, Томаса Куна и философа Карла Поппера. Конечно, список этот неполон даже в рамках западной социологии: в нем отсутствуют имена Джона Бернала, Дирека Прайса, Уильяма Гарвея, Белвера Гриффита, Генри Смола. Если же обратиться к отечественному науковедению, то вряд ли можно не упоминать работы Г. М. Доброва, В. В. Налимова, Н. И. Родного, С. Р. Микулинского, Э. М. Мирского и других исследователей, внесших существенный вклад в современное представление о положении науки в обществе.

Неадекватность традиционных концепций науки Дж. Займан справедливо видит в неправомерности противопоставления науки и техники, теоретических исследований и прикладных разработок, которые в настоящее время уже не имеют четких демаркационных линий. Более важное значение для понимания современного положения науки и техники в обществе он придает особенностям научного сообщества, развивающегося в определенном направлении с середины XVII в.

М. Поланьи (его книга "Личностное знание" 1958 г. издана у нас в 1985 г. издательством "Прогресс") впервые показал, что научное со-

общество – уникальный продукт цивилизации. В нем нет постулированной системы формальных убеждений, но есть неписаная этика, определяющая безоговорочные нормы поведения ученых. Как почти в любой общественной институции, в научном сообществе существуют свои ранги престижа и иерархия, но они устанавливаются в соответствии с оценкой достижений ученого и его признанием в своей среде. Их престиж не наследуется и не приобретается какой-либо вне-научной деятельностью. Быть ученым, по его мнению, значит постоянно быть учеником, а достичь компетентности можно лишь под руководством учителей.

«В ведущих исследовательских школах воспитываются наиболее важные предпосылки научного открытия. Они раскрываются перед способным студентом в ходе каждодневного труда учителя: интуиция, которой он руководствуется в работе, передается ученику... Именно поэтому так часто великие ученые в качестве подмастерьев следуют за великими учителями».

Будучи идеалистом, М. Поланьи отрицает определяющее влияние материальных потребностей общества на развитие науки. Эта его позиция особенно четко выявилась в споре с Дж. Берналом, который в этом вопросе придерживался материалистических взглядов. «Мы должны подтвердить, писал М. Поланьи, что сущность науки есть тяга к знанию и что практическая польза знания не должна нас занимать... Мы, ученые, посвятили себя служению ценностям, более дорогим и делу более насущному, чем материальные блага».

Р. Мертон кодифицировал принцип независимости научного исследования и выделил в нем четыре аспекта: универсальность, коммунальность, беспристрастность и скептицизм. *Универсальность* предполагает доступность научной профессии каждому, кто обладает соответствующими способностями, независимо от расы, национальности, происхождения, принадлежности к классу или личных качеств. *Коммунальность* подразумевает, что знание является общественным продуктом, выделяемым из наследия прошлого и свободно передаваемым в будущее. Можно купить патент и извлекать прибыль из использования изобретения, но не из теории, которая привела к изобретению. Научная теория является достоянием общества, а полная и свободная информация о научных достижениях выступает необходимым условием прогресса в обновлении знаний.

Беспристрастность оценки научных достижений также является одним из условий развития науки. Требование объективности имеет твердый базис в социальном характере науки, и именно это определяет высокую репутацию людей науки. Ученые, разумеется, стремятся к известности, но сама природа научного исследования дает науке гарантии против фальсификации его результатов. *Скептицизм* подразумевает организованную систему проверки этих результатов, добровольный отказ от принятия их на веру, отказ от слепой приверженности принятым догмам или прежним авторитетам.

Т. Кун ввел в научный оборот понятия научной парадигмы и нормальной науки, которые позволили по-новому взглянуть на деятельность научного сообщества. Понятие научной парадигмы, сложное и по-разному определяемое автором, возникло по аналогии с грамматической парадигмой спряжения глаголов или склонения существительных. В книге 1962 г. "Структура научных революций" (перевод 2-го издания 1970 г. опубликован издательством "Прогресс" в 1975 г. и 1976 г.) он пишет: «Под парадигмами я подразумеваю признанные всеми научные достижения, которые в течение определенного времени дают модель постановки проблем и их решений научному сообществу...

Успех парадигмы, будь то аристотелевский анализ движения, расчеты планет у Птолемея, применение весов Лавуазье или математическое описание электромагнитного поля Максвеллом, вначале представляет собой в основном открывающуюся перспективу успеха в ряде проблем особого рода. Заранее неизвестно исчерпывающе, каковы будут эти проблемы. Нормальная наука состоит в реализации этой перспективы по мере расширения частично намечаемого в рамках парадигмы знания о фактах. Реализация указанной перспективы достигается также благодаря все более широкому сопоставлению этих фактов с предсказаниями на основе парадигмы и благодаря разработке самой парадигмы

Немногие из тех, кто фактически не принадлежит к числу исследователей в русле зрелой науки, осознают, как много очистительной работы такого рода осуществляется в рамках парадигмы и как приятно выполнять эту работу... Именно наведением порядка занято большинство ученых в ходе их научной деятельности. Последняя и составляет то, что я называю здесь нормальной наукой... Ученые в русле нормальной науки не ставят себе цели создания новых теорий, обычно к

тому же они нетерпимы и к созданию таких теорий другими. Напротив, исследование в нормальной науке направлено на разработку тех явлений и теорий, существование которых парадигма заведомо предполагает» (с.11, 43-44).

.

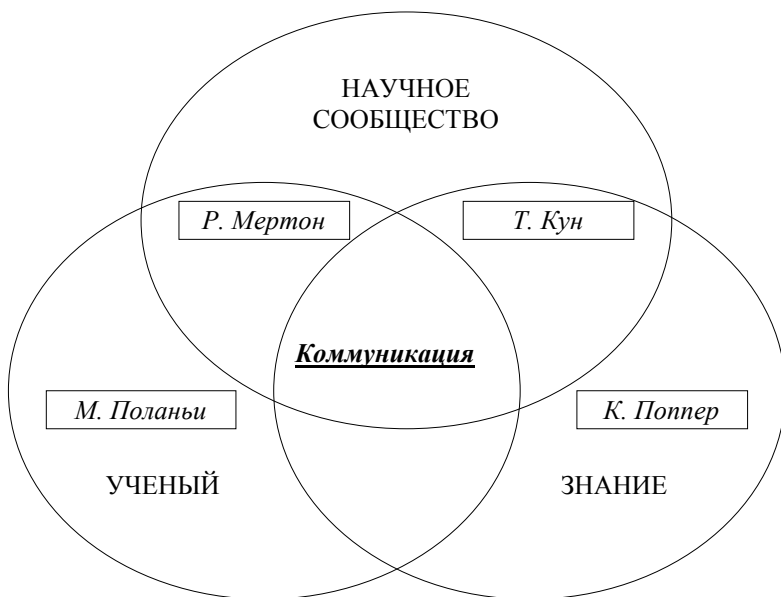


Рис. 2. Внутренние параметры науки (по Дж. Займану)

Таким образом, научные революции, понимаемые Т. Куном как смена одних парадигм другими, означает ломку привычных коммуникаций между учеными. Как правило, научная революция меняет не только круг проблем, которые научное сообщество считает достойными исследования, но и научный язык, на котором описываются результаты исследований, и методы их проведения, и верификации их истинности. Поэтому такая ломка оказывается крайне болезненной для представителей определенного научного сообщества, а само это явление стало весьма ощутимым для социального здоровья науки в связи с увеличением темпов ее роста и частотой смены парадигм, которая в наше время все больше захватывает и механизм передачи знаний в науке.

К. Поппер, английский философ, логик и социолог, является создателем так называемого критического рационализма. В основе этой философской концепции лежит абсолютизация принципа фальсификации как главного метода отделения научного знания от ненаучного. Этот принцип предполагает отказ от признания объективной истинности научных знаний, релятивизм в истолковании его роста, конвенциализм в трактовке исходных оснований знаний и объективно-идеалистическую теорию "трех миров": мира физического, мира субъективного знания и мира ментального объективного знания. К. Поппер доказывает, что наука расширяет наши знания не путем верификации существующих теорий, а путем их фальсификации. Его книга "Логика научного исследования" (1934 г.) и отрывки из книг "Предположения и опровержения" и "Объективное знание" (1963 г. и 1972 г.) опубликованы на русском языке в 1993 г. издательством "Прогресс" под заглавием "Логика и рост научного знания".

В книге "Объективное знание" он пишет: «Феномен человеческого знания, вне всякого сомнения, является самым большим чудом в нашей Вселенной. Оно представляет собой проблему, которая будет решена не скоро. Со времен Декарта теория человеческого знания была в основном субъективистской: знание рассматривалось как особо надежный вид человеческой веры, а научное знание как особо надежный вид знания. Положения этой книги порывают с традицией, восходящей еще к Аристотелю, с традицией общепринятой теории знания как здравого смысла, которую я рассматриваю как субъективистскую ошибку. Эта ошибка господствует в западной философии. Я сделал

попытку искоренить ее и заменить объективной теорией преимущественно предположительного знания».

Многие западные теоретики информатики в своих гносеологических построениях опираются на онтологическую схему К. Поппера, которая привлекает их наличием в ней мира объективного знания как среды, в которой происходит информационная деятельность. По мнению Б. Брукса, например, ученые-естественники и инженеры трудятся в сфере "физического мира", гуманитарии изучают "мир субъективного знания" и его связи с "физическим миром", а результаты их деятельности помещаются в "ментальный мир", рамками которого ограничивается сфера математиков-теоретиков.

Таким образом, сущность информационной деятельности заключается в сборе и организации записей ментального "мира" объективного знания для дальнейшего их использования, а теоретическая задача информатики в изучении, описании и объяснении взаимодействия между "мирами" субъективного и объективного знания и в оказании содействия в организации знаний. Поскольку записи человеческого знания у К. Поппера независимы от познающего субъекта, объективное знание доступно для всеобщего изучения, которое и создает предпосылки для возникновения информатики как науки.

Центральное место в науковедческих исследованиях последних десятилетий занимает концепция научной коммуникации, понятие которой не случайно и в модели Дж. Займана находится на скрещении трех внутренних измерений науки. Концепция научной коммуникации была глубоко разработана в 60-е и 70-е годы. Она получила яркое выражение в трудах Д. Прайса, У. Гарвея, Б. Гриффита и многих других ученых. Ей была посвящена и монография А. И. Михайлова, А. И. Черного и Р. С. Гиляревского "Научные коммуникации и информатика" (1976 г.).

В рамках этой концепции были выяснены закономерности передачи научной информации внутри научного сообщества и за его пределами, установлено, что система научной коммуникации является важной частью социального механизма науки, что она служит одним из существенных объектов исследования информатики. Были глубже поняты процессы оценки новизны и значимости результатов научных исследований, роль публикаций в журналах и их рефератов как формальных каналов передачи знания, разработаны методы анализа цитирования для поиска нужной информации, определения вклада отдель-

ных ученых и школ в науку и динамики развития научных направлений.

В тенденциях самой научной коммуникации были вскрыты противоречия, служащие фактором развития всей информационной сферы. Одним из важных достижений в изучении научной коммуникации был вывод о том, что эффективное развитие науки в значительной степени зависит от прогресса в системе научной коммуникации.

К трем внутренним измерениям науки, по мнению Дж. Займана, следует добавить четвертое – техническое оснащение, которое и теперь играет существенную роль в исследованиях и разработках, а в ближайшем будущем займет еще большее место. Уже сейчас электронный микроскоп и миникомпьютер входят в необходимое оборудование каждой естественнонаучной лаборатории. В дальнейшем сложные приборы станут использоваться во всех областях науки. Они будут способствовать смягчению традиционного противопоставления науки и техники.

Однако наибольший интерес в новой модели науки вызывают не эти, уже довольно глубоко изученные, ее внутренние измерения. Значительно более существенными являются ее внешние аспекты, или элементы, связывающие ее со сложной жизнью современного общества (рис. 3). Такими аспектами служат научная политика и экономика науки, управление наукой и техникой, технический прогресс и пропаганда научных достижений, тесно связанная с системой подготовки научных кадров и образованием как таковым. Следовательно, в модель науки оказываются вовлеченными промышленность и народное хозяйство в целом, органы государственного управления, включая специальные подразделения для управления наукой и техникой, системы высшей и средней школы, а также система массовой коммуникации с ее разветвленными средствами информации и пропаганды.

Описываемая модель предполагает, что на науку будет оказывать все более значительное влияние экономика, политика, идеология, культура, мораль, другими словами, вся надстройка, определяемая общественным строем государства. При всей ограниченности и условности рассматриваемой модели науки, как, впрочем, любой графической или даже математической модели, она является шагом вперед по сравнению с традиционными концепциями науки, так как устанавливает новые связи между ее внутренними измерениями и внешними элементами, а также с другими, более широкими аспектами жизни общества.

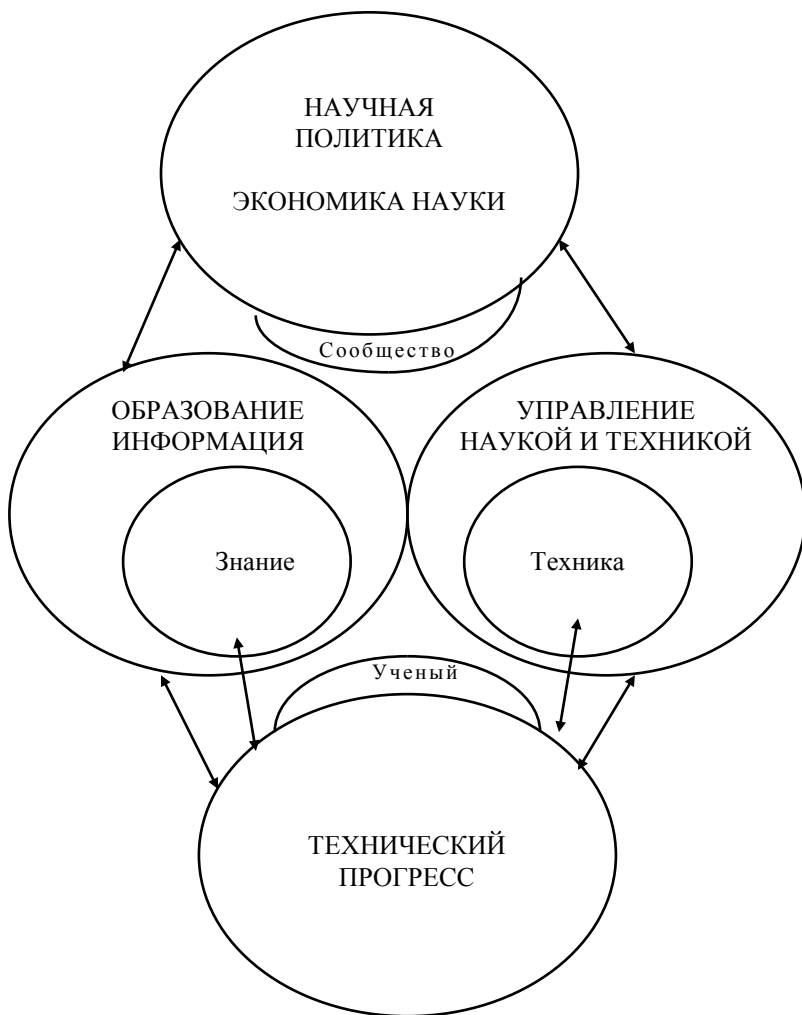


Рис. 3. Внешние параметры науки (по Дж. Займану)

Важной особенностью современной "большой" науки является то новое обстоятельство, что на ее внутреннюю структуру накладываются внешняя иерархия, проявляющаяся в управляющих воздействиях общества, в частности, органов государственного управления. Эта иерархия и ее воздействия становятся все более разветвленными и жесткими по мере роста всех параметров науки.

Перспективы развития науки

Чтобы прогнозировать дальнейшее развитие науки как социального явления недостаточно экстраполировать в будущее действие выявленных тенденций. Более важным представляется анализ тех противоречий, которые могут оказаться решающими факторами этого развития. По всей вероятности, одним из таких противоречий, заложенных в описанной модели, выступает несовместимость этоса науки в его внутренних измерениях и требований "большой науки", связанных с деятельностью внешних элементов ее модели.

Действительно, научное сообщество, его "незримые коллективы" стремятся сохранить выработанные в течение трех столетий нормы поведения ученых. Они направлены на сохранение независимости в выборе объектов и методов исследования в рамках парадигмы, внутреннюю свободу и объективность в оценке научных достижений и присвоении рангов во внутренней иерархии науки. Соблюдение этих норм обеспечивает устойчивость стимулов к публикации научных работ, что, в свою очередь, стабилизирует систему научной коммуникации.

Однако внешние аспекты современной модели науки во многом не соответствуют интуитивным устремлениям научного сообщества. Научная политика "большой науки" не всегда следует логике исследования, направленного на понимание законов природы, общества и мышления. Она направляет исследования и разработки на удовлетворение насущных потребностей человечества, на решение таких глобальных проблем, как поиски новых энергетических и сырьевых ресурсов, источников питания, регулирование роста народонаселения, охрана окружающей среды. Государства стремятся к использованию научных достижений в целях повышения эффективности производства, укрепления экономики и обороноспособности своих стран. Решение этих задач требует перехода от свободного экстенсивного развития науки к планируемому интенсивному развитию. Однако социологиче-

ский смысл происходящих в "большой науке" процессов не вполне однозначен.

Прежние тенденции, восходящие к "малой науке" университетского типа, ведут к неконтролируемому росту числа небольших исследовательских групп, их интернациональному взаимодействию, гласности, т. е. открытой публикуемости результатов научного труда. Поскольку научное сообщество столетиями развивалось в рамках этих традиций, они обеспечивают высокую эффективность труда по научному познанию мира и приращению научного знания, а также стабильность самого социального института науки.

Новые тенденции, связанные с быстрым развитием "большой науки" промышленного типа, устремлены к решению технологических задач и во многом насаждают в организации науки методы крупного производства. Это выражается в укрупнении научных коллективов и тематики их исследований и разработок, жесткости административного управления ими. При этом коллективы, занятые разработкой комплексных целевых и других важных государственных программ, состоят далеко не только из исследователей и разработчиков. Они включают большое число инженерно-технических работников, наряду с учеными являющихся полноправными участниками выполнения научно-технической программы.

В сфере "большой науки" предпринимаются усилия к повышению эффективности научного труда за счет лучшего его технического оснащения, повышения уровня подготовки специалистов, реорганизации формальных каналов научной коммуникации. Это необходимо, так как прежние резервы роста числа людей, занятых в науке, и увеличения ассигнований на исследования и разработки оказываются исчерпанными.

В развитых странах наблюдается стабилизация, а в некоторых странах сокращение численности научных кадров, сокращение бюджетных ассигнований на науку, уменьшение темпов роста подготавливаемых специалистов. Это дает повод некоторым социологам говорить о потере общественного доверия к науке, падении ее социального престижа.

Социологи науки еще в 60-е годы выяснили, что все показатели науки как социального явления за последние триста лет росли экспоненциально, причем каждые полстолетия их абсолютные значения увеличивались на порядок. Это относилось и к ассигнованиям на научные

исследования, и к числу людей, занятых в сфере науки, и к объемам опубликованной ими научно-технической литературы. Такой быстрый темп роста этих показателей науки при значительно более медленном изменении валового национального продукта и численности занятого населения ведет к существенному увеличению доли, приходящейся на науку.

Это хорошо видно по росту затрат на науку в США, которые за 1920–1950 гг. выросли по отношению к валовому национальному продукту в 10 раз и уже к середине 30-х годов достигли 3 %. Экстраполяция такого роста уже тогда приводила к абсурду, так как уже к началу XXI в. весь национальный продукт поглощался бы затратами на науку. Подобная картина возникала бы и при дальнейшем сохранении темпов роста числа научных работников. Если бы число людей, занятых в науке, продолжало увеличиваться в 10 раз каждые полстолетия, а население Земли за эти периоды только удваивалось бы, уже в первой четверти нашего века все люди стали бы работать в науке, что также абсурдно. Из этих подсчетов были сделаны выводы, что рост как абсолютных, так и особенно относительных показателей должен замедлиться, а затем и прекратиться. Статистика последних лет подтверждает справедливость этих выводов.

Этот процесс в первую очередь затрагивает крупные научные учреждения, ориентированные на государственные субсидии. Более жизнестойкими оказываются небольшие научные коллективы без жесткой административной структуры и долговременной программы исследований и разработок, а также исследовательские группы в высших учебных заведениях. Научное сообщество может стать более "интимным". Узкие научные кружки будут превалировать над обширными научными школами с менее демократической структурой. Небольшие рабочие симпозиумы без обязательной публикации материалов будут вытеснять малоэффективные международные конгрессы с тысячами участников. Это поможет науке следующих десятилетий стать более лабильной, обеспечить гибкость организации, способность быстро откликаться на изменение и расширение фронта исследований, проникать в те сферы, которые подолгу выпадают из поля зрения "большой науки".

Подводя итог, можно сказать в терминах Т. Куна, что научное сообщество будет стремиться сохранять способность к совершению научных революций, тогда как "большая наука" промышленного типа,

совокупная сфера науки и техники будет все больше сосредотачиваться на поддержании "нормальной науки". Предполагаемые социологические изменения в науке не могут не сказаться и на всех звеньях научной коммуникации. Расширительные тенденции новой модели науки имеют для нас большое значение, так как научная информация в новых условиях становится все более необходимой не только ученым и инженерам, но представителям всех профессий.

Интеллектуальная коммуникация

Основные понятия

Интеллектуальной коммуникацией называют обмен информацией между индивидами посредством общей для них знаковой системы при помощи жестов, устной речи, письменности и т. п. Различают сферы массовой коммуникации (основными средствами которой являются печать, радио, телевидение) и научной коммуникации.

Под научной коммуникацией мы понимаем совокупность процессов представления, передачи и получения научной информации, образующих основной механизм существования и развития науки. Во всех процессах научной коммуникации участвуют ученые и специалисты. Степень их участия различна в зависимости от каналов коммуникации, которые делят на формальные и неформальные.

К неформальным относят непосредственный диалог между учеными и специалистами о проводимых ими исследованиях и разработках, посещение лабораторий коллег и научно-технических выставок, выступления перед аудиторией, обмен письмами и препринтами.

К формальным каналам относят процессы обмена научной информацией посредством литературы. В них непосредственное участие ученых и специалистов, создавших научную информацию, значительно меньше, так как эти процессы оформились в самостоятельные виды деятельности: научно-информационную, библиотечно-библиографическую, архивную, редакционно-издательскую, журналистскую.

Следует отметить, что лишь первый канал специфичен для научной коммуникации, а остальные выполняют роль общих информационных каналов для всех видов семантической информации. В этом направлении развивается и научно-информационная деятельность, методы и средства которой все шире используются во всей сфере информационной коммуникации.

В связи с быстрым развитием технических средств коммуникации на Западе существуют различные теории, определяющие роль этих средств и самой информационной коммуникации. Среди них особого внимания заслуживает "маклюэнизм", философская концепция средств общения, сформулированная в 60-е годы канадцем Маршаллом Маклюэном. Три его последовательно выпущенные книги, ("Исследования в области средств коммуникации", "Гутенбергова галактика" и "Средства понимания: расширение возможностей человека") вызывали бурные дискуссии среди философов и социологов, Его называли "оракулом электронного века" и сравнивали с Ньютоном, Дарвиным и Фрейдом.

Отвлекаясь от этих эмоциональных оценок, нельзя не признать, что М. Маклюэн первым предпринял серьезную попытку исследовать роль новейших средств коммуникации в человеческом обществе и выявить их свойства и закономерности развития. Ему удалось сделать глубокие и верные наблюдения относительно природы, значения и сферы использования таких средств массовой коммуникации, как кино, радио, телевидение, печать, звукозапись, телефонная и телеграфная связь и т. п.

Широко трактует М. Маклюэн значение книгопечатания для последующего развития человеческого общества. Он считает, что с книгой в общественную жизнь внедрились принципы непрерывности, единообразия и повторности, которые стали основой вычислений и торговли, промышленного производства, искусства и науки.

Основная его идея заключается в том, что средства коммуникации являются определяющим фактором в развитии человеческого общества, а общественная жизнь в большей мере зависит от характера и свойств средств передачи информации, чем от содержания передаваемых сообщений. Крайность таких утверждений неоднократно подвергалась справедливой критике. Однако маклюэнизм заслуживает внимания как одна из немногих попыток создания всеобщей теории коммуникации.

Система научной коммуникации

Система научной коммуникации является высокоупорядоченной социальной системой, основанной на взаимодействии людей, функционировании социальных институтов и подчинении определенным общественным нормам. Элементы этой системы динамически

связаны таким образом, что изменение одного из них неизбежно влияет на функционирование других. Взаимосвязь эта проявляется, по крайней мере, в трех аспектах: функциональном, социальном и экономическом.

Определяя свойства системы научной коммуникации в духе требований общей теории систем, следует констатировать, что она является *открытой* системой, так как обладает всеми соответствующими признаками. Система научной коммуникации не изолирована от окружающей среды, испытывает внешние воздействия, и реакция на них отличается стохастичностью. Она является самоприспосабливающейся, самоорганизующейся, что затрудняет целенаправленное ее изменение. Ее изучение также затруднено тем, что в процессе любого эксперимента система оказывает обратное воздействие на экспериментатора, поскольку он всегда находится внутри нее.

Система научной коммуникации отличается устойчивостью по отношению к внешним воздействиям, на которые она реагирует медленно и незначительно. Это обусловлено стабильностью навыков ученых при решении ими информационных задач, малой изменчивостью основных побудительных мотивов в поведении ученых и специалистов, а также решающей ролью письменности и научно-технической литературы при передаче научной информации в пространстве и времени. Наиболее устойчивое состояние, к которому стремится система в процессе самоорганизации, достигается ею тогда, когда эта система приобретает иерархическую структуру. Она является сложной системой, которая включает множество подсистем с различными и часто противоречивыми целями. При изучении таких систем использовать их описание для прогноза их развития и поведения в новых условиях можно только, если известны основные свойства и закономерности их функционирования,

Среди названных подсистем наибольшее значение для рассматриваемых дисциплин имеет подсистема научной литературы, поскольку научные документы и издания входят в состав основных объектов изучения этих дисциплин и соответствующих им сфер деятельности. Системный подход к изучению документальных источников научной информации предполагает глубокий анализ социальных функций подсистемы научной литературы. Работники умственного труда для повышения эффективности своей деятельности должны знать, каким образом научная литература способствует сохранению единства науки и

поддержанию установившегося ее уровня, преодолению информационных барьеров и другим позитивным процессам.

Сформулированная в работах по информатике концепция *научного документа как совокупности логически завершенного фрагмента научной информации и материального носителя, на котором эта информация закреплена*, позволила всесторонне раскрыть его функции. Они не сводятся к решению только информационной задачи служить средством распространения информации в пространстве и времени, но включают и такие общественные функции, как формирование научного стиля изложения, утверждение приоритета исследователя, стимулирование научной деятельности и обеспечение продуктивности творческого труда.

Одной из важных закономерностей развития средств коммуникации как массовой, так и научной является то, что с появлением новых средств, реализующих процессы коммуникации, прежние не отмирают, а сохраняются в сфере человеческого общения. При этом старые средства, отдав часть своих функций вновь возникшим средствам, как правило, начинают новую жизнь, эффективнее выполняя более ограниченный круг функций. Эту закономерность можно объяснить тем, что новые потребности в средствах коммуникации не ведут к ликвидации тех потребностей, которые удовлетворялись существовавшими средствами.

Потребность в сохранении и распространении знаний, вызвавшая к жизни письменность, не отменила потребности в личном общении, удовлетворяемой устной речью. Нужда в большом числе идентичных экземпляров учебных и богослужебных произведений, приведшая в середине XV в. к изобретению Гутенберга, не ликвидировала необходимости в написании рукописи, научной переписке и личном общении между учеными. Появление в середине XVII в. и быстрое развитие научных журналов не привело к отмиранию научной книги. Ни кино, ни радио, ни телевидение не только не вытеснили книгу как средство научного познания и художественного отображения мира, но, напротив, сделали ее воздействие более эффективным, а распространение более широким.

Исходя из описанных общих свойств системы информационной коммуникации как устойчивой открытой системы целенаправленного действия, можно утверждать, что неоднократно высказывавшиеся опасения о замене библиотек информационными службами, традицион-

ных форм произведений печати микро- или электронными формами изданий, информации на естественном языке данными в машиночитаемой форме, информационно-поисковых систем экспертными и т. п. являются несостоятельными.

Новые процессы, методы и средства научной и массовой коммуникации, которые появляются непрерывно и призваны повышать эффективность интеллектуальной деятельности человека, будут лишь дополнять и обогащать систему коммуникации, не приводя к отмиранию существующих ее форм. Методы и средства коммуникации характеризуются большой стабильностью и консерватизмом в силу того, что они служат для материального воплощения общественного сознания, а их использование в некоторой сфере означает создание устойчивого канала связи между членами определенного сообщества. Этими же причинами вызваны, по всей вероятности, и трудности изменения письменностей, алфавитов, орфографии, систем транскрипции и транслитерации, специальных знаковых систем и т. п.

Широкую известность получили результаты исследований системы коммуникации, проводившихся в 1961–1994 гг. У. Гарвеем и Б. Гриффитом. Они изучили механизмы распространения информации по психологии в США. Было установлено, что из 30 тыс. психологов, работавших в США, около 40 % являлись сотрудниками научных учреждений, а остальные были заняты в промышленности, медицине и административном управлении. Более половины из них были членами Американской ассоциации психологов, которая публикует ведущие журналы по психологии и собирает на своих ежегодных конференциях до 10 тыс. участников (около 2 тыс. докладов). Американские психологи читают 1 тыс. различных названий журналов (каждый из которых публикует около 500 статей в год), 200–300 книг и научно-технические отчеты 400 учреждений. Задача заключалась в том, чтобы выявить каналы распространения информации в этом большом коллективе, оценить их значимость и временные характеристики. Некоторые итоги этого исследования представлены на рис. 4. Выяснилось, что работы, результаты которых опубликованы в журнальных статьях (отметка «0» на шкале времени), начинаются в среднем за 36 месяцев до опубликования. В середине этого периода они достигают той стадии, на которой составляется отчет о работе. С этого времени на протяжении трех с половиной лет по различным каналам непрерывно циркулируют в различной форме сведения об этих работах.

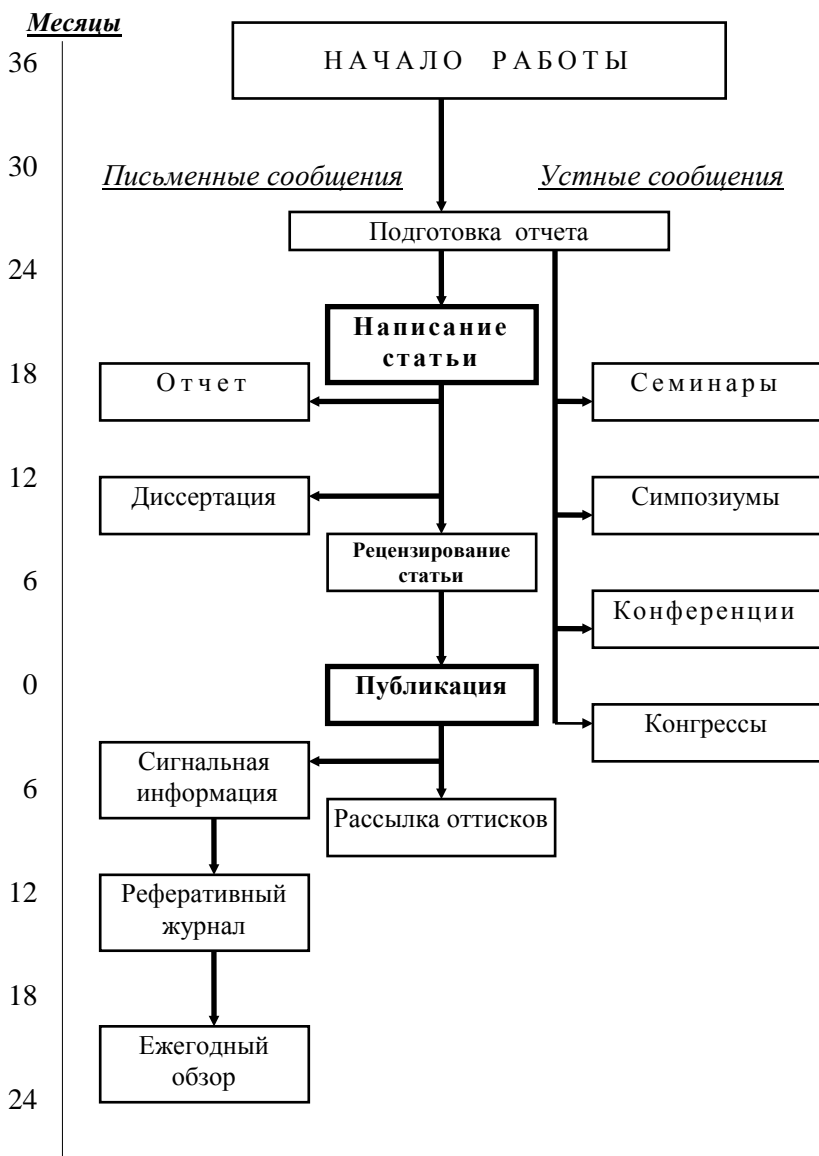


Рис. 4. Каналы коммуникации (по У. Гарвею и Б. Гриффиту)

На схеме видно, что первоначально эти сведения распространяются по неформальным (неофициальным) каналам (наиболее активно – за год до публикации), а с момента опубликования статьи почти целиком переходят в сферу информационных (вторичных) изданий. Исследования, о которых идет речь, показали, что имеется значительно больше неформальных каналов, по которым информация поступает к ограниченным группам специалистов, чем формальных каналов, предназначенных для широких кругов ученых. При этом по первым каналам передается значительно более свежая информация, чем по вторым. Практические шаги, которые могут быть предприняты на основе этих исследований, не потеряли своего значения до наших дней. И теперь существует необходимость ускорять и расширять для научных работников и специалистов доступ к материалам ограниченного распространения, в частности, к докладам на научных конгрессах, конференциях, симпозиумах.

Мы пришли к пониманию того, что система научной коммуникации является высокоупорядоченной и что действие разных ее элементов можно предсказывать очень точно. Она основана на взаимодействии людей, функционировании соответствующих общественных институтов и подчинении определенным социальным нормам. Ученый и информация – два важных компонента науки связаны не только прямой, но и обратной связью, так как ученые являются не только создателями, но и потребителями информации.

В течение долгого времени непосредственное общение между учеными, переписка между ними служили основным информационным каналом. Недаром условная форма письма одного ученого к другому много лет сохранялась в научных публикациях. Личные контакты между учеными и теперь продолжают оставаться одним из действенных способов передачи информации. Они дополняются взаимным обменом оттисками и препринтами статей в научных журналах, участием в научных конгрессах, конференциях, симпозиумах.

Научоведческие исследования последнего времени, в частности, те, которые опираются на метод библиографического сочетания документов с помощью ссылок, свидетельствуют о важности этих каналов распространения научной информации. Оказывается, фронт научных исследований распадется на не-

большие отрезки. Каждый из них соответствует одновременной работе над какой-либо проблемой небольшой группы в 100–150 ведущих ученых разных учреждений и стран, которые образуют как бы незримый коллектив, "невидимый колледж".

Библиотечная и информационная деятельность

На рис. 5. схематически показаны основные каналы распространения научной информации от создателей к потребителям. Внешние линии на схеме обозначают личные контакты ученых, их переписку, обмен отпечатками и препринтами, а также участие в научных конференциях, конгрессах, симпозиумах. Внутренние линии обозначают распространение неопубликованных документов, получающих апробацию и утверждение в научных учреждениях, в которых работают их составители. Затем они, как правило, доводятся до потребителей с помощью органов научной информации. Диссертации, защищаемые соискателями ученых степеней в научно-исследовательских институтах или высших учебных заведениях, становятся доступны потребителям в специальных библиотеках (один экземпляр хранится в библиотеке соответствующего института, а другой в национальной библиотеке страны, у нас – в Российской государственной библиотеке, копия во Всероссийском научно-техническом информационном центре).

Основную массу научных документов составляют публикации. Каналы их распространения занимают центральную часть схемы. Создатели этих документов, другими словами, авторы книг, журнальных статей или других готовых к опубликованию материалов передают их рукописи в книжные издательства или редакции научных журналов, которые в данном случае выступают в качестве апробирующих учреждений. Часто авторы передают в издательства рукописи своих работ после обсуждения и по рекомендации научных учреждений или высших учебных заведений, сотрудниками которыми они являются, но это обстоятельство на нашей схеме не отражено. Органы научной информации и специальные библиотеки служат важнейшими вспомогательными учреждениями науки.

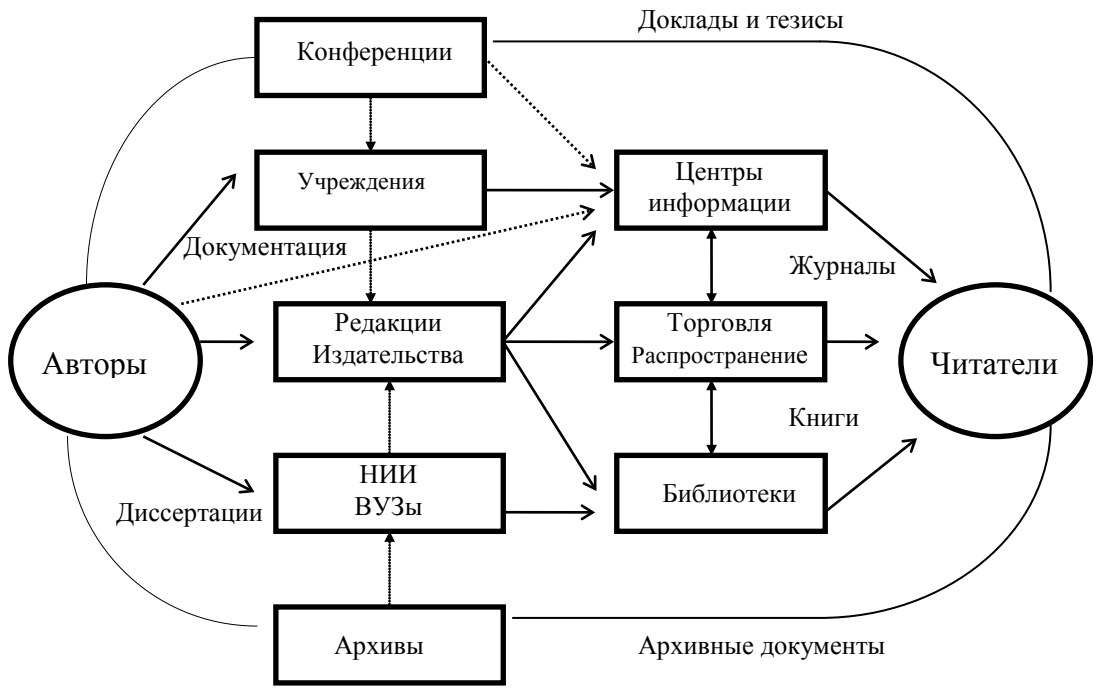


Рис. 5. Каналы распространения научной и другой специальной информации

Эти учреждения призваны выявлять и собирать научную информацию, анализировать и перерабатывать ее в формы, удобные для хранения и последующего поиска, обеспечивать ее хранение и распространение, включая выдачу по запросам. Однако осуществляя общие функции, они решают разные конкретные задачи, при помощи специалистов разного профиля и разными методами. Эти различия надо понимать, чтобы лучше использовать те возможности, которые предоставляют информационные службы и библиотеки.

Одно из таких различий, касающееся информационных работников и библиотекарей, было удачно сформулировано несколько десятилетий назад группой ведущих американских ученых в докладе президенту США: "Для перегруженного текущей работой специалиста гораздо важнее получить помощь от квалифицированного ученого-информатора, чем самому разбираться в огромном количестве литературы, полученной им из информационного центра. Квалифицированные ученые, работающие в информационном центре и вносящие свой вклад в науку, составляют ядро такого центра. Именно они превращают специализированный информационный центр в научно-исследовательское учреждение, а не в техническую библиотеку"¹.

Таким образом, информационный работник и библиотекарь – специалисты разного профиля. Первый из них это специалист в определенной области, участвующий в решении проблемы и обеспечивающий свой коллектив информацией. Библиотекарь же, который тоже может иметь образование и квалификацию в одной из отраслей знания, это специалист, владеющий методами библиографического контроля, ориентирующийся в огромном потоке книжно-журнальной продукции, понимающий психологию читателя и обладающий педагогическим мастерством руководства его чтением.

Конечные задачи информационной службы оповещать о новых фактах и идеях, отвечать на фактографические запросы, например, *что* известно по данному вопросу, *какими свойствами* обладает данный объект, *какие объекты* обладают данными свойствами и т. п.. Поскольку сегодня ответы на эти вопросы содержатся в научных документах и базах данных, информационные службы работают с ними и часто вынуждены вместо фактографических давать "документальные"

¹ Science, government, and information: A Report of The President's Science Advisory Committee. – Washington, D. C.: The White House, 1963. – P. 33.

ответы: какие сведения содержатся в тех или иных документах, в каких документах содержится необходимая информация. Следовательно, работа с научными документами – не основная деятельность информационных служб, а лишь средство для получения необходимой информации, ее интеллектуальной переработки и предоставления потребителю в удобной для использования форме.

Библиотека же – культурно-просветительное и научно-вспомогательное учреждение, организующее общественное пользование произведениями печати и письменности и призванное помогать читателю в выборе книг, руководить его чтением. Таким образом, библиотеки и информационные службы имеют разные конечные цели и задачи, хотя во многом работают с одними и теми же научными документами.

Еще одно различие касается людей, на которых направлена деятельность этих учреждений. И хотя это одни и те же люди – научные работники, специалисты-практики, преподаватели – называют их по-разному: информационные работники "потребителями информации", а библиотекари "читателями". Надо признать, что в данном случае существование разных терминов оправдано, так как они обозначают разные понятия. Потребитель информации обращается к информационным работникам за конкретными сведениями. Он вправе требовать от них точного и полного ответа на узкоспециальный запрос, поскольку они обязаны следить по документальным источникам за развитием обслуживаемых ими направлений. В библиотеке тот же самый человек выступает как читатель, которому наряду с произведениями, соответствующими его сегодняшним узким интересам, нужно дать прочитать книги и статьи, расширяющие его кругозор, повышающие его квалификацию, ориентирующие его в смежных областях, удовлетворяющие его общенаучные и широкие общественные интересы.

Наконец, если рассматривать работу этих двух типов учреждений по крупному, то нельзя не отметить, что они работают и в разных режимах. Хотя это различие не носит принципиального характера, но такова преимущественная практика и, по-видимому, она не случайна. Большинство информационных служб работает в режиме регулярного распространения научной информации. Он соответствует режиму поступательного развития науки, периодического выхода новых публикаций, появления непубликуемых научных документов.

Органы информации стремятся к максимально полному ознакомлению с новыми документами, осуществляют их аналитико-

синтетическую переработку, а затем равномерными порциями доводят их основное содержание до сведения потребителей в виде отдельных выпусков периодически издаваемых реферативных журналов, обзоров, экспресс-информации или библиографических указателей. Одновременно информационные службы стремятся обеспечить долговременное хранение документов с целью их последующего поиска и выдачи потребителям по запросу.

Библиотеки же преимущественно работают в режиме "запрос-ответ". Они накапливают произведения печати и письменности, раскрывают их содержание и организуют их хранение таким образом, чтобы выдавать затем по запросам читателей. Правда, почти каждая библиотека или группа библиотек ведет регулярное информирование читателей о своих новых поступлениях, но эта их деятельность по объему значительно уступает обслуживанию читателей по запросам.

Научно-информационная деятельность

Мы не случайно рассматриваем информационную деятельность на примере ее научной разновидности, поскольку в последние десятилетия ее организации придавалось большое значение в национальном и международном масштабе. В большинстве развитых стран существуют национальные или даже государственные системы научно-технической информации, а их правительства создают специальные органы для управления этими системами. По некоторым оценкам, в мире на организацию этой деятельности ежегодно затрачивается не менее 10 млрд. долларов.

Научно-информационная деятельность – это организационно оформленная разновидность научного труда, который выполняется в целях повышения собственно исследований и разработок и заключается в сборе, аналитико-синтетической переработке, хранении и поиске закрепленной в документах научной информации, а также в предоставлении этой информации потребителям в оптимальные сроки и в удобной для них форме.

Для уяснения этого определения важно отметить, что под аналитико-синтетической переработкой информации здесь понимается ее анализ и синтез, не доходящие до получения нового знания, что является задачей исследования. Дело в том, что при обсуждении вопроса о месте и роли научно-информационной деятельности высказываются две крайние точки зрения. Одна из них сводит эту деятельность к сово-

купности чисто технических операций, которые облегчают специалистам доступ к необходимой информации или же современной разновидности библиотечно-библиографического обслуживания. Этой позиции часто придерживаются ученые-теоретики и библиотекари.

Другая точка зрения, высказываемая некоторыми информатиками, инженерами и менеджерами, заключается в том, что, поскольку нет четкой границы между информационным и исследовательским анализом и синтезом, нет и черты, разделяющей информационную и исследовательскую деятельность. Такая позиция отражает растущую потребность специалистов в аналитических и прогнозных обзорах и не делает различий между информированием и получением нового знания.

На наш взгляд, обе точки зрения являются крайними и неверными. Они ведут к отрицанию самостоятельного существования информационной деятельности. Однако в настоящее время стало понятно, что проблема информационного кризиса не может быть решена как в рамках чистого исследования, так и одними библиотечно-библиографическими методами и средствами. Для существенного повышения эффективности любой творческой работы нужна разработка принципиально новых информационных методов и средств, что является практической задачей современной информатики.

Этапы и задачи коммуникации

В коммуникационном процессе обычно выделяют ряд этапов, которые соответствуют его основным функциям (или задачам). Они показаны на рис. 6. Прежде всего, необходимо с максимальной полнотой осуществить *сбор* выявленной заранее информации. Это значит, что из всей массы сведений должны быть отобраны и систематически пополняться все те и только те сведения, которые могут понадобиться для научной или практической работы по нескольким отраслям, одной отрасли, направлению или даже теме какому-либо специалисту, коллективу или учреждению.

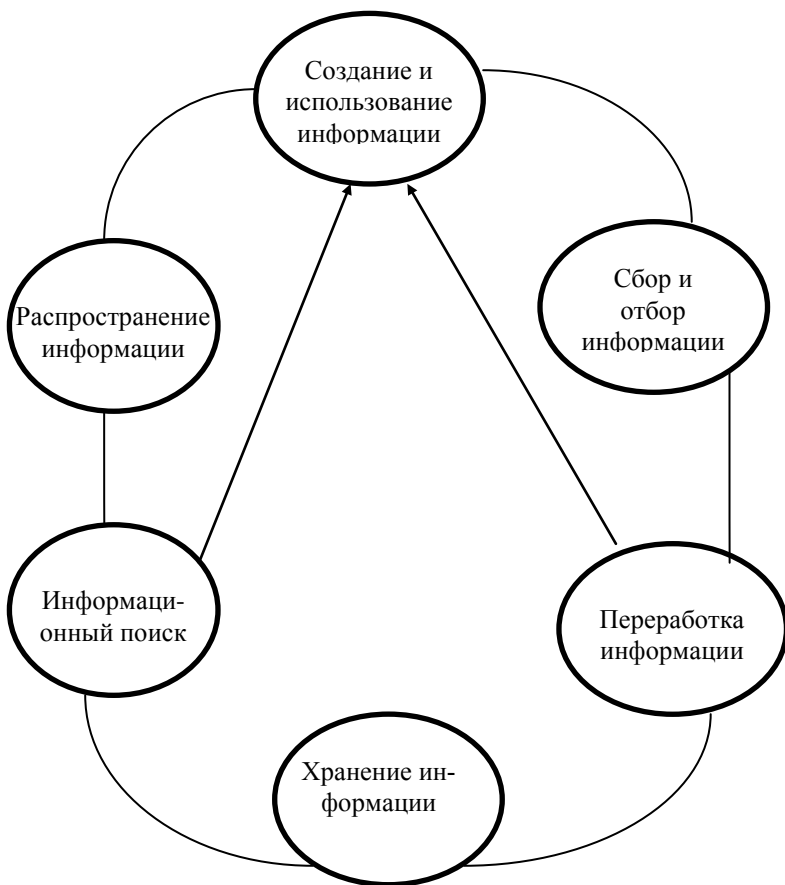


Рис. 6. Этапы (функции, задачи) коммуникации

Затем необходимо выполнить *аналитико-синтетическую переработку* информации. Каждую группу сведений, имеющихся в издании или документе, нужно проанализировать с точки зрения содержания, а сам документ – с точки зрения его формы.

Результаты этого анализа записываются кратко: они синтезируются в специально разработанных для этого условных обозначениях (при помощи слов естественного языка, букв и/или цифр, формул, таблиц, графиков). Переработанная таким образом информация становится удобной для ее последующего хранения и поиска, а также для использования в творческой работе.

Следующая задача – долговременное *хранение* информации. Ее нужно решить так, чтобы собранные сведения занимали как можно меньше места. Материальные носители, на которых они записаны, нужно сохранять как можно дольше, и организовать в стройную систему, позволяющую быстро и без затруднений находить необходимые сведения.

Информационный поиск – следующий этап и одна из важнейших проблем информационной деятельности. Она представляет собой ряд логических операций, обеспечивающих нахождение необходимой информации. Решение этой задачи с помощью соответствующего процесса позволяет в любой момент найти в массе собранных сведений те и только те из них, которые являются прямым ответом на вопрос потребителя.

Поскольку результаты всех предыдущих процессов (и особенно информационного поиска) могут представлять интерес одновременно для большого числа специалистов, возникает задача *распространения* информации. Издания и документы (или их части) с необходимыми сведениями должны быть репродуцированы в достаточном числе идентичных экземпляров и распределены в кратчайшие сроки между всеми заинтересованными потребителями.

Наконец, для информационной деятельности не безразлично, как используются ее результаты, хотя *использование* информации с целью *создания* новой информации, строго говоря, в задачи этой деятельности не входит. Однако в ее задачи входит пропаганда достижений науки и техники, наиболее рациональных методов работы с документами, оценка эффективности использования этих документов и изучение того влияния, которое информационная работа оказывает на науку, технику, производство и все народное хозяйство. Ясно, что вопросы технологии

создания и использования информации, входящие в сферу организации умственного труда, играют важнейшую роль для всей информационной деятельности.

Информационное обслуживание

Поскольку почти с каждого этапа коммуникационных процессов любая информация может быть доступна потребителю, а передача ему этой информации и составляет основной смысл самих этих процессов, вопросы информационного обслуживания находятся в центре внимания информационных работников. Основные виды информационного обслуживания можно различать по многим признакам:

- * от кого исходит инициатива при обслуживании – от информационной службы или от потребителя,
- * как осуществляется обслуживание – посредством опубликованных или непубликуемых документов,
- * на кого направлено обслуживание – на единичного или массового потребителя,
- * осуществляется ли обслуживание постоянно или лишь при необходимости,
- * как доводится информация до потребителя – непосредственно или в результате отклика на сигнальное оповещение,
- * с какой целью осуществляется обслуживание – для ознакомления или практического использования.

Классификация основных видов информационного обслуживания и строится по этим признакам, используемым как основания деления. При этом некоторые виды имеют специальные названия, а другие обозначаются описательно.

Как будет сказано дальше, эффективное информационное обслуживание возможно только на основе глубокого знания информационных потребностей ученых и специалистов. Вместе с тем, наиболее успешное изучение этих потребностей ведется в процессе самого обслуживания. Наиболее подходящим для этого видом обслуживания является избирательное распространение информации (ИРИ). Оно является одним из наиболее распространенных в информационной деятельности и исторически одним из первых видов обслуживания, которое стало осуществляться автоматически при помощи компьютеров.

Избирательное распространение информации (и сам термин "ИРИ") было введено сотрудником американской фирмы ИВМ

Х.-П. Луном, который в 1958 г. предложил соответствующую автоматизированную систему с использованием вычислительной машины. Он определил этот вид обслуживания следующим образом: "Избирательное распространение информации это такая служба в какой-либо организации, которая направляет новые элементы информации, получаемые из любого источника, в такие места, где вероятность их использования для текущей работы или заинтересованность в них наиболее высоки"¹.

При избирательном распространении информации потребитель как бы доверяет информационной службе поиск необходимых ему сведений в текущей научно-технической литературе. Он сообщает признаки, по которым следует отбирать ему документы: классификационные индексы, предметные заголовки, дескрипторы, фамилии авторов необходимых работ или цитируемых в них документов. Вся литература, поступающая в информационный орган, анализируется по этим признакам и в случае их совпадения с признаками, указанными потребителем, ему посылаются только его одного интересующие сведения. Система обратной связи с потребителем позволяет постоянно корректировать состав отбираемых материалов и тем самым улучшать индивидуальное информационное обслуживание.

Этот вид информационного обслуживания издавна практиковался библиотеками. Библиотекарь просматривал поступающие в библиотеку книги и журналы и сообщал читателям о публикациях, которые, по его мнению, могли представлять для них интерес. Здесь все зависело от памяти и доброй воли библиотекаря, а также от наличия у него свободного времени. Поэтому такая форма информационного обслуживания была нерегулярной, неформальной и весьма ограниченной по масштабам. В современной системе ИРИ поступающие документы индексируются (а иногда аннотируются или реферируются) специалистами, после чего они вводятся в компьютер, в котором содержатся введенные ранее "профили" абонентов. Если поисковые признаки документа совпадают с "профилем" абонента, ему посылается сигнальное извещение, по которому он может запросить документ или его копию, а затем обязан сообщить, насколько этот документ был ему полезен. Сведения из этих карт обратной связи также вводятся в машину, что позволяет оптимизировать работу системы.

¹ IBM Journal of Research and Development. – 1958. – Vol. 3, 4.

Перспективы развития

В заключение данной лекции хотелось бы высказать два соображения. Первое относится к необходимости различать информационную деятельность как общественную институцию от той информационной работы, которую неизбежно ведет каждый творческий работник и особенно ученые. Информационная работа ученого как неотъемлемая часть научных исследований является прерогативой научного сообщества, входит в систему науки и регулируется ее закономерностями. Научно-информационная деятельность выделилась в самостоятельную разновидность научного труда, принадлежит системе научных коммуникаций и подчиняется ее законам. Водораздел между этими системами крайне условен, трудно различим, но он существует и имеет принципиальное значение. В условиях новой информационной технологии этот водораздел будет смещаться в сторону большей интеллектуализации процессов научно-информационной деятельности, которая постепенно осваивает ряд процессов, прежде присущих только научному исследованию.

Второе соображение связано с перспективами информационных профессий. Новая информационная технология, дающая ученым и специалистам возможность, используя персональный компьютер, самим вести непосредственный диалог с информационной системой, не только расширяет их возможности в самостоятельной работе с информацией, но и в значительной степени освобождает их от посредничества информационных работников. Во всяком случае, это посредничество все больше ограничивается техническими и технологическими аспектами. Появились публикации¹, в которых говорится, что успехи информатики могут привести к отмиранию информационной профессии, к ухудшению перспектив развития научно-информационной деятельности как социального института. Одна из библиотечных статей на эту тему заканчивалась словами: "2001 год может стать великим годом для информатики, но тяжелым годом для информационных специалистов".

¹ News and Library World. – 1981. – Vol. 82, N 967. – P. 6.

Человек в процессе коммуникации

Потребители информации

Термин "потребитель информации" является не очень удачным переводом английского термина *information user*, который по отношению к информационным системам переводится как "пользователь", что немногим лучше, поскольку ассоциируется с архаичным словом "пользовать" в смысле "лечить". Потребителем информации считается лицо или коллектив, получающее или использующее информацию в научной или практической работе.

Мы уже говорили о том, что "потребление" информации в корне отличается от потребления товаров материального характера. И хотя результаты исследований и разработок, например, могут являться объектом купли-продажи, а их использование ведет к получению прибыли, выражения типа "производство знаний", "потребление информации" являются метафорами, затемняющими суть дела.

Информация, как вы уже поняли, сама по себе не является товаром и, удовлетворяя определенные потребности, не исчезает, как это происходит при потреблении любого товара. Больше того, увеличение числа источников информации не понижает, а повышает ее потребительскую стоимость, так как увеличивает общественную эффективность ее использования, тогда как расширение производства любого товара приводит к обратному результату.

Тем не менее, термин "потребитель информации" укоренился в информатике, широко используется и дал много производных терминов, в частности, термин "информационные потребности". Он применяется для обозначения наиболее интенсивно изучавшегося и наименее понятного явления, хотя именно его мы признаем определяющим фактором развития информационной деятельности.

Теперь мы понимаем, что любая творческая деятельность в информационном аспекте может рассматриваться как сбор, анализ, хра-

нение, поиск и использование информации. Различны лишь цели и результаты ее использования: ученый создает новую научную информацию, инженер – машину или технологию, врач – методы лечения, преподаватель – методы обучения и воспитания, руководитель принимает решения и т. д. Все они являются потребителями информации, хотя понимание этого пришло не сразу.

Исторически первыми потребителями информации были признаны создатели научной информации – ученые. На протяжении многих веков они пользовались при создании новой информации не только одними наблюдениями и размышлениями, они черпали сведения также из книг и журналов своих личных, а впоследствии и общественных библиотек. В среде ученых возникли и первые информационные издания – реферативные журналы. При этом наше представление о времени первого их появления постепенно отодвигается в глубь веков

В начале нашего века появился крупнейший отраслевой журнал по химии *Chemical Abstracts* (США, 1907), но столетием раньше стал выходить его предшественник в Германии *Pharmaceutisches Central-Blatt* (1830). Как установил М. Бониц (Германия), его первым издателем был известный немецкий физик, психолог, философ-идеалист и писатель-сатирик Г. Т. Фехнер, а самым ранним реферативным журналом можно считать *Aufrichtige und Unpartheyische Gedancken, Über die Journale, Extracte und Monats-Schriften, Worinnen Dieselben extrahiret, wann es nutzlich suppliret und wo es nöthig emendiret werden* («Искренние и внепартийные мысли о журналах, экстрактах и ежемесячниках, которые извлекаются и по необходимости дополняются и исправляются», 1714). В России рефераты также публиковались в первой половине XVIII в.: "Краткое описание комментариев имп. Академии наук" (1728) и "Содержание ученых рассуждений" (1750–1759) – и были тесно связаны с деятельностью М. В. Ломоносова.

Другая крупная категория потребителей информации – инженеры – была идентифицирована с появлением в конце прошлого века в Европе так называемых специальных библиотек и информационных бюро, которые возникли в связи с быстрым ростом техники и промышленности и необходимостью информационного обеспечения специалистов. В нашей стране первые крупные информационные органы появились именно в промышленности.

Как потребители информации инженеры и вообще специалисты-практики во всех сферах народного хозяйства отличаются от на-

учных работников не только значительно большей численностью, но и совершенно иным подходом к отбору и использованию информации. Научные работники стремятся найти закономерность в наблюдаемых явлениях, понять их природу, привести структуру современных знаний в соответствие с научной парадигмой, в рамках которой они работают. Им нужна информация по определенной проблеме с большой полнотой и широким охватом. Они стремятся самостоятельно осуществлять ее поиск, основываясь на ассоциациях, которые иногда трудно объяснить. Ученые в рамках научного сообщества сами устанавливают методы и нормы верификации создаваемых ими знаний и сроки выполнения этапов научных исследований.

Специалисты-практики решают конкретные задачи, сформулированные заказчиками, в сроки, установленные народнохозяйственными планами. Они призваны проектировать предприятия, машины, аппараты, обеспечивать их эффективное функционирование, разрабатывать и внедрять новые, более совершенные технологии, выводить новые сорта растений, искать и осуществлять новые методы лечения больных, обучать новые кадры и повышать их квалификацию. Им обычно нужна научно-техническая и производственная информация, ориентированная на ту задачу, которую они решают в данный момент и в определенные сроки.

Эта информация должна быть комплексной, относящейся к разным научным дисциплинам и отраслям знания. Она должна содержать конкретные сведения о методах и средствах выполнения названных задач. У практиков нет времени на исчерпывающий просмотр литературы, им в большей степени, чем ученым, нужно помочь отобрать информацию, обработать ее и представить в удобном для использования виде.

В 60-е годы информационные работники осознали необходимость выделения такой категории потребителей информации, как руководители. Появилась задача информационного обеспечения управления в таких социальных системах, как наука, техника, промышленность, экономика, торговля, здравоохранение. Эта задача оказалась крайне необычной.

Во-первых, при ее решении работники информационной службы впервые столкнулись с ситуацией, в которой научно-техническая информация является хотя и важным, но далеко не главным компонентом в структуре информационного обеспечения. Даже в управлении наукой

и научным учреждением наряду с научной требуются экономическая, социально-политическая, идеологическая, административная и другие виды информации. В еще большей степени это относится к иным сферам управления.

Во-вторых, та информация о достижениях науки и техники, которая необходима руководителям, значительно отличается по своему характеру от научно-технической информации, используемой учеными и инженерами. Это информация о вариантах решений и возможных последствиях их принятия. Ее содержат не статьи и книги, а обзоры, прогнозы, таблицы, графики. При информационном обеспечении организационного управления информационные службы, пожалуй, впервые встретились с необходимостью полной переработки информации. Исходя из рассмотренных нами особенностей дальнейшего развития науки и техники, подобная необходимость в будущем возрастет и будет охватывать все новые категории людей, которые прежде не осознавались в качестве потребителей информации, да и не были ими в привычном понимании.

На сегодняшний день в качестве потребителей научно-технической информации рассматриваются следующие три категории, выделенные в соответствии с характером деятельности:

- а) ученые в области фундаментальных и прикладных наук,
- б) специалисты-практики, занимающиеся разработками и/или практической деятельностью в различных областях техники, промышленности, сельского хозяйства, медицины, образования и т. п.,
- в) руководители (управляющие, менеджеры), плановики и другие лица, принимающие решения и координирующие различные виды деятельности на местном, национальном и международном уровнях частного или государственного секторов экономики.

В последние годы информационные службы достигли определенных успехов в обеспечении информацией специфических профессий. Получили развитие информационные системы для практических врачей конкретных специальностей (хирургов, педиатров, дантистов), для преподавателей и студентов высших учебных заведений определенных профилей. Разрабатываются системы информационного обслуживания для исследователей и работников различных сфер культуры, связанных с использованием информации в области общественных и гуманитарных наук.

Прогноз изменений в количественном составе потребителей информации нельзя ограничивать рассмотрением только научных работников или даже специалистов, занятых в науке и технике. Научно-технический прогресс ведет к интеллектуализации трудовой деятельности. Этот процесс характерен для всех стран, но его интенсивность зависит от развитости экономики и народного хозяйства страны. Активными потребителями информации становятся не только специалисты с высшим и специальным средним образованием, но также мастера и рабочие на производстве, работники сельского хозяйства, торговля, связи, планирования и учета, средний медицинский персонал.

Из всех подлежащих учету статистических совокупностей наиболее близка к такой трактовке потребителей информации численность работников, занятых преимущественно умственным трудом в народном хозяйстве. Анализ данных за последние десятилетия показывает, что рост потребителей информации в нашей стране происходит экспоненциально, удваиваясь примерно каждые 20 лет.

Эгалитарность информационного обслуживания

До сих пор мы рассматривали потребителей информации как элемент таких общих социальных систем, как наука, культура, производство, стремясь охарактеризовать качественный и отчасти количественный аспекты этого явления в целом. Но реальный потребитель информации предстает перед нами как конкретный абонент информационной службы, и это накладывает на наши рассуждения существенные ограничения. Далекое не все потенциальные потребители информации являются абонентами информационных систем. Статистика показывает, что в течение года в органы информации и научно-технические библиотеки обращается хотя бы один раз лишь половина потенциальных потребителей информации, причем не каждый из них может быть признан абонентом информационной системы.

Здесь действует общий для всех информационных процессов логарифмический закон, по которому лишь незначительная часть потребителей информации интенсивно использует услуги информационных органов. Вероятно, половина всех усилий этих органов затрачивается на обслуживание такого числа их абонентов, которое составляет корень квадратный из их общей численности.

Мы уже отмечали, потребители информации не являются однородной массой людей, они заняты различными видами творческой дея-

тельности, каждая из которых требует специфического информационного обслуживания. Потребители информации в рамках одной определенной деятельности, находятся на разных уровнях иерархии, выполняют различную по характеру работу с далеко не одинаковой значимостью для конечного результата. Их информационные потребности и навыки в получении и использовании информации зависят от их служебного положения, образования, знания языков, возраста и от многих других, часто не поддающихся строгому учету факторов.

Информационная служба достаточно дорога, и по мере усложнения услуг и внедрения в информационные системы современной техники стоимость этих услуг и систем возрастает. В связи с этим все более актуальным становится повышение экономической эффективности информационной деятельности. С самого начала ее интуитивно основывали на традиционной библиотечной этике бесплатного и демократического обслуживания с предоставлением всем абонентам равных возможностей. Однако с течением времени возникли сомнения в правильности такого подхода. Если абоненты информационных систем в разной степени влияют на прогресс в науке, технике, промышленном производстве, то не следует ли дифференцировать и уровень их информационного обслуживания?

Среди подходов к ответу на этот вопрос нельзя не заметить стремления информационных работников оценивать эффективность своей деятельности не только по ее собственным параметрам, но и по результату, достигаемому потребителем, получившим от системы необходимую ему информацию. Один из видных представителей отраслевой информационной службы М. Л. Колчинский давно предлагал определять эффект совершенствования информационной службы относительным сокращением времени, которое потребитель тратит на получение этой информации. Ясно, что общественная стоимость этого времени зависит от значимости результатов труда, выполняемого той или иной категорией потребителей.

Поскольку информационная служба переходит на новую информационную технологию, необходимо уметь определять эффективность производимых изменений. Понятно, что такие изменения происходят постепенно, так как невозможно перевести сразу всех абонентов информационных систем на тот уровень обеспеченности, который в данный момент признается общественно необходимым. В свое время научные сотрудники ВИНТИ М. В. Арапов и Ю. А. Шрейдер предла-

гали решить эту проблему путем ранжирования потребителей по значимости их роли в народном хозяйстве страны. При всех трудностях такого ранжирования руководители, работники директивных и плановых органов, аппарата управления народным хозяйством, ведущие специалисты учреждений, организаций и предприятий имели бы приоритет перед рядовыми научными работниками и специалистами-практиками.

В данном случае эффект совершенствовании информационной службы можно измерять объемом предотвращаемых потерь, происходящих от недостаточной информированности определенной части потребителей на общественно необходимом уровне. При таком подходе эффективность совершенствования оказывается "нелинейной", т. е. потери сокращаются как бы непропорционально затраченным средствам и усилиям. Например, обеспечение современными средствами поиска информации в диалоговом режиме только одной десятой части потребителей, играющих наиболее важную роль в создании национального дохода, может сократить предотвращаемые потери вдвое.

Подобную практику мы и наблюдаем в настоящее время. Возможности информационных систем, обеспечивающих непосредственный доступ к машиночитаемым базам данных, предоставляются в первую очередь руководителям и ведущим специалистам. Подразумевается, что подобное положение имеет временный характер, и по мере оснащения информационных органов современными техническими средствами доступ к ним получают все потребители информации. Но ведь предела совершенствованию нет, и нынешняя ситуация, представляющаяся нам временной, может оказаться постоянной. Возникает вопрос, не утверждаем ли мы де-факто *элитарность* информационного обслуживания, не пытаемся ли косвенно обосновать новую информационную этику, в которой потребители не должны располагать равными возможностями получения информации и доступа к ней? Если это так, то необходимо подумать о последствиях, к которым это может повести. Лучше всего это видно на примере науки.

В течение многих десятилетий было распространено мнение, будто научное сообщество организовано таким образом, что каждый его член, независимо от ранга в науке, может вносить небольшой вклад в общую сокровищницу знаний, а на долю выдающихся его представителей выпадает успех крупных открытий и научных переворотов.

Эту гипотезу обычно связывают с известным высказыванием испанского философа Х. Ортеги-и-Гассета: "Необходимо еще раз подчеркнуть этот чрезвычайный, но неопровержимый факт: экспериментальная наука развивалась в значительной мере благодаря усилиям людей на редкость посредственных, и даже менее чем посредственных. Иными словами, современная наука, этот исток и символ нашей нынешней цивилизации, находит место для интеллектуально заурядных людей и предоставляет им возможность с успехом трудиться в ней. Таким образом, большинство ученых способствует общему прогрессу науки, оставаясь замкнутыми в узких клетках своих лабораторий подобно пчеле в ячейке улья, либо белке в колесе"¹.

До последнего времени почти все пишущие о науке явно или неявно придерживались этого мнения. На нем основано "общественное знание" Дж. Займана, "нормальная наука" Т. Куна, "коммуникативное существо науки" У. Гарвея. Однако все чаще справедливость этого общего мнения подвергается сомнению. Оно было поколеблено впечатляющими исследованиями Р. Мертона, показавшего, что в науке действует принцип, провозглашенный еще в Евангелии от Матфея ("Кто имеет, тому дано будет и приумножится, а кто не имеет, у того отнимется и то, что имеет"). Об этом же говорится в работах Д. Прайса, раскрывших социальный механизм цитирования научных публикаций, и С. и Д. Коулов, обнаруживших тесную связь между интенсивностью цитирования и реальным вкладом ученого в науку.

В ходе этих и других примыкающих к ним социологических исследований, проводившихся главным образом на материале цитирования работ физиков, было доказано, что большинство публикуемых в них результатов почти никогда не используется наиболее видными учеными, делающими самые значительные открытия.

Следующие из этого выводы хорошо сформулировал Дж. Коул: "Из всех социальных институтов наука оказывается в числе наиболее стратифицированных. Наблюдается существенный разрыв между небольшой "элитной" группировкой и основной массой ученых. В то же время, несмотря на эту стратификацию, в науке постоянно поддерживается система ценностей, зиждущаяся на принципах эгалитаризма и неуклонном подчеркивании коллективистского духа"².

¹ Цит. по кн.: Коммуникация в современной науке. – М.: Прогресс, 1975. – С. 390.

² Коул Дж. Схемы интеллектуального влияния в научных исследованиях // Коммуникация в современной науке. – М.: Прогресс, 1976. – С. 422–423.

Когда в 60-е годы были впервые сделаны эти выводы, они были поняты буквально. Из признания большой роли неформальных коммуникаций и якобы бесполезности значительной части публикаций вытекало требование сократить объем и число публикуемых статей, заменить их неформальными "группами обмена" информацией, увеличить количество симпозиумов и конференций. Все конкретные мероприятия в этом направлении не имели успеха, а работы Дж. Займана, У. Гарвея и Б. Гриффита восстановили равновесие формальной и неформальной коммуникации в науке.

Принцип *эгалитаризма* в науке, да и в других общественных институтах выполняет ряд важных системообразующих функций. Признание этого обстоятельства содержалось в известном высказывании Н. Винера о том, что хотя 95 % оригинальных научных работ принадлежит менее чем 5 % ученых, они никогда не были бы написаны, если бы остальные 95 % занятых в науке не содействовали бы созданию общего ее высокого уровня.

Нельзя не заметить, что большинство крупных зарубежных социологов науки – в прошлом физики. Они неявно считают высшим видом научной деятельности эксперимент, а единственным достойным научным результатом – открытие. В этом невольно проявляется физический снобизм, связанный с великими достижениями физики недавнего времени. Они не отдают себе полного отчета в том, что не меньшее значение для "нормальной" науки имеет деятельность по "расфасовке" и "переупаковке" научных знаний, все то, что ими презрительно называется компиляцией и популяризацией и несправедливо трактуется как нетворческая, а порой и как бесполезная работа.

Элитаризм, возведенный в принцип информационного обслуживания, угрожает серьезной опасностью для развития науки, так как он не учитывает внутренней иерархии информирования в научном обществе. Дело в том, что как раз руководители и ведущие специалисты живут в густой информационной среде, создаваемой их сотрудниками. На всех уровнях официальной научной иерархии, и особенно на ее низших ступенях, среди аспирантов, молодых научных работников образуется тип ученого-информатора, который все читает, следит за новыми публикациями, сообщает о них на рабочих семинарах. В процессе обсуждений эта информация подвергается оценке, интерпретации, затем по неформальным каналам она в сублимированном виде поднимается на верхние уровни научной иерархии.

Игнорирование этих внутренних механизмов информирования в научном сообществе ведет к резкому снижению эффективности информационного обеспечения научных исследований и разработок. Поэтому вынужденную ситуацию, при которой все потенциальные потребители информации не могут сразу быть переведены на современные методы информационного обслуживания, основанные на новой технологии, не следует возводить в принцип.

Единственно правильным принципом информационного обеспечения специалистов в науке и народном хозяйстве, культуре и образовании, во всех сферах человеческой деятельности должен оставаться эгалитаризм, признание равных для всех потребителей информации возможностей ее получения и доступа к ней. Это, разумеется, не исключает дифференцированного подхода к различным категориям потребителей информации при их обслуживании, учета особенностей их нужд, запросов и привычек.

Какие пути возможны для реализации этого принципа в сложившихся условиях? Прежде всего, планируя, создавая и внедряя новые виды и средства информационного обеспечения, нельзя произвольно отменять прежние. Пока возможности ретроспективного поиска или избирательного распространения информации в режиме диалога с автоматизированной системой при непосредственном доступе к машинночитаемым базам данных существуют лишь для ограниченного числа пользователей, не следует сокращать выпуск информационных изданий в традиционной форме. Необходимо продолжать издание бюллетеней сигнальной информации, кумулятивных библиографических указателей, реферативных журналов и вспомогательных указателей к ним, указателей цитированной литературы и всех других видов текущего оповещения о новых достижениях.

Одним словом, и в данном случае сохраняет силу общий информационный закон, по которому введение новых элементов в систему научной коммуникации не отменяет и не заменяет старых, традиционных средств. Эволюция в научной коммуникации осуществляется путем перераспределения функций между видами и средствами информационного обеспечения и, конечно, их объемами и тиражами.

Информационные потребности в развитии

Наиболее общие типы информационных потребностей были выявлены давно и, несмотря на интенсивные исследования последних

десятилетий в этой области, их формулировка остается почти неизменной. Творческие работники обращаются к источникам информации главным образом в трех случаях:

а) при текущем ознакомлении с новыми сведениями, как по их узкой специальности, так и в смежных областях,

б) при повседневной работе, когда необходимо получить конкретные фактические сведения о числах, количествах, методах, средствах, концепциях, гипотезах и т. п.,

в) в начале работы над новой проблемой или темой и при ее завершении и литературном оформлении, когда необходимо путем ретроспективного поиска с возможной полнотой выявить и учесть все опубликованные и неопубликованные источники по данной теме.

Такое постоянство свидетельствует о стабильности мотивов обращения специалистов к источникам информации, что не исключает изменчивости самих информационных потребностей. Исследования информационных потребностей различных категорий потребителей закономерно сопровождают разработку любой крупной информационной системы.

В прошедшее десятилетие существенно продвинулось социологическое понимание информационных потребностей, тесно связанное с трактовкой существа и перспектив развития информационной деятельности. По этому поводу отечественными и зарубежными исследователями высказано несколько заслуживающих внимание точек зрения. Одна из них связана с традиционным пониманием неопределимости информационных потребностей ученого, исследующего фундаментальные проблемы науки. Известно, что полное и адекватное формулирование ученым своей потребности в информации равносильно решению им соответствующей проблемы. Он ищет необходимую информацию не только по теме, проблеме, узкому ее аспекту, но и по ассоциации, по аналогии, по методу исследования и по многим другим трудно формулируемым аспектам.

Новая информационная технология, позволяющая пользователю вести активный диалог с информационной системой, не только расширяет его возможности при поиске информации, но и освобождает его от посредничества информационных работников в процессе этого поиска. Во всяком случае, это посредничество все больше ограничивается технологическими аспектами и становится все менее содержательным. Как вы уже знаете, некоторые специалисты за рубе-

жом, доводя этот ход рассуждений до крайности, приходят к парадоксальному выводу, что успехи информатики могут привести к отмиранию самой информационной профессии. Существует и противоположная точка зрения, по которой создание диалоговых информационных систем упрочивает положение информаторов, давая им в руки мощное средство управления информационной средой.

Дело в том, что первоначальная концепция научно-информационной деятельности, исходившая из понимания информации только как некоего продукта, одного из ресурсов науки, давно признана ограниченной. Более плодотворной оказалась концепция информационной среды философа Ю. А. Шрейдера, связанная с изучением социального контекста научной коммуникации: "Ученый активно взаимодействует со своей информационной средой так же, как и с любой другой средой – природной, социальной и т. п. Он, в частности, динамично перестраивает свое поведение в этой среде – варьирует представления о собственных информационных потребностях, обращая то к одним, то к другим ее элементам в разной последовательности и сочетаниях, получая информацию в самых разнообразных ракурсах, проверяя незапрограммированные связи между фактами и, главное, самостоятельно управляя характеристиками получаемой информации, полнотой, темпом поступления, степенью детализации, тематикой и т. п."¹.

Концепция информационной среды ориентирует информационную деятельность на то, чтобы сделать эту среду более благоприятной, дружественной, комфортной для ученого. Эта концепция связана с тенденцией науки к гуманитаризации, с повышением интереса ученых к мотивации творчества, анализу ценностных и познавательных установок. В конечном счете, этот интерес к личности ученого, создающего информацию, ведет к усилению роли социологии науки, психологии научного творчества, к выявлению места науки в общей культуре. Личностный интерес к субъективной компоненте вовсе не противостоит традиционной тенденции науки к объективизации и математизации, а дополняет ее. С точки зрения информационных потребностей, это означает, что прежние вопросы, "что" и "как" сделано в науке, дополняются вопросами, "кем" и "почему" это сделано.

¹ Шрейдер Ю. А. Гуманитаризация знания и управление информационной средой // Вестник АН СССР. – 1978. – № 9. – С. 88.

Своеобразное представление о развитии информационных потребностей разработчиков новой техники выдвинул Э. С. Бернштейн, многие годы работавший в области информационного обеспечения машиностроения. Он поделил всех потребителей информации на "наблюдателей" и "аналитиков". Первые, по его мнению, делают изобретения в рамках готового набора "ситуаций", т. е. по сложившимся канонам, по традиции. Вторые открывают новые традиции, создавая свои каноны, критически оценивая существующие концепции решения стоящей перед ними задачи. Именно они являются двигателями технического прогресса и нуждаются в нетривиальном информационном обслуживании, которое состоит в выяснении "пробелов" в существующем знании. Истинная информационная потребность этого типа специалистов состоит не только в том, чтобы получить информацию об уже имеющихся достижениях, но в том, чтобы выяснить, *где* существуют пробелы в знаниях, *что* пока еще неизвестно.

О несовершенстве социологических методов исследования, таких как анкетирование и интервьюирование, давно известно, хотя они успешно применяются во многих сферах и постепенно совершенствуются. При изучении информационных потребностей эти общие методы дополняются анализом обратной связи в системах информационного обслуживания. Однако исследовать подлинные потребности специалистов в информации очень трудно, так как они осознают их на уровне имеющихся возможностей. Более того, новые виды информационного обслуживания с трудом пробивают себе дорогу, наталкиваясь на устойчивые привычки и навыки потребителей в получении информации. По этой же причине и новая информационная технология внедряется значительно медленнее, чем это позволяют имеющиеся технические средства.

В самом деле, потребности специалистов в научной, технической, экономической, политической, конъюнктурной и другой информации значительно превышают те, которые они осознают как разумные для предъявления информационной службе. Да и обслуживаемые ныне контингенты намного меньше действительно нуждающихся в информации на общественно необходимом уровне. Именно это противоречие между реальными информационными потребностями и возможностью их удовлетворения в рамках информационной деятельности является определяющим фактором ее развития.

Литература как источник информации

Основные понятия, эволюция и типология

Строго говоря, источником любой информации является человек или группа людей. Ответственность за достоверность тех или иных сведений могут брать на себя учреждения или организации, которые в этом случае также служат как бы источниками информации. Но в профессиональной информационной деятельности информация циркулирует в виде документов, и именно они считаются источниками информации.

Под *документом* понимается совокупность логически завершенных сведений и материального носителя, на котором они записаны, с неизменным указанием кем, где и когда документ был создан. Понятие документа является альтернативным и обобщающим по отношению к понятию произведения письменности и печати, часто употребляемому в гуманитарных науках.

Произведением письменности считается результат целенаправленной познавательной деятельности (факты, идеи, образы), имеющий определенную логическую взаимосвязь частей, завершенность в целом и изложенный в письменном виде. *Произведение печати*, кроме того, получает обязательную редакционную апробацию и имеет точный библиографический адрес, состоящий из стандартного набора выходных сведений (автор, заглавие, город, издательство и год публикации или название публикующего органа).

Поскольку способы хранения и передачи информации играют большую роль в развитии коммуникации, да и цивилизации в целом, интересно проследить эволюцию этих способов.

С развитием каждой цивилизации по мере усовершенствования ее языка и письменности вырабатывались и определенные типы документов, как по форме, так и по содержанию. Самые древние из дошедших до нас документов – глинописные плитки Месопотамии – да-

тируются четвертым тысячелетием до н. э. От шумерской культуры того времени и вавилоно-ассирийской, расцвет которой приходится на второе тысячелетие до н. э., сохранились сочинения в области астрономии, географии, истории, права, торговли. Наряду с небольшими глиняными плитками высотой в 2,5 см изготавливались плитки высотой до 40 см, содержащие до 400 строк в 4 колонки с двух сторон. Основным недостатком этих глиняных книг, в течение тысячелетий обслуживавших культурные и научные потребности человека, была их громоздкость и недостаточная емкость. Отдельные произведения занимали до 10 плиток. Собрания, насчитывавшие десятки тысяч плиток (а до нас дошли остатки подобных библиотек), требовали огромных помещений.

Папирусный свиток – более компактная форма документа, позволявшая накапливать большие собрания произведений письменности. Начиная с третьего тысячелетия до н. э. в Египте изготавливались именно такие книги. Текст на папирусном свитке располагался перпендикулярно его длине колонками от 25 до 45 строк. Хрупкость и недолговечность папируса обусловила незначительное число дошедших до нас образцов древнеегипетских документов (древнейший из них восходит к XVIII в. до н. э.). Это, главным образом, ритуальные "книги мертвых", извлекаемые археологами из пирамид и других захоронений.

У греков и римлян на протяжении долгого времени сведения (особенно научного характера) распространялись устным путем. Известно, что лишь после Аристотеля для этих целей стали широко применяться рукописи. До середины нашего века мы не располагали ни одним оригиналом времен античных классиков, часто столетия отделяют последнюю сохранившуюся копию от возможной даты написания текста. Исключение составляет лишь найденные во время второй мировой войны кумранские рукописи ("рукописи Мертвого моря"), датируемые первым веком н. э. Сведения о греческих и римских папирусах почерпнуты из малоазиатских и позднеегипетских образцов. Средняя длина свитка не превышала 10 м, ширина – 30 см.

Пергамент, изготовлявшийся из телячьих шкур, известен как материал для письма с III в. до н. э. Он позволял писать с двух сторон, был более долговечен, чем папирус, и обусловил переход к современной блочной форме книги – кодексу. Кодексы из папируса изготовля-

лись еще в первые века н. э., но к V в. были вытеснены пергаментным кодексом.

Кодекс – более емкая форма книги, чем свиток, он удобнее для записи больших текстов и для наведения справок, Пергаментные кодексы вплоть до XII в. были единственной формой книги в Европе. Основным их недостатком, препятствовавшим широкому распространению письменных документов, была дороговизна. Для одного экземпляра пергаментной книги требовались шкуры целого стада. На смену пергаменту пришла бумага, изобретенная в 105 г. в Китае. На Ближнем Востоке ее начали изготавливать с 751 г., а в XII в. через арабских завоевателей Испании она проникла в Европу.

Бумажная книга, значительно более дешевая, чем пергаментная, стала широко использоваться в научных и образовательных целях. Опыт нескольких тысячелетий развития письменных документов показывает, что их форма менялась главным образом под влиянием потребностей общества: документы становились все более емкими, удобными для использования и дешевыми. Именно эта тенденция развития средств, служивших для закрепления информации, привела к появлению бумажного книжного блока, который до нашего времени оставался основной материальной формой документа.

В XV в. стала повсеместно ощущаться потребность в новом способе изготовления документов. Рукописная книга перестала удовлетворять культурные и научные запросы общества по двум причинам. Во-первых, она изготавливалась слишком долго и требовала значительных затрат труда. Во-вторых, переписка текстов от руки не давала возможности получить большое число экземпляров идентичного содержания, так как копии одного и того же текста отличались друг от друга из-за искажений, вносимых переписчиками.

Именно к этому времени относится историческое изобретение в 1448 г. немецким ремесленником И. Гутенбергом книгопечатания подвижными литерами. Технические средства, которые легли в основу книгопечатания, – граверная и литейная техника и винодельческий пресс, преобразованный Гутенбергом в печатный станок, – были известны еще в античные времена. Однако лишь настоятельная потребность общества в быстром и точном механическом воспроизведении текстов вызвала к жизни это изобретение, совершенству которого мы не перестаем удивляться и основными принципами которого продолжаем пользоваться для размножения документов.

Мы знаем, что теперь появились технические средства, значительно повысившие наши возможности хранения больших массивов информации с быстрым доступом к любой единице этой информации. Речь идет о компактных оптических дисках, используемых в качестве внешней памяти компьютера (CD-ROM – Compact Disc Read Only Memory). В этой области прогресс происходит так быстро, что рискованно приводить какие-либо точные данные. Но объем одного диска измеряется тысячами мегабайт, т. е. миллионами страниц текста, время записи и считывания одной страницы не превышает 1–2 сек., а в "библиотеке" из 64 дисков поиск и выдача информации по запросу занимает 10–15 сек. Выведенные на экран тексты и изображения можно изменять по мере необходимости. Эти новые средства начинают широко использоваться, и мы еще будем о них говорить.

В широком смысле документами иногда считают не только надписи, рукописи и печатные издания, но и произведения искусства, нумизматические памятники, музейные экспонаты минерального, ботанического, зоологического или антропологического характера. П. Отле считал документом любой материальный объект, который фиксирует или подтверждает какие-либо знания и может быть включен в определенное собрание.

Различные виды научных документов возникали в разное время и на протяжении последних столетий и даже десятилетий претерпевают значительную эволюцию. Книга существует уже несколько тысячелетий, описание изобретений – 500 лет, научный журнал – немногим менее 350 лет, а журнальная статья в ее настоящем виде – 100–150 лет. Типология документов также существенно меняется. До последнего времени наиболее важным считалось деление научных документов на опубликованные и неопубликуемые. Еще несколько десятилетий назад идеи и факты признавались введенными в оборот только после их опубликования, означавшего широкое распространение и официальную регистрацию документов, в которых они содержались.

Для информационной деятельности это разграничение менее существенно, так как, во-первых, в неопубликованных документах содержится много ценной информации, опережающей сведения, появляющиеся в публикациях, а, во-вторых, новые средства репродуцирования делают это разграничение очень условным. Такие научные документы, считающиеся обычно неопубликуемыми, как отчеты, диссер-

тации, переводы, часто распространяются в сотнях и даже тысячах экземпляров.

Информатика выдвинула на первый план деление документов на первичные и вторичные. Деление это также очень условно и приблизительно, поскольку оно главным образом относится к самой информации, а не к документам, в которых она содержится. Считается, что в первичных документах отражаются непосредственные результаты познания, а во вторичных – результаты аналитико-синтетической переработки информации, содержащейся в первичных документах. Однако исторически сложившаяся система научных документов такова, что многие из них содержат одновременно и результаты научных исследований и переработку прежних сведений, содержащихся в ранее опубликованных документах. Примером могут служить и статьи в научных журналах, и монографии, и учебники, и особенно – справочная литература.

Тем не менее, деление это удобно, так как позволяет характеризовать различные потоки документов в информационной деятельности. Мы придерживаемся его в информатике, считая *первичными те документы и издания, в которых преимущественно содержатся новые сведения или новое осмысление известных идей и фактов, а вторичными те документы и издания, в которых содержатся сведения о первичных документах*. С учетом сделанных оговорок к первичным документам и изданиям можно отнести большинство книг (за исключением справочников), журналы, газеты и сериальные издания, описания изобретений, стандарты, отчеты, диссертации, переводы, а ко вторичным – справочники и энциклопедии, обзоры, реферативные журналы, библиотечные каталоги, библиографические указатели и картотеки.

Основные виды первичных документов и изданий охарактеризованы в книговедческих курсах, а вторичным будет посвящена в данном курсе лекция об информационных изданиях и услугах. Здесь хотелось бы высказать некоторые соображения лишь о журналах, так как на их примере дальше будут анализироваться закономерности роста, старения и распределения научных публикаций. *Журналом мы будем называть периодическое (сериальное) издание, регулярно публикуемое в течение одного года выпусками, одинаково оформленными и содержащими статьи или иные материалы научно-технического или общественно-политического содержания, а также произведения художе-*

ственной литературы. В журналах содержится новейшая информация, освещаются последние достижения науки и техники. Журналы появились почти триста пятьдесят лет назад: точной датой этого события считается 5 января 1665 г., когда был опубликован первый номер французского еженедельника "Журнал ученых", который дал название этому виду периодических изданий.

Основным назначением этого журнала, предопределившим характер научных журналов на 150 лет вперед, стало оповещение о новых книгах по всем отраслям науки, литературы и искусства, однако с особым вниманием к естественным наукам и технике. Для раскрытия содержания книг в то время широко пользовались прямыми заимствованиями и цитатами из текста. Вначале научная хроника играла в журнале второстепенную роль, постепенно все больше места в нем стали занимать сообщения об экспериментах в области естественных наук и вновь открытых явлениях природы. Оригинальные статьи в течение всего XVIII в. публиковались в журналах редко. Обычно они принадлежали крупным ученым и имели традиционно-условную форму писем одного ученого к другому: так было принято сообщать о научных открытиях в предшествующие эпохи.

Начиная с XIX в. журнал становится основным источником научной информации. В нашем столетии установилась исключительно важная практика упоминания в каждой журнальной статье всех научных работ, которые использовались при ее написании. Статьи в научных журналах являются в настоящее время основным источником научной информации, они прочно занимают первое место среди всех других документов. Обследование библиографических запросов нескольких тысяч ученых и инженеров показало, что до 70 % всех используемых ими источников составляют журнальные статьи.

Однако рост числа журналов, их недостаточная профилированность, быстрое старение опубликованных в них материалов привели к тому, что уже с 30-х годов журнал как источник информации стал подвергаться критике ученых. Они выдвинули множество проектов замены научных журналов другими средствами распространения знаний. В их основе лежит предложение вместо издания журналов депонировать разрозненные статьи в специальных отраслевых центрах и отражать их в реферативных журналах.

Один из первых проектов такого рода был выдвинут отечественными учеными, делегатами Международного геологического конгресс-

са в 1933 г. В этом же году англичанин У. Дэвис сделал аналогичное предложение, которое легло в основу известного "плана Бернала", опубликованного в 1939 г. в книге Дж. Бернала "Социальная функция науки". Этот план был предметом обсуждения в 1948 г. на Конференции по научной информации, созванной английским Королевским обществом, а в 1958 г. на Международной конференции по научной информации в Вашингтоне. Дж. Бернал предложил отказаться и от самой статьи как формы сообщения результатов научного исследования, поскольку она не обеспечивает их быстрого и адекватного отражения. Рациональные моменты в этих его идеях были воплощены при создании системы депонирования неопубликованных научно-технических документов. Однако полностью план Бернала вряд ли когда-либо будет реализован, так как он не учитывает многих закономерностей системы научных публикаций.

Закономерности роста и старения

С развитием информатики наступил новый этап в изучении научных публикаций, поскольку основное внимание стали уделять закономерностям, характеризующим внутреннюю связь изданий с развитием науки, количественные зависимости между числом публикаций и показателями роста науки. Закономерности эти связаны со структурой и свойствами научной информации, но проявляются несколько иначе и могут быть достаточно точно измерены.

Выяснилось, что для числа авторов, публикующих определенное количество работ в течение своей жизни, числа журналов, ежегодно публикующих определенное количество статей, числа публикаций, содержащих определенное количество ссылок на другие публикации, существует общая закономерность распределения. "Они следуют тому же типу распределения, который характеризует соотношение миллионеров и бедняков в условиях высокоразвитой капиталистической экономики: огромная доля богатств находится в руках узкого круга сверхбогачей, а небольшой остаток – в руках несметного множества мелких производителей. Является ли точная форма распределения логарифмической, экспоненциальной, описывается ли она законом Ципфа или обратной квадратной функцией это предмет особого рассмотрения в каждом отдельном случае"¹.

¹ Прайс Д. Регулярные закономерности в организации науки // Органон. – 1965. – № 2. – С. 246–247.

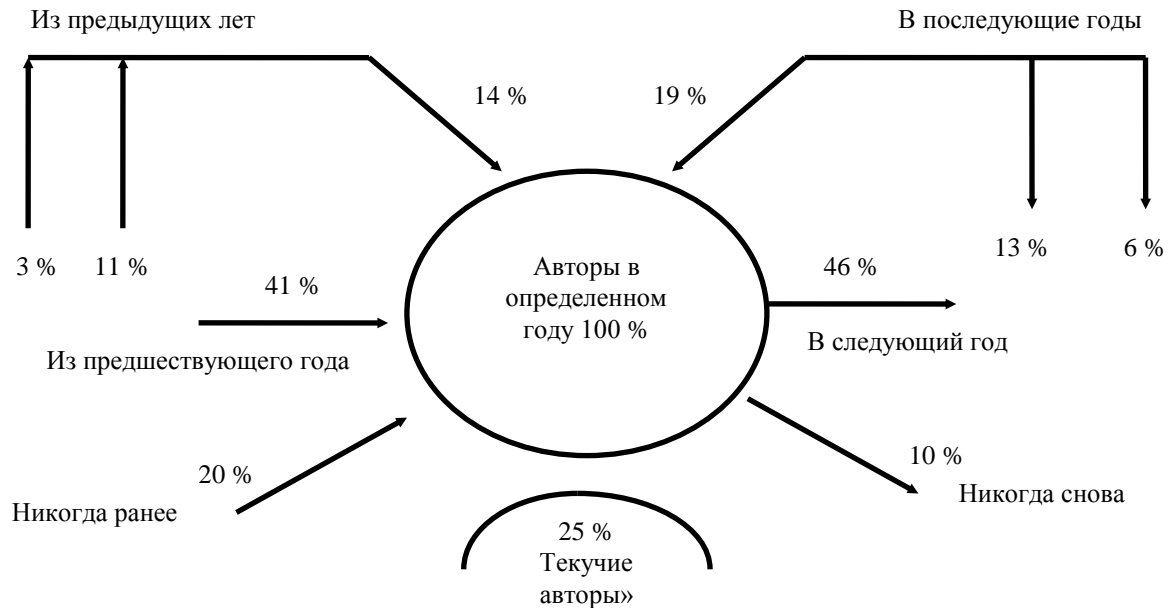


Рис. 7. Постоянство авторства (по Д. Прайсу)

Это означает, что большинство авторов за всю жизнь публикует лишь одну или две статьи, тогда как небольшая группа авторов отличается плодовитостью, публикуя по несколько десятков или даже сотен работ. По большей части прекращают выходить в свет периодические издания, успевшие выпустить несколько годовых комплектов, тогда как небольшое число давно выходящих изданий публикует львиную долю всех статей. Примерно половина опубликованной литературы обязана такому числу авторов или журналов, которое составляет квадратный корень общего их количества. "Короче, если, например, в мире выходит 30 тыс. журналов или в какой-либо стране имеется 1 млн научных работников, то лишь небольшое ядро в 175 журналов и 1 тыс. ученых ответственно за половину всей литературы по количеству и, вероятно, за 70–80 % по важности содержания"¹.

Эти положения Д. Прайс иллюстрирует схемой, которую он назвал "подходом страхового агента" к проблеме авторства научных статей. На рис. 7 круг обозначает 100 % авторов в какой-либо области или стране, опубликовавших статьи в определенном году. Левая сторона схемы показывает распределение авторов этих статей по их публикуемости в предыдущие годы (год назад, два-три года назад и ранее не публиковавшихся). Правая сторона показывает то же распределение в последующие годы. Под кругом изображено движение "постоянных" авторов, т. е. публикующихся на протяжении ряда лет (и не учитывавшихся в приведенном выше распределении). Обобщенный смысл этой схемы заключается в том, что для увеличения числа постоянных авторов на одну единицу необходимо появление примерно четырех новых авторов. Один из них заменяет постоянного автора, переставшего публиковать свои труды, два других выбывают по причине "детской смертности". Это те, кто приходят в систему научных публикаций и уходят из нее в течение одного года. И только четвертый остается на более или менее длительный срок печатающимся автором².

Рост литературы выражается в непрерывном увеличении числа вновь появляющихся изданий и публикаций. Широко распространенное представление об экспоненциальном росте основных видов литературы справедливо лишь для ее суммарного количества, причем без

¹ Там же. – С. 248.

² Прайс Д., Гюрсей Г. Текучесть и постоянство в авторстве научных публикаций // Международный форум по информации и документации. – 1976. – № 2. – С. 18–27.

учета старения. Реальной моделью такого представления могут служить крупные научные библиотеки, комплекующие литературу по широкому профилю и выполняющие функции архивного хранения литературы. Но если нас интересует ежегодный прирост новой литературы, то приходится быть более осторожными в оценках.

Статистика мирового книжного рынка за последние полстолетия показывает, что число ежегодно выпускаемых на рынок книг увеличивается в арифметической прогрессии, а именно, на 15–20 тыс. названий. В 1955 г. по данным ЮНЕСКО опубликовано 269 тыс. названий книг, в 1960 г. – 332 тыс., в 1970 г. – 521 тыс., в 1980 г. – 715 тыс., в 1990 г. – 842 тыс., в 2000 г. – 1,25 млн. Справедливость требует заметить, что это далеко не все выходящие книги, а только те, которые поступают в продажу. Если бы было возможно учесть заказные, бесплатные, ведомственные, учебно-методические и другие издания ограниченного распространения, то приведенные цифры можно было бы удвоить. Следует также иметь в виду, что научные книги (т. е. содержащие научную информацию в нашем понимании) составляют 20–25 % от общего их числа.

Подсчет числа журналов значительно сложнее, так как они в отличие от книг все время находятся в процессе изменений. Журналы возникают, прекращаются, сливаются, дробятся, меняют название, издателей, периодичность и т. д. По мнению специалистов, только по естественным, точным и прикладным наукам ежедневно три новых журнала возникает, а один перестает выходить.

Наиболее достоверные сведения о числе выходящих журналов можно получить из "Международной библиографии периодических изданий", выходящей под именем Констанции Ульрик. В последнем издании этого справочника зарегистрировано 164400 названий журналов, из которых около 50 % падает на издания по общественным и гуманитарным наукам, 40 % составляют журналы по точным, естественным и прикладным наукам и лишь 10 % – литературно-художественные и общественно-политические журналы. Некоторое представление о темпах роста числа журналов можно получить, сравнивая объемы различных изданий библиографии К. Ульрик: 13-е изд. (1969–70) – 40 тыс. названий, 15-е изд. (1971–72) – 55 тыс., 17-е изд. (1973–74) – 60 тыс., 19-е изд. (1980) – 62 тыс., 21-е изд. (1982) – 63 тыс., 24-е изд. (1988) – 100 тыс., 30-е изд. (1995) – 120 тыс., 39-е изд. (2001) – 164 тыс.

Старение публикаций заключается в том, что они с увеличением своего "возраста" теряют ценность как источники информации и все меньше используются специалистами. Степень этого использования можно устанавливать при помощи учета цитирования. В данном случае стареет не сама информация, а содержащие ее публикации, поскольку (как мы выяснили) в свежих работах эта информация может быть "упакована" более плотно вместе с новой.

Для измерения скорости старения публикаций американские ученые Р. Бартон и Р. Кеблер предложили в 1960 г. меру, названную "периодом полужизни" публикаций по аналогии с показателем скорости распада радиоактивных веществ. *Период полужизни публикаций* – это время, в течение которого была опубликована половина всей используемой в настоящее время литературы по какой – либо отрасли или предмету. Например, если этот период равен 5, то это значит, что 50 % всех процитированных в текущем году по данному предмету работ не старше пяти лет. Ниже приводятся данные разных авторов о периодах полужизни публикаций в различных отраслях науки:

Биомедицина 3,0	Химия 8,1
Металлургия 4,6	Ботаника 10,0
Хим. технология 4,8	Математика 10,5
Социология 5,0	Геология 11,8
Машиностроение 5,2	География 16,0
Физиология 7,2	

Достоверность приведенных цифр зависит от величины выборки цитирования, от типа и характера публикаций, поэтому даже в пределах одной науки данные разных авторов могут существенно расходиться. Но дело не только в этом. В 70-е и 80-е годы проблема старения литературы подверглась интенсивным исследованиям, в результате которых ее понимание стало сильно отличаться от концепции "периода полужизни". Начало этому пересмотру положил М. Лайн, который ввел в расчет характеристики старения литературы темпы ее экспоненциального роста. Сущность того, что произошло в трактовке старения, как всегда ярко выразил Д. Прайс: "В течение нескольких лет после публикации спрашиваемость статьи или ее относительная цитируемость уменьшается крайне медленно (по параболе, если считать по логарифмам прошедших лет). Даже через столетие возможность цитирования уменьшается только на порядок. Большинство ссылок падает на работы последних лет потому, что этих работ большинство, и очень

сомнительно, чтобы это вызывалось эффектом немедленности, связанным с быстрым старением..."¹.

Этой проблеме до сих пор и у нас и в ряде зарубежных стран посвящается много серьезных работ, которые убеждают в том, что частота использования определенной совокупности литературы одного года издания меняется очень медленно. Использование публикаций, определяемое по их цитированию или на основе запросов читателей, отражает не только старение литературы, но и ее рост.

Для теоретиков информатики и историков науки важно учитывать старение литературы в чистом виде, для информаторов и библиотечарей период полужизни служит важным практическим показателем и продолжает широко использоваться. Следует также иметь в виду, что цитируются далеко не все научные публикации. Половина статей в определенной области в текущем году, как правило, не упоминается, а еще 40 % цитируется лишь один раз (обычно самим автором). Таким образом, активный исследовательский фронт, т. е. число работ, цитируемых более одного раза в году, на порядок меньше корпуса опубликованной литературы.

Закон рассеяния

Еще одним важным свойством научных публикаций является их *рассеяние*. Закон рассеяния научных статей в журналах был открыт в 1934 г. С. Бредфордом, который в 1948 г. дал ему следующую формулировку²: *"Если научные журналы расположить в порядке убывания числа помещенных в них статей по какому-либо заданному предмету, то в полученном списке можно выделить ядро журналов, посвященных непосредственно этому предмету, и несколько групп или зон, каждая из которых содержит столько же статей, что и ядро. Тогда числа журналов в ядре и в последующих зонах будут относиться как $1 : a : a^2$ ".*

В соответствии с этим законом журналы по продуктивности можно сгруппировать так, чтобы они как бы образовали три зоны. Включенные в каждую такую зону журналы содержали бы одну треть публикаций по данному предмету, помещенных во всех этих журналах.

¹ Price D. General theory of bibliometrical and other cumulative processes // Journal of the American Society for Information Science. – 1976. – Vol. 29, N 5 – P. 292.

² Bradford S. Documentation. – 2nd ed. – London: Lockwood, 1953. – P. 154.

Первая, ядерная зона содержит публикации из небольшого числа самых продуктивных журналов – Т1. Вторая зона содержит публикации из большего числа журналов средней продуктивности – Т2, а третья зона – из еще большего числа журналов с низкой продуктивностью – Т3. Тогда в соответствии с рассматриваемым законом

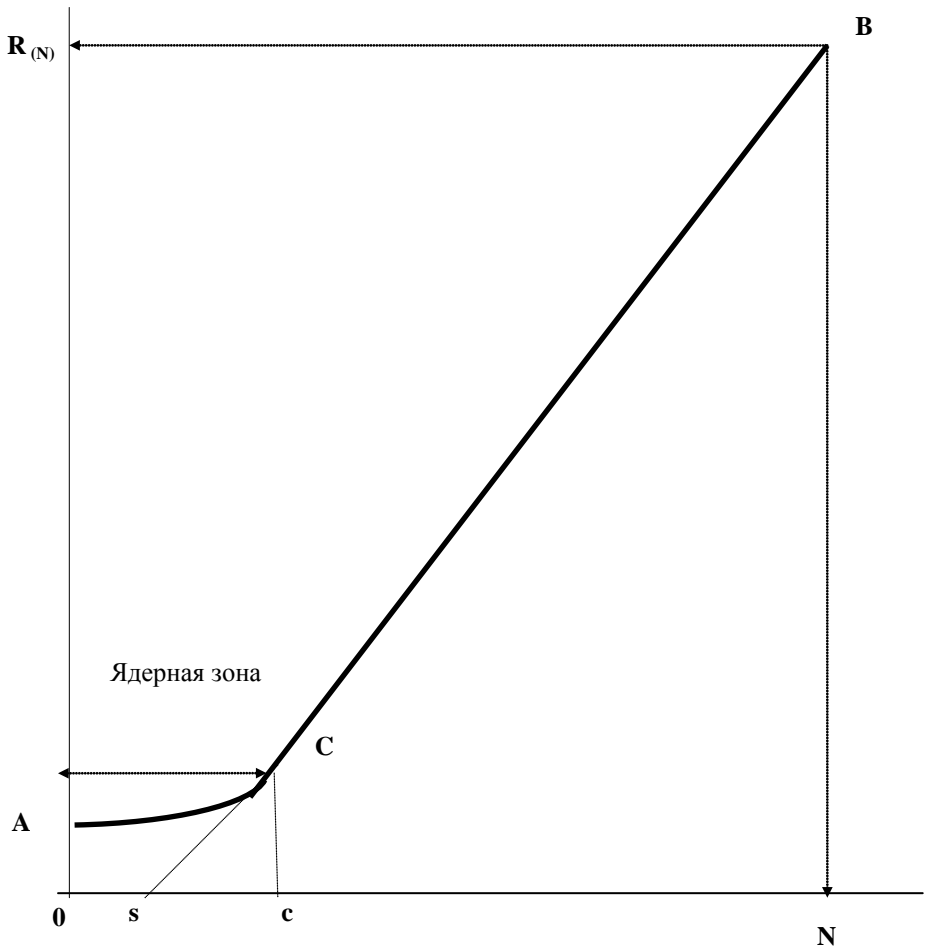
$$T1 : T2 : T3 = 1 : a : a^2,$$

где a является коэффициентом рассеяния, т. е. величиной для данного предмета и времени постоянной.

Для 248 журналов по электрохимии, которые были проанализированы С. Бредфордом, численное значение a составляло примерно 5. В ядерной зоне содержалось 8 самых продуктивных журналов, во второй зоне $8 \times 5 = 40$ журналов средней продуктивности и в третьей зоне $8 \times 25 = 200$ журналов. В каждой из этих трех зон содержалось по 220 релевантных публикаций, общее число которых составляло 660. Кривая рассеяния публикаций в соответствии с законом Бредфорда представлена на рис. 8. Другими словами, если совокупность всех публикаций по какому-либо вопросу принять за целое, то в специальных журналах данного профиля (число которых невелико) помещается лишь одна треть этих публикаций.

Вторая треть статей по данному вопросу оказывается опубликованной в значительно большем числе тематически родственных (смежных) журналов. Последняя треть этих публикаций рассеяна в огромном числе периодических изданий, в которых появление статей данной тематики трудно предвидеть, так как эти издания имеют широкий профиль или общенаучный характер.

За годы, прошедшие со времени открытия этого закона, проведены сотни исследований с целью проверить его истинность и найти для него строгое математическое выражение. Они показали, что закон этот выполняется только при определенных условиях, когда предмет или тема четко сформулированы, учитываются все релевантные документы в полном перечне изданий и строго ограничено время выхода этих изданий.



По горизонтали: Ранг периодического издания ($\log_e n$)
 По вертикали: Число релевантных публикаций, $R(n)$

Рис. 8. Рассеяние публикаций (по С. Бредфорду)

Последнее условие имеет особый смысл, так как закон этот характеризует рассеяние в определенный момент. Он является частным случаем более общего распределения, описываемого законом Ципфа. Дж. Ципф установил, что если к достаточно большому тексту составить список всех встретившихся в нем слов и расположить их в порядке убывания частоты встречаемости в данном тексте, то для любого слова произведение его порядкового номера (ранга) на эту частоту есть постоянная величина, имеющая одинаковое численное значение в данном тексте. Этому закону подчиняется распределение не только слов во всех языках мира, но и других явлений социального характера: ученых по числу опубликованных ими работ, городов по численности населения, людей по размерам дохода и даже биологических родов по числу входящих в них видов.

Следует отметить, что многие попытки объяснить механизм осуществления закона Бредфорда оказались неубедительными из-за того, что их авторы распространяли этот закон на процессы, происходящие во времени (т. е. в диахронии), тогда как он справедлив только для определенного временного среза (т. е. в синхронии). Закон Бредфорда отражает одно из свойств открытой социальной системы, какой и является научная литература по предмету, а именно стабильность ее иерархической структуры.

Некоторое отличие этого закона от ципфовского распределения объясняется спецификой периодических изданий как формы квантования научной литературы. Эти издания обладают большой инерционностью: изменения в их профилях и номенклатуре происходят значительно медленнее, чем в содержании статей, которые непосредственно отражают все процессы в науке и технике.

Закон рассеяния публикаций имеет большое практическое значение. Из него следует, что охват всех публикаций по какой-либо отрасли или предмету не может быть обеспечен, если ограничиться просмотром лишь профильных журналов и журналов по родственной тематике – для этого приходится просматривать значительную часть научно-технических журналов. Этот закон учитывается при организации национальных информационных систем. Он позволяет решить ряд практических задач информационной деятельности:

* определять число журналов, которые обеспечивают тот или иной процент всех публикаций по какой-либо отрасли или предмету,

- * составлять списки журнальных публикаций по определенной теме с гарантированной степенью полноты,
- * оценивать полноту библиографических списков журнальных публикаций,
- * комплектовать журнальные фонды при фиксированных ассигнованиях на подписку,
- * вычислять длину полок, необходимых для хранения оптимального фонда журналов.

Рассмотренные нами закономерности далеко не исчерпывают всех достижений информатики в изучении средств информационной коммуникации. К ним следует также добавить и результаты исследований последних десятилетий в области социологии массовой коммуникации и информации: распределение типов и видов передач на радио и телевидении, организацию газетных банков данных и информационно-поисковых систем и т. п.

Перспективы развития

В последнее время наблюдается много признаков того, что система периодических изданий утрачивает значение основного средства распространения научной информации. Об этом свидетельствует быстрый рост удельного веса непубликуемой научной литературы – отчетов, докладов, обзоров, справок и т. п., а также все увеличивающаяся роль неформальных коммуникаций – всевозможных конференций и совещаний, выставок, посещений учеными лабораторий своих коллег, обмена препринтами и другими непубликуемыми материалами. Если на научно-информационную деятельность в США расходуется до 10 % правительственных ассигнований на науку, то более 10 % из них тратится на научно-технические совещания. Каждый день там происходит несколько конференций или симпозиумов. Совершенно ясно, что это значительно ускоряет распространение информации и стало возможным в результате достижений в области транспорта и связи.

Но это вовсе не означает, что журнальные публикации вообще утрачивают свое значение. По мнению Д. Прайса, "80 % ценности и функционального назначения статьи лежат вне области коммуникации. *Статья и коммуникация перекрывают друг друга лишь на 20 %*". Статья, а тем более монография становятся для ученого средством кристаллизации и формулирования своих идей, самовыражения и утверждения своей личности в науке, закрепления своего приоритета на ре-

зультаты исследования¹. Недаром суммарное число опубликованных работ служит общепринятым показателем продуктивности труда ученого. Кроме того, если рассматривать каждый документ как составную часть литературы, то его опубликование означает индивидуальный вклад в общечеловеческое знание и возможность его сохранения для последующих поколений, т. е. передачу не только в пространстве, но и во времени.

Мы уже говорили раньше о том, что наряду с источниками информации в традиционной литературной форме новая информационная технология выдвигает на передний план документы в электронной машиночитаемой форме. Именно это дало основание французскому социологу Ж. Андерла утверждать, что темпы роста информации будут возрастать, несмотря на ограничение выпуска журналов и книг. По его мнению, в 1955 г. ежегодный прирост не дублируемых единиц научной информации составил 0,5 млн, в 1970 г. он вырос до 2 млн, а к 1985 г. – до 12 млн.

Увеличение темпов роста первичной и вторичной информации он оценивал в 12–13 % в год. При этом он считал, что от 1/4 до 1/3 всей информации сосредоточено в автоматизированных системах и циркулирует в их сетях, а прирост этих электронных источников информации является наиболее интенсивным и достигает 40 % в год. Хотя этот прогноз был встречен скептически большинством специалистов, в действительности темп этого роста оказался еще более динамичным.

Следует отметить, что электронные источники информации отличаются от традиционных двумя существенными чертами: во-первых, более широким составом авторов, и, во-вторых, большими изобразительными возможностями. В отношении первого обстоятельства выдвигается следующая гипотеза: "В настоящее время установлено, что в распространении новой информации важную роль играют такие категории авторов, которые не являются собственно научными работниками. На первом месте здесь должны быть названы экономисты, демографы, юристы, социологи, психологи, географы, педагоги и другие представители общественных наук, которых обычно не учитывают в статистических исследованиях численности научных работников.

¹ Прайс Д. Тенденции в развитии научной коммуникации – прошлое, настоящее, будущее // Коммуникация в науке. – М.: Прогресс, 1972. – С. 96.

Вполне вероятно также, что значительная и без сомнения все возрастающая часть общего объема информации производится инженерами, техниками и другими специалистами-практиками, работающими в различных областях и на разных уровнях"¹.

Второе обстоятельство обсуждается в связи с *электронной книгой*. Эта книга интерактивна, с ней можно взаимодействовать как с диалоговой информационной системой, вносить в нее собственные изменения и комментарии, компоновать фрагменты текста в соответствии с заданной читателем логикой. Кроме того, она не ограничивается текстом и статичными иллюстрациями, в ней возможны видеосюжеты в цвете и со звуком в цифровой записи. К примеру, вы обращаетесь к электронной книге с записью энциклопедии на компакт-диске. Под заголовком "Бетховен" вы можете вывести на экран компьютера не только обычную статью о жизни и творчестве композитора, но и все его прижизненные портреты, фрагменты посвященных ему кинофильмов, партитуры и звукозаписи его произведений и другую необходимую информацию. Разумеется, реализация этих возможностей потребовала решения многих проблем. Некоторых из них мы коснемся в связи с информационной технологией в отдельной лекции, посвященной этой теме.

¹ Anderla J. Information in 1985. – Paris: OECR, 1973. – P. 25.

Информационные издания и услуги

Основные виды

Основными видами информационных изданий являются:

* *библиографические указатели (БУ), бюллетени сигнальной информации (СИ)*, в которых приводятся в систематизированном виде библиографические данные о новейших журнальных и других публикациях по определенной отрасли, проблеме или предмету;

* *реферативные журналы (РЖ)*, в которых приводятся в систематизированном виде рефераты и аннотации журнальных и других публикаций и непубликуемых документов по определенной отрасли, проблеме или предмету;

* *продолжающиеся обзорно-аналитические издания* по избранным отраслям, предметам и проблемам (типа «Advances in...», «Progress in...», "Итоги науки и техники" ВИНТИ), в которых дается обобщение важнейших достижений по публикациям и непубликуемым документам за год или за несколько лет.

Бюллетени СИ обычно издаются еженедельно, дважды в месяц или ежемесячно и снабжаются, по крайней мере, пономерными авторскими указателями. Наилучшие бюллетени СИ отражают материалы за четыре недели со времени их опубликования. РЖ обычно издаются раз в две недели или ежемесячно и снабжаются, по крайней мере, пономерными авторским и предметным указателями, а также соответствующими годовыми указателями. Наилучшие РЖ отражают 50% материалов за срок, не превышающий 60 календарных дней со времени их выхода в свет.

Все виды информационных изданий могут выпускаться как в традиционной печатной форме, так и на машиночитаемых носителях информации (магнитных лентах, дискетах, оптических компакт-дисках) и на микрофильмах. Особенно большой популярностью среди

ученых пользуются бюллетени СИ, в которых приводятся оглавления важнейших журналов по соответствующим отраслям (типа "Current Contents»). Обычно такого рода бюллетени издаются по важнейшим отраслям естественных, технических и общественных наук, сельскому хозяйству и медицине и снабжаются авторскими и предметными указателями.

Реферирование и библиографирование

Основой информационных изданий служат библиографические описания и рефераты. Методику их составления трудно излагать на лекциях, поэтому она изучается на практических занятиях. Здесь хотелось бы обсудить некоторые проблемы, связанные с эти и важными видами информационной деятельности. Они свидетельствуют о том, что это не просто рутинная техника, а сложные интеллектуальные процессы.

Даже самый беглый взгляд на историю реферирования, которая уходит корнями в глубокую древность, показывает, что этот вид информационной деятельности возник из потребности кратко изложить существо, основное смысловое содержание того, что мы теперь называем первичным документом. Его развитие шло от простого извлечения наиболее содержательных фрагментов текста к концентрированному изложению его определенных аспектов. К настоящему времени мы располагаем арсеналом специализированных методов как самого реферирования, включая использование формализованных схем (анкетное и избирательное реферирование), содержательного изложения текста (экстрагирование, перефразирование, интерпретацию), так и оценки рефератов. Продолжают совершенствоваться статистические, логические и лингвистические методы автоматизации реферирования.

Однако грандиозный и все возрастающий объем реферирования, исчисляемый миллионами документов в год, заставляет задуматься о целях реферирования и их возможной модификации в будущем. В соответствии с известным принципом оценивать перспективы развития как разрешение существующих противоречий, нужно эти противоречия выявить. Одним из них является многофункциональность реферата, в которой основные его функции – служить средством текущего оповещения о новых достижениях и одновременно средством ретроспективного поиска – выдвигают противоречивые требования.

При использовании реферата как средства текущего оповещения он составляется в расчете на узкий круг специалистов, занятых разработкой проблемы, которой посвящено основное содержание реферируемого документа. Состав извлекаемых из документов фактических сведений и логика изложения нацелены на решение конкретных текущих задач исследования, известных и понятных референту, который чаще всего сам принадлежит к числу активных исследователей данной проблемы.

При этом реферат является не просто индикатором необходимости обратиться к первичному документу. Он становится как бы консолидированной информацией, облегчающей принятие того или иного научного, технического или иного решения. Внимательный анализ деятельности специалиста меняет традиционный взгляд на реферат. Он выступает не как промежуточный этап на пути к первоисточнику, а наоборот, как документ, который ближе к конечному продукту информационного обеспечения, чем первоисточник.

Как средство ретроспективного поиска информации, реферат, напротив, не имеет определенного адреса и должен быть рассчитан на длительное использование и непредвиденно широкий круг специалистов, по преимуществу исследующих смежные проблемы. В этом случае для реферата важен охват всех, в том числе и побочных линий содержания реферируемого документа, а его ценность тем выше, чем более общее представление он дает не только о методах и результатах исследования или разработки, но и о позиции автора первичного документа, ходе его рассуждений и т.п.

Возникает явное противоречие, связанное с тем, что первая функция тяготеет к аспектному, анкетному, специализированному, формализованному реферированию, а вторая – к реферированию общенаучному, системному, понятийному, интерпретационному. Для первого нужен в качестве референта узкий специалист, а для второго специалист более широкого профиля и высокого уровня.

Другое важное противоречие связано с тем, что издание реферативных журналов становится частью интегральных автоматизированных систем НТИ, их побочным продуктом при генерировании машинночитаемых баз данных. Как мы убедились, прогресс в информационной технологии в последние десятилетия намного опережает совершенствование интеллектуальных процессов. Методы реферирования не являются исключением. В течение долгого времени они развивались

в расчете на логику человеческого восприятия. и до сих пор продукция реферативных служб целиком состоит из рефератов, предназначенных для традиционных реферативных журналов, т. е. исключительно для чтения их человеком (с их рубрикацией, расположением материала, дублированием рефератов в разных разделах и рубриках и т. п.).

Реферативный журнал и другие информационные издания на машиночитаемых носителях (не вполне правомерно называемые базами данных) должны быть рассчитаны, в первую очередь, как показывает и их название, на переработку электронными машинами, для которых не только не имеет смысла порядок расположения рефератов и их дублирование, но требуются и определенные правила составления. В настоящее время остро встал вопрос о необходимости автоматического извлечения из рефератов сведений для пополнения специализированных баз данных и автоматизированных интеллектуальных информационных систем (экспертных, диагностических и т. п.). Это предполагает разработку особых методов реферирования и представления данных в реферате, которые могут существенно отличаться от традиционных.

Вместе с тем реферативные журналы и в машиночитаемой форме сохраняют многие функции, которые пока не может взять на себя никакое другое издание и никакой другой вид информационного обслуживания. Они делают для исследователя обозримой не только ту область, которой он непосредственно занимается, но и смежные области, компенсируют рассеивание информации, способствуя интеграции и сохранению единства науки, служат средством косвенной оценки научного качества публикаций. Выполнение этих функций предполагает совершенствование традиционных методов реферирования и издания реферативных журналов.

Дальнейшее развитие реферирования и реферативных служб будет происходить в преодолении этих противоречий. Можно констатировать, что общенаучные функции реферативных журналов все больше превалируют над их оперативными и специальными функциями. Это может повести к разделению этих изданий на специализированные (машиночитаемые) и общенаучные (*человекочитаемые*), подготавливаемые в рамках единых реферативных центров по разной методике. Следует подчеркнуть, что изучение этого вопроса должно стать одной из "горячих" точек современной информатики.

Аналогичное положение наблюдается и в области библиографирования, где теория библиографического описания, в течение многих лет развивавшаяся под влиянием библиотечной каталогизации, пришла в противоречие с новой информационной технологией. Основные принципы книгоописания, сложившиеся к настоящему времени, нацелены на выбор сведений, которые содержат самые важные для идентификации данного документа признаки и определяют место его библиографического описания в алфавитном ряду других описаний.

Библиографическое описание в научно-информационной деятельности выполняет ряд важных функций, из которых две считались до сих пор основными – адресная и сигнальная. Функция адресности выполняется благодаря тому, что описание содержит необходимые и достаточные сведения для отыскания и отождествления определенных документов по их основным библиографическим признакам: фамилиям авторов и других лиц, принимавших участие в создании документа, наименованиям учреждений, ответственных за содержание и издание документа, первым словам заглавия.

При использовании традиционных библиотечно-библиографических средств информационного поиска – каталогов, картотек и указателей – обычно приходится ограничивать число авторов, имена которых служат характеристиками для поиска документов. Необходимость в таком ограничении отпадает, когда информационный поиск ведется при помощи компьютера. Это дает возможность отразить в описании и использовать в качестве признаков для поиска не только имена всех авторов, но и имена других лиц, участвовавших в создании документа – составителей, авторов отдельных глав или статей, редакторов, переводчиков и т. п.

Сигнальная функция описания заключается в том, что оно может использоваться в качестве средства оповещения о появлении документа, имеющего определенное научное содержание. Из всех элементов описания этой цели в наибольшей степени служат заглавия и подзаголовочные данные, хотя они и не всегда в достаточной мере раскрывают содержание документа. Новая информационная технология позволяет использовать значимые слова заглавия в качестве поисковых признаков, а также снимает жесткость требований к строгой после последовательности элементов описания. Однако все эти соображения, давно предвиденные в связи с использованием компьютеров, далеко не исчерпывают круга проблем, встающих перед информаци-

онными работниками в связи с необходимостью перехода на новую информационную технологию.

Новейшие исследования показывают, что сформулированные нами десятилетия назад основные функции описания сохраняются и поныне, независимо от предложений называть их по-другому: поисковую функцию – идентификационной, а сигнальную – информационной. Долгое время казалось, что книгоописание развивалось под влиянием только интеллектуальных потребностей социальной коммуникации. Лишь теперь пришло понимание того, как много для него значила ориентированность на определенные технические средства – карточные каталоги и указатели в форме списков. Именно под влиянием широкого внедрения компьютеров стали по-новому подходить к обоснованию функций библиографического описания, определению факторов, влияющих на его структуру.

Прежде всего, функция идентификации до настоящего времени выполнявшаяся всеми элементами описания, теперь во многих процессах удовлетворяется международным стандартным номером для книг и периодических изданий (ISBN и ISSN). Это в значительной мере снимает идентификационную нагрузку с других элементов описания. Новшество состоит в том, что заголовок описания, в течение целого столетия составлявший главную заботу профессиональных каталогизаторов, исключается из состава библиографического описания и переносится во вспомогательные элементы библиографической записи, т.е. попадает в один разряд с классификационными индексами, предметными заголовками, дескрипторами и ключевыми словами.

Однако в практическом применении этого положения остается некоторая неясность. Все хорошо, пока мы не выходим за пределы применения описания в каталогах, указателях и других поисковых системах. Но ведь мы забываем обширную область литературного, так сказать, использования описания: упоминание произведений печати в тексте и подстрочных примечаниях других произведений. Здесь роль заголовка как привычного элемента в начале описания и как связующего элемента "бытового" описания с "профессиональным" (т.е. для связи ссылок в тексте и описаний в каталоге) настолько велика, что полное исчезновение заголовка из описания в будущем представляется сомнительным.

Есть еще один, на первый взгляд мелкий, но на самом деле важный аспект библиографического описания, связанный с автоматизи-

зацией библиографирования. Это пунктуация библиографического описания, выраженная ныне сложной системой так называемых разделительных знаков. Традиционно роль этих знаков, отделяющих один элемент библиографического описания от другого, исполняли обычные знаки препинания: точка, запятая, тире, скобки. На первых этапах автоматизации появилась необходимость обозначать каждое поле различающимися метками, которые заменили обычные разделительные знаки. Эти метки (разделительные знаки) были взяты из англо-американских правил каталогизации, которые восходят к первым печатным каталогам библиотеки Британского музея. В наборных кассах середины прошлого века не было достаточного числа обычных знаков препинания, и тогдашний директор Британского музея А. Паницци решил употребить для этой цели неиспользуемые литеры наборных касс: "косую" черту", "знак равенства", "плюс" и т.п. По традиции они перешли в Британскую национальную библиографию, а оттуда в международные правила книгоописания.

В нашей стране эта система разделительных знаков, которая имеет тенденцию усложняться, была введена государственным стандартом. В аналитическом описании, например, сведения о статье отделяются от названия источника знаком "двойная косая черта" (что можно видеть и в сносках этой книги). Мотивы, которыми оправдывают необходимость этих нелепых знаков, малоубедительны. Из требования автоматизации иметь при каждом элементе описания метку при вводе его в машину вовсе не вытекает необходимость выводить аналог этой метки в *человекочитаемую* форму описания.

Что же касается стремления сделать описание на малопонятных языках доступным для различения элементов, то это может иметь значение лишь для каталогизаторов, да и то в редких случаях. И ради этого мы заставляем миллионы людей составлять и читать тексты, в которых "неграмотно" употребляются знаки препинания. Нам эти мотивы не представляются достаточными для превращения описания в искусственную и неудобочитаемую запись. По всей вероятности, профессиональные каталогизаторы руководствуются в своей нормативной деятельности принципами, которые не в полной мере учитывают общекультурное значение библиографического описания и его фактическую роль в научной и деловой прозе. Следует заметить, что в условиях новой информационной технологии, когда масштабы библиографи-

рования измеряются десятками и сотнями миллионов записей в год, эти мелочи оборачиваются серьезными экономическими потерями.

Этот частный пример свидетельствует о том, что и в нашей сфере развитие не всегда происходит по прямой линии целесообразности, что в нем возможны такие несообразности, как стандартизация практики, восходящей к историческим анекдотам, дань консерватизму, фетишизация узких технических требований, ослепление международной всеобщностью и т.п.

Говоря о перспективах развития этого наиболее древнего вида информационной деятельности, долгое время развивавшегося в рамках библиотеко- и библиографаведения, необходимо остановиться на тенденции, наметившейся в последние годы, рассматривать библиографическое описание как особый информационно поисковый язык. Такого рода предложения неоднократно высказывались как в советской, так и в зарубежной литературе.

Изучение библиографического описания как особого информационного языка (библиографического языка) сделалось по-настоящему актуальным после того как библиографическая информация стала служить материалом для создания поисковых систем нетрадиционного типа, в которых элементы описания позволяют искать информацию по содержанию документов, воссоздавать структуру научных коммуникаций, строить кластеры научных сообществ и дисциплин.

Методы семиотики позволяют исследовать подобные знаковые системы в разных аспектах. С точки зрения *синтактики* (отношения знаков и их систем между собой) выявляются правила построения библиографического языка. В новых стандартах описания, в отличие от прежних каталогизационных правил, знаки препинания имеют не разделительный, а различительный смысл, т. е. не отделяют один элемент описания от другого, а маркируют следующий за знаком элемент. Однако это осуществляется весьма непоследовательно и с нарушением законов семиотики.

В аспекте *семантики* (отношения знаков к их смыслу) используют семиотические понятия денотата (обозначаемого знаком конкретного объекта) и концепта (смысла этого знака). Сравнение в этом плане библиографического языка с дескрипторным и классификационным выявляет то преимущество первого, что только библиографическое описание однозначно соответствует своему денотату, тогда как

дескриптор и классификационная рубрика находятся со своими денотатами в размытых отношениях.

В плане *прагматики* (отношения знаковой системы к обозначенной ею реальной действительности) библиографический язык позволяет моделировать структуру научных публикаций, коллективов и коммуникаций, примеры чего были неоднократно представлены в этой лекции.

Реферативный журнал ВИНТИ

В мире выходит свыше полутора тысяч реферативных журналов, которые продолжают служить специалистам основным средством доступа к мировой научной литературе. В нашей стране таким средством служит Реферативный журнал ВИНТИ, выпускаемый с 1953 г. Первые выпуски этого журнала охватывали астрономию, математику, механику и химию.

Потребовалось почти десять лет, чтобы этот журнал сложился как ведущий многоотраслевой реферативный журнал мира. В настоящее время он отражает около 1 млн. статей из 10 тыс. журналов 50 стран. Для удобства читателей РЖ ВИНТИ выходит в 22 сводных томах, 250 входящих в них выпусках и 100 отдельных выпусках, не входящих в сводные тома.

Все опубликованные материалы обязательно проходят научную экспертизу (путем их рецензирования и рассмотрения редколлегией или научными редакторами) и потому удовлетворяют требованиям, которые предъявляются современной наукой к материалам, пропускаемым в формальные каналы научной коммуникации. Из этого следует, что мнение о “ненаучности” опубликованной статьи или книги, как правило, не может быть достаточным основанием для отказа от отражения ее в виде реферата, аннотации или библиографического описания в информационных изданиях.

При этом следует помнить, что если какая-либо публикация не будет отражена в РЖ или другом информационном издании, то при нынешних потоках научных публикаций она будет практически безвозвратно потеряна для будущих поколений читателей.

Любая публикация может иметь своих читателей. Поэтому при подготовке информационных изданий недопустимо отбраковывать опубликованные документы на том основании, что они “не представ-

ляют интереса”, т. е. мало полезны. К числу важных недостатков многих информационных изданий, в том числе и изданий ВИНТИ, следует отнести то, что в них не отражаются рекламные материалы, заметки технико-экономического и хроникального характера, реферируются не все 100% описаний изобретений приоритетной заявки и не помещаются рисунки. Если информационные издания действительно призваны служить главным источником сведений о мировых достижениях в науке, то в них в той или иной форме должны отражаться все без исключения материалы мировой литературы по этой тематике.

Реферативные журналы служат своего рода путеводителями по мировой научно-технической литературе. Прочтение ученым или специалистом реферата какой-либо публикации позволяет лишь определить, может ли эта публикация содержать интересующие его сведения и следует ли ее читать, но не заменяет чтения такой публикации. Поэтому нецелесообразно стремиться к помещению в РЖ чрезмерно расширенных рефератов, если это не вызвано достаточно важными причинами (например, труднодоступностью языка исходного документа).

В РЖ и других информационных изданиях, которые не являются узкоспециализированными, а рассчитаны на широкие круги пользователей, имеющих разные задачи и разные уровни профессиональной подготовки, должна отражаться лишь *центральная тема* или *предмет* каждой публикации и непосредственно относящиеся к нему сведения, а не сопутствующие темы, предметы, сведения. Это означает, что в информационных изданиях общего назначения исходная публикация может быть вполне адекватно отражена одним рефератом или аннотацией несмотря на то, что данная публикация относится одновременно к двум и более отраслям науки, техники или производства.

Авторские рефераты или резюме, помещаемые и в отечественных, и в зарубежных журналах, становятся все более информативными и пригодными для публикации в РЖ без существенных изменений (но, разумеется, под редакторским контролем). Сам факт использования какого-либо элемента публикации в информационных изданиях способствует его совершенствованию.

При современных возможностях электронного сканирования текста и его перевода в машиночитаемую форму авторские резюме позволяют существенно сократить затраты труда и времени на подготовку РЖ. В печатных изданиях в связи с ростом стоимости бумаги

сокращение затрат на 10–15% возможно за счет более компактной верстки, совершенствования структурной схемы реферата и библиографического описания (например, отказа от заглавий на языке оригинала).

Зарубежные реферативные журналы

Для характеристики этого важнейшего вида информационных изданий выбраны 10 реферативных журналов, отражающих мировую литературу по естественным и техническим наукам.

Applied Mechanics Reviews (AMR) – ежемесячный РЖ по техническим наукам, издается Американским обществом инженеров-механиков с 1948 г., публикует около 15 тыс. рефератов в год. С середины 80-х годов перешел на частичную публикацию авторских резюме. Рефераты упорядочиваются по трехуровневой классификационной схеме, содержащей по 90 рубрик на двух первых уровнях и 1,2 тыс. – на нижнем. Основные элементы и форма представления реферата: номер реферата (сквозной по всем выпускам в пределах года, имя первого автора (нелатинские алфавиты транслитерируются), адрес первого автора, имена остальных авторов, заглавие статьи (только на английском языке), название и выходные данные журнала-первоисточника, текст реферата, имя референта с указанием страны. Имеет пономерные авторские указатели и годовой по отдельным выпускам авторский с ключевыми словами – AKWAS (Authors and Key Words in Alphabetical Sequence). Том годового указателя включает руководство для пользователя, перечни индексационных терминов, классификационных рубрик первого уровня, отражаемых первоисточников (свыше 1 тыс.), референтов и таблица транслитерации. В этом РЖ отражаются статьи из журналов, научно-технические отчеты, книги, труды конференций, ежегодники.

Astronomy and Astrophysics Abstracts (AAA) – РЖ по астрономии, астрофизике и смежным областям, выходит 2 раза в год в виде двух полугодовых томов (в переплете с суперобложкой), издается с 1969 г. Астрономическим вычислительным институтом (Astronomisches Recheninstitut, Гейдельберг, Германия) под эгидой Международного астрономического союза. В этом РЖ отражаются около 20 тыс. публикаций в год по их авторским резюме; реферируются только статьи без резюме. Статьи популярного характера отражаются только в виде библиографического описания. Сроки прохождения материалов – не более

8 месяцев. Рефераты располагаются по перечню предметных рубрик (108) и получают порядковые номера внутри раздела. Основные элементы и форма представления: номер реферата (шесть цифр: код рубрики и порядковый номер в разделе), заглавие статьи на английском языке, имя автора (нелатинский алфавит транслитерируется), название первоисточника на языке оригинала (или в транслитерации) и выходные данные, язык публикации, текст реферата (без подписи референта и указания количества ссылок). Выпускаются годовые авторский и предметный указатели с двухтомной пятилетней кумуляцией. В каждом томе дается перечень свыше 750 первоисточников, из которых 150 реферируются полностью. Перепечатываются рефераты из других РЖ, в том числе из отдельных выпусков РЖ ВИНТИ.

Biological Abstracts (BA) – РЖ по биологическим наукам, издается информационной службой BIOSIS (BioSciences Information Service, США) с 1926 г., выходит 2 раза в месяц (2 тома в год по 12 выпусков), публикует около 600 тыс. рефератов в год. Рефераты упорядочиваются по собственному рубрикатору. В начале каждого выпуска приводится перечень предметных рубрик с синонимами из смежных областей. Основные элементы и форма представления реферата: номер реферата (сквозной по всем выпускам каждого тома), имена авторов на языке оригинала (транслитерируются для нелатинских алфавитов), адрес автора, отмеченного *, на языке оригинала, название первоисточника и выходные данные, язык публикации, заглавие статьи на английском языке, текст реферата. Каждый номер имеет указатели: авторский, биосистематический, биологических родов, понятий, пермутационный и сокращений. Издаются также полугодовые кумулятивные указатели в трех отдельно издаваемых частях. Периодические и продолжающиеся издания, отражаемые в *BA* (около 7 тыс.), приводятся в ежегодно издаваемом указателе источников *Serial Sources for BIOSIS Data Base*, в котором указывается полное название издания (в алфавитном порядке сокращенных названий, выделенных в полном полужирном шрифтом), перевод названия на английский язык, примечания об изменениях, прекращении издания или его отражения в *BA*, шифр CODEN, периодичность, издатель (два последних в закодированном виде).

Chemical Abstracts (CA) – издается службой Chemical Abstracts Service (CAS) Американского химического общества с 1907 г. и отражает выпускаемую в мире литературу по химии и химической технологии. Публикует около 1 млн. рефератов в год в виде еженедельных вы-

пусков (2 тома по 26 номеров). Каждый еженедельный выпуск состоит из двух частей – собственно рефератов и указателей. Рефераты распределяются по 80 тематическим разделам и внутри каждого раздела по 7 видам первоисточников: журнальные статьи, труды конференций и сборники, научно-технические отчеты, депонированные рукописи, диссертации, сообщения о новых книгах, патентные описания. Структура библиографического описания (БО) определяется видом реферируемого документа. БО журнальной статьи включает:

- номер реферата (сквозной во всех выпусках полуугодового тома, начинается с номера тома, заканчивается /с 1967 г./ контрольной буквой для автоматической проверки правильности написания номера);

- заглавие реферата (заглавие реферируемого документа на английском языке воспроизводится полужирным шрифтом, на других языках дается в переводе на английский, для книг приводится также заглавие на языке оригинала);

- полные имена авторов в инвертированной форме, приводятся до 10 имен, при большем числе – 9 с сокращением *et al.* (и др.), нелатинский алфавит транслитерируется);

- место проведения исследования или адрес для переписки;

- сокращенное название первоисточника (курсивом в соответствии с *International List of Periodical Title Word Abbreviation*);

- год издания (полужирным шрифтом);

- номер тома и (в скобках) номер выпуска;

- начальная и последняя страницы публикации;

- язык публикации.

Библиографические описания других видов документов отличаются от приведенного дополнительными элементами (такими, например, как дата проведения конференции, номер научно-технического отчета, место депонирования рукописи и т. п.). Каждый еженедельный выпуск включает указатель ключевых слов, авторский и патентный указатели. Указатель ключевых слов, выбранных или составленных по заглавиям и/или текстам рефератов, представляет собой их алфавитный перечень с пояснительными записями. В патентном указателе перечисляются по странам и номерам описания всех впервые реферируемых патентных документов, даются перекрестные ссылки на первый документ, если их несколько, и при первом документе – перечень всех других, относящихся к этому изобретению. Патентные документы

систематизируются также по кодам видов документов, принятым всеми странами. К каждому тому *СА* издаются кумулятивные годовые указатели: авторский, ключевых слов, патентный и предметный из двух частей: общий и химических соединений.

Отражаемые периодические и другие издания и документы включены в Указатель источников (*The Chemical Abstracts Service Source Index – CASSI*), кумулятивный выпуск с 1907 г. и квартальные (с кумуляцией за год) дополнения. Сведения о новых журналах и об изменении названий публикуются в *CASSI* и в каждом еженедельном выпуске *СА* непосредственно за авторским указателем. В начале кумулятивного выпуска *CASSI* за 1907–1979 гг. приведен перечень 1 тыс. наиболее продуктивных журналов, алфавитный и ранжированный по продуктивности. Всего в *CASSI* описано 50 тыс. источников информации.

Computer and Control Abstracts (CCA) – ежемесячный РЖ по вычислительной технике и управлению, издается с 1966 г., публикует более 50 тыс. рефератов в год. Полугодовые кумулятивные указатели издаются в виде одного тома, включающего указатели: авторский, библиографий, книг, трудов конференций, коллективных авторов, предметный, отражаемых журналов и (отдельно) других периодических и продолжающихся изданий. Авторские и предметные указатели кумулируются каждые три года.

Electrical and Electronics Abstracts (EEA) – ежемесячный РЖ по электротехнике и электронике, издается с 1903 г., отражает свыше 80 тыс. публикаций в год. Полугодовые кумулятивные указатели издаются двумя отдельными томами: предметный составляется на основе тезауруса *INSPEC*, в авторском имена авторов сопровождаются заглавиями статей, а также приведены “малые” указатели – библиографий, книг, трудов конференций, коллективных авторов и источников. Авторские и предметные указатели кумулируются каждые четыре года.

Physics Abstracts (PA) – реферативный журнал по физике издается с 1903 г. по два выпуска в месяц, ежегодно отражает свыше 150 тыс. публикаций. Полугодовые кумулятивные указатели издаются тремя отдельными томами: два тома предметного (A–L, M–Z) и один авторский. В остальном указатели аналогичны предыдущему, включая четырехлетнюю кумуляцию. В конце 1994 г. в РЖ *Physics Abstracts* влился другой РЖ по физике – *Physics Briefs/Physikalische Berichte*, который выходил в Германии под разными названиями с 1848 г. С января 1995 г. этот объединенный и расширенный журнал выходит под

названием *Physics Abstracts* в виде двухнедельных выпусков.

Все три РЖ – *CCA*, *EEA*, *PA* – издаются Институтом инженеров-электриков (Institute of Electrical Engineers – IEE) Великобритании и в настоящее время подготавливаются при помощи автоматизированной информационной системы INSPEC (Information Services in Physics, Electrotechnology, Computers and Control), функционирующей с 1969 г. Они имеют одинаковую структуру, формы представления реферируемых материалов и систему указателей. С 1898 г. IEE совместно с Лондонским физическим обществом издавал РЖ по физике и электротехнике *Science Abstracts*, который в 1903 г. разделился на две серии – *PA* и *EEA*, а с 1966 г. к ним прибавилась третья – *CCA*. Рефераты в каждом выпуске этих РЖ упорядочиваются по четырехуровневой рубрикации соответствующих отраслевых рубрикаторов.

В начале каждого выпуска дается перечень предметных рубрик с подрубриками и классификационными кодами, а за ним – предметный указатель. Обязательные элементы реферата журнальной статьи: номер реферата (сквозной по всем выпускам за год), заглавие статьи (на английском языке в оригинале или переводе, для статей на другом языке он указывается), имя и адрес автора (нелатинские алфавиты транслитерируются), название и выходные данные первоисточника, текст реферата, количество ссылок. Пономерные указатели: авторский, библиографий (перечень статей с обширными пристатейными библиографиями), книг, трудов конференций, коллективных авторов (организаций, ответственных за содержание документа), новых журналов (дополнительный перечень). В указателе источников INSPEC приведено свыше 3 тыс. журналов, из которых около 600 реферируются полностью, а также свыше 400 нежурнальных периодических и продолжающихся изданий. Все содержание этих РЖ доступно в машиночитаемом виде.

Engineering Index Monthly (Ei Monthly) – ежемесячный РЖ, издается с 1884 г. (до 1962 г. – под заглавием *Engineering Index*) фирмой Engineering Information, Inc. (до 1981 г. – Engineering Index, Inc., США) и охватывает литературу, публикуемую в мире по всем техническим отраслям (около 15 тыс. рефератов в каждом выпуске). Рефераты упорядочиваются по основным предметным рубрикам, выбираемым из отдельно изданного перечня индексационных терминов *Subject Headings for Engineering (SHE)*, из которого используется около 12 тыс. рубрик, подрубрик и их сочетаний.

Рефераты, опубликованные в 12 выпусках *Ei Monthly* за каждый календарный год, кумулируются в алфавите предметных рубрик и под- рубрик и издаются с 1959 г. в нескольких переплетенных томах издания *Engineering Index Annual (Ei Annual)* – массив с 1884 г. доступен на микрофильмах. Номера рефератов в *Ei Monthly* и *Ei Annual* не совпадают, основные элементы и форма представления реферата – общие: номер реферата (сквозной по всем выпускам за год), заглавие статьи на английском языке (или в транслитерации с переводом на английский в скобках), текст реферата, количество ссылок, имена авторов и адрес первого, сокращенное название первоисточника, том, выпуск, дата и страницы, предметная рубрика, подрубрика, перекрестная ссылка.

Каждый выпуск *Ei Annual* включает перечень сокращений, а *Ei Monthly* – авторский указатель с заглавиями статей при именах авторов. Для помощи читателям в определении предметных рубрик и под- рубрик отдельно издается *Subject Headings Guide to Engineering Categories*. Реферативный журнал *Ei Annual* помимо рефератов включает указатели: источников, отраженных в текущем году (*Publication Index for Engineering – PIE*), кумулятивный авторский, адресов авторов, соответствия номеров *Ei Monthly* и *Ei Annual*.

Указатель *PIE* включает перечень источников по сокращенным в соответствии с американским стандартом названиям (с шифрами CODEN, новыми и измененными шифрами), источники без шифров, труды конференций, перекрестные ссылки от названия организации к сокращенному названию, от транслитерированных названий на неанглийском языке, от переводов на английский язык названий на японском и китайском языках, от названий, начинающихся с сокращений. Всего в *PIE* описано свыше 2,5 тыс. первоисточников.

Выпускаются также специализированные ежемесячные РЖ *Energy Abstracts* (с 1974 г.) – по традиционным и новым источникам энергии, *Bioengineering Abstracts* (с 1975 г.) – по биотехнике; библио- графическая серия *Technical Bulletins* – по актуальным вопросам техники; пятилетние кумулятивные указатели к *Ei Annual*. На машиночи- таемых носителях доступны БД COMPENDEX (аналог *Ei Monthly*) и *Ei Engineering Meetings*, включающая рефераты материалов около 2 тыс. конференций (с 1984 г. издается и в печатном виде).

Mathematical Reviews (MR) – ежемесячный РЖ, издается Амери- канским математическим обществом с 1940 г. Охватывает все разделы теоретической и прикладной математики, публикует ежегодно около

50 тыс. рефератов. Отражает полностью 8 журналов Американского математического общества, 7 российских (в переводе на английский язык), а всего около 1,5 тыс. периодических и продолжающихся изданий. Перепечатывает также (с указанием источника) рефераты других РЖ: *Applied Mechanics Reviews*, *Computing Reviews* (США), *Physics Abstracts* (Великобритания), *Zentralblatt für Mathematik und ihre Grenzgebiete* (ФРГ), *РЖ ВИНТИ* (Россия). Время опубликования рефератов с момента издания первоисточника – 7 месяцев, средний объем реферата – 200 слов.

Рефераты систематизируются по трехуровневой рубрикации; под названием рубрики, которая может быть и пустой, указываются номера рефератов, частично относящихся к данному разделу, но помещенных в других разделах. Выпуски могут охватывать от одного до четырех разделов. Нумерация рефератов производится по разделам: год выхода (2 цифры), буквенный код выпуска (1), “двоеточие”, шифр раздела (2), “пробел”, порядковый номер реферата в разделе (3).

БО журнальных статей имеет следующую структуру: имя автора, заглавие статьи (на английском, немецком и французском языках, заглавия на других языках – в английском переводе), язык оригинала, резюме на других языках, сокращенное название источника, том, год, выпуск, страницы. Дается указание о наличии перевода статьи на английский язык. В конце текста реферата указывается количество рисунков, таблиц и библиографических ссылок, а также имя и шифр референта. В описании книги или периодического издания как целого ставится астериск (*), а имя автора и заглавие приводятся в оригинале и переводе на английский язык.

В каждом выпуске РЖ два указателя: авторский (все авторы) и ключевой (предметные заголовки, названия конференций и семинаров, географические названия и виды изданий – словарь, биография, некролог и т. п.; ключевые фразы до 8 слов). В последнем номере приводятся годовые кумулятивные указатели: полный авторский (с заглавиями публикаций) и систематический (по рубрикатору). Авторские указатели кумулируются за несколько (5–20) лет. Создана БД MATNFILe, охватывающая все рефераты с 1973 г.

Meteorological and Geostrophysical Abstracts (MGA) – ежемесячный РЖ по метеорологии и геоастрофизике, издается Американским метеорологическим обществом с 1960 г. (до этого назывался *Meteorological Abstracts and Bibliography*); ежегодно отражает свыше

7 тыс. публикаций из 150 журналов и нескольких десятков ежегодников и продолжающихся изданий, названия которых приводятся в начале каждого выпуска РЖ. Рефераты упорядочиваются в соответствии с рубрикацией (с перекрестными ссылками). Обязательные элементы реферата: номер (в пределах одного выпуска), имя автора, его адрес, заглавие статьи на языке оригинала (нелатинские алфавиты транслитерируются, а заглавие переводится на английский язык), полное название и выходные данные первоисточника (географические названия – только по-английски), количество ссылок, иллюстраций, таблиц, код библиотеки (где хранится источник), текст реферата, предметные рубрики, имя референта, код рубрикации. Каждый выпуск содержит авторский, предметный и географический указатели.

Электронная информация и базы данных

Все более важным видом источников первичной и вторичной информации становятся *базы данных* (БД), представляющие собой в большинстве случаев электронные версии печатных изданий – газет, научных, научно-технических и общественно-политических журналов и бюллетеней, энциклопедий, справочников и т. п. В настоящее время принята следующая типология БД: 1) *первичные* (source) БД, которые подразделяются на цифровые, текст-цифровые, полнотекстовые и по физико-химическим свойствам веществ; 2) *вторичные* (reference) БД, которые подразделяются на библиографические (в том числе реферативные) и адресно-справочные. Эти БД распространяются на машиночитаемых носителях (магнитных лентах, дискетах, оптических компакт-дисках) и после загрузки в ЭВМ используются для автоматизированного поиска информации. В мире насчитывается около 20 тыс. общедоступных БД, и их число ежегодно возрастает.

Составной частью системы информационных изданий является служба быстрого изготовления и доставки читателям (потребителям информации) *копий первоисточников*, отраженных в этих изданиях. Сроки изготовления и доставки копий первоисточников соразмерны времени, затрачиваемому на проведение поиска информации о них в банках данных.

Большое значение в научно-информационной практике имеют *указатели цитирования литературы*, которые являются не только эффективным инструментом библиографического информационного по-

иска, но и уникальным средством для наукометрических исследований. Они позволяют оценивать вклад отдельных ученых и стран в мировую науку, выявлять наиболее активные точки роста в науке, определять взаимосвязи между конкретными исследованиями (публикациями) и научными дисциплинами, решать другие важные задачи. Печатные и машиночитаемые версии таких указателей, выпускаются в США фирмой Institute for Scientific Information с 1961 г.

Внедрение средств вычислительной техники в сферу подготовки печатных изданий привело к появлению электронных аналогов этих изданий – журналов, справочников, энциклопедий и т. п. В таких электронных изданиях текст и иллюстрации представлены в цифровой форме, что позволяет производить их автоматизированную обработку и передавать по каналам электросвязи.

В электронных изданиях с помощью компьютеров можно проводить также быстрый поиск нужной информации по сочетаниям разных признаков – ключевых слов, имен и т. д. В настоящее время все крупнейшие газеты и журналы США, а также других развитых стран издаются как в печатной, так и в электронной форме. Электронные версии печатных изданий называются также "полнотекстовыми базами данных".

В последнее время начали выходить научные журналы, уже не имеющие печатных версий. В связи с появлением все большего числа электронных журналов и книг некоторые специалисты предсказывают, что электронные издания будут вытеснять печатные журналы и книги и что поэтому библиотеки в их традиционном понимании не имеют будущего. Однако такие предсказания противоречат историческому опыту, который свидетельствует, что новые носители и средства распространения информации обычно дополняют, а не заменяют уже существующие.

Базы данных как средства поиска и распространения научной и иной информации появились 40 лет назад. *База данных (БД) представляет собой упорядоченную совокупность информационных сообщений – библиографических описаний статей, рефератов, записей фактов или иных текстов, относящихся к какой-либо теме и представленных на машиночитаемом носителе (магнитной ленте, дискете, компакт-диске).* Поиск информации в БД можно производить как на своем компьютере, так и в режиме теледоступа по каналам электросвязи к центральному компьютеру, в котором она имеется.

В России крупнейшим производителем БД по естественным и техническим наукам является Всероссийский институт научной и технической информации (ВИНИТИ), а по общественным наукам – Институт научной информации по общественным наукам (ИНИОН).

Для сбора и хранения БД, а также для обеспечения поиска в них информации созданы специальные службы или "банки данных", число которых в мире составляет около 800, причем наблюдается устойчивая тенденция к их укрупнению. Крупнейшими зарубежными службами автоматизированного поиска информации в БД являются американские Dialog и Mead Data Central, французская Télésystèmes -Questel и международная STN International.

Служба Dialog (1972–) находится в г. Пало-Альто (шт. Калифорния), предоставляет доступ более чем к 400 БД и имеет 155 тыс. абонентов в 100 странах мира. Служба Mead Data Central (1968–) находится в г. Дейтон (шт. Огайо) и предоставляет доступ к 450 БД. Она располагает наиболее полным собранием полнотекстовых БД и большим опытом поиска информации в них. Только в двух БД этой службы – NEXIS (наука, техника, экономика) и LEXIS (судебно-юридическая информация) – ежедневно проводится примерно 200 тыс. поисков.

Служба Télésystèmes -Questel (1979–) находится в Париже и предоставляет доступ более чем к 150 БД. В начале 1994 г. она приобрела американскую службу ORBIT (1970–, свыше 70 БД), которая включена в ее североамериканский филиал (г. Александрия, шт. Виргиния).

Служба STN International (1987–) представляет собой сеть из трех банков данных – Chemical Abstracts Service (г. Колумбус, шт. Огайо, США), Информационного центра по энергетике, физике и математике (г. Карлсруэ, Германия) и Японского информационного центра по науке и технике (г. Токио). Эта служба предоставляет доступ более чем к 170 БД. С 1992 г. центры теледоступа к STN International открыты в Москве и Новосибирске.

Сети передачи и средства хранения и обработки данных

Для теледоступа к БД, находящимся в специальных службах поиска информации, кроме персонального компьютера необходимы со-

ответствующие каналы и сети связи. И такие сети были созданы. В настоящее время крупнейшей сетью, позволяющей компьютерам взаимодействовать друг с другом в масштабе реального времени, является Internet, началом которой послужила сеть ARPANET, созданная в 1969 г. министерством обороны США. Сегодня Internet представляет собой фактически объединение примерно из 9 тыс. других сетей. Если в 1981 г. сеть Internet связывала всего 213 компьютеров, а в 1989 г. – 80 тыс., то в январе 1993 г. число подключенных к ней компьютеров превысило 1,3 млн, в январе 1995 г. – 3,5 млн, а в настоящее время их число перевалило за 1 млрд., причем они находятся в 160 странах мира. Сеть Internet все больше используется учеными и специалистами для получения научной информации, тем более что для многих из них пользование этой сетью является бесплатным. В настоящее время через сеть Internet открыт доступ к электронным каталогам более 200 библиотек, в том числе к каталогу Библиотеки конгресса США (свыше 30 млн каталожных карточек), а также к службам поиска информации в БД. В настоящее время к Internet подключено более четверти всех публичных библиотек, обслуживающих население численностью не менее 250 тыс. человек каждая.

В США завершается создание еще более мощной Национальной сети для исследований и образования (National Research and Education Network – NREN). Сейчас эта сеть позволяет передавать данные со скоростью 1,5 млн бит/с, что эквивалентно примерно 50 страницам текста. К 1996 г. быстродействие NREN было доведено до 3 млрд. бит/с. Это позволило за одну секунду передавать 100 тыс. страниц текста, т. е. все 32 тома Британской энциклопедии. Аналогичные быстродействующие сети цифровой связи созданы и развиваются в Великобритании, Германии, Франции, Японии и других передовых странах мира.

Создание быстродействующих цифровых сетей передачи данных позволило разработать и внедрить видеографические системы типа "видеотекс". Особенно большие успехи в этой области достигнуты во Франции: созданная там видеографическая система Télétel основана на использовании телефонных линий связи. Видеотерминалы этой системы, получившие название Minitel, устанавливаются в учреждениях, гостиницах и частных домах. Эти видеотерминалы, число которых превысило 5 млн, позволяют запрашивать и получать самую разнообразную информацию – сводки погоды, сведения о репертуарах театров,

котировки акций, расписания самолетов и поездов, номера телефонов и т. п., а также осуществлять поиск информации в БД службы Télésystèmes-Questel. Кроме того, такая система позволяет абонентам вести переписку друг с другом.

Информационные услуги

Современные информационные услуги можно разделить на три взаимодействующих области:

- специальная информация,
- электронные сделки,
- электронная коммуникация.

В области *специальной информации* выделяются следующие основные сектора:

1. *Сектор деловой информации* (биржевой, финансовой, коммерческой, экономической, статистической), охватывающий:

- биржевую и финансовую информацию – информацию о котировках ценных бумаг, валютных курсах, учетных ставках, рынке товаров и капиталов, инвестициях, ценах, предоставляемую биржами, специальными службами биржевой и финансовой информации, брокерскими компаниями, банками;

- экономическую и статистическую информацию – числовую экономическую, демографическую, социальную информацию в виде рядов динамики, прогнозных моделей и оценок, предоставляемую государственными службами, а также компаниями, занятыми исследованиями, разработками и консалтингом;

- коммерческую информацию – информацию по компаниям, фирмам, корпорациям, ценам, о финансовом состоянии, связях, сделках, руководителях, о направлениях их работы и продукции и т. п.;

- деловые новости в области экономики и бизнеса, предоставляемые специальными информационными службами.

2. *Сектор научно-профессиональной информации*: научно-технической, медицинской, юридической и другой информации, охватывающий документальную, библиографическую, реферативную, справочную информацию и данные в области фундаментальных и прикладных, естественных, технических и общественных наук, отраслей производства и сфер человеческой деятельности. В этом секторе предусмотрена организация доступа к первоисточникам информации через библиотеки и специализированные службы, возможность приобрете-

ния первоисточников, их получения по межбиблиотечному абонементу в виде полноразмерных и микрокопий.

3. *Сектор массовой, потребительской информации*, охватывающий новости и литературу, справочники, энциклопедии, потребительскую и развлекательную информацию, ориентированную на домашнее, а не служебное использование, в том числе сведения о погоде, расписании транспорта, игры, программное обеспечение, предложения по обмену, покупкам и продажам, справочники отелей и ресторанов, информацию по курсам валют, аренде, турах, а также услуги телетекста и видеотекста и т. п.

4. *Сектор социально-политической информации*, обслуживающий органы государственной власти и управления статистической, социальной, архивной и другой специальной информацией.

В России понятие информационной деятельности в течение долгого времени связывалось в основном с научно-технической, а также некоторыми видами социальной информации, в то время как на них приходится около 15–25 % мирового рынка информационных услуг.

В бывшем СССР были созданы достаточно развитые, но замкнутые автоматизированные информационные системы, направленные в основном на реализацию внутренних задач тех ведомств, в рамках которых и на средства которых они существовали. Задачи обеспечения научной и технической информацией решались в рамках Государственной системы научной и технической информации (ГСНТИ), где ведущая роль генератора информации принадлежала ВИНТИ, статистическая информация по состоянию экономики была сосредоточена исключительно в ведомствах Государственного комитета по статистике и т. д. Открытая информация, сосредоточенная в основном в рамках системы научной и технической информации, была изолирована от системы информации в оборонных отраслях и системы массовой информации.

Что касается экономической, коммерческой (конъюнктурной), массовой и прочей информации, то командно-административная система удовлетворялась ею в том виде и количестве, которое было необходимо для целей директивного управления. Такие важнейшие для нужд рыночной экономики информационные составляющие, как биржевая и финансовая информация, демографическая и потребительская информация, еще предстоит создавать. В условиях централизованно планируемой экономики основная часть информации, необходимой

предприятиям и организациям в их хозяйственной деятельности, поступала в виде директив административной системы и носила служебный характер. С разрушением административной системы предприятия столкнулись с острой нехваткой практически любой информации делового характера, необходимой для работы в условиях формирующегося рынка. Отсутствие такой деловой информации не способствовало восстановлению внутриотраслевых и межотраслевых экономических связей и возникновению негосударственных экономических партнеров (кооперативов, малых предприятий, банков, бирж).

Мировая практика последних десятилетий показала, что информация в электронной форме превратилась в важнейший компонент современной рыночной инфраструктуры. Доступность информации тесно связывается западными экономистами со свободой конкуренции и рассматривается как одно из базисных условий эффективного функционирования рыночной экономики.

Независимый, самостоятельный производитель товаров или услуг (а именно эта фигура наилучшим образом характеризует рыночную экономику и ее основного субъекта), а также все те, кто обеспечивает непрерывность цикла "наука–техника–производство–сбыт–потребление" не могут успешно действовать на рынке, не имея информации. Это информация о других производителях, о возможных потребителях продукции, о поставщиках сырья, комплектующих и технологиях, о положении на товарных рынках и рынках капитала, наконец об общей экономической и политической ситуации не только в собственной стране, но и во всем мире, тенденциях развития экономики, перспективах развития науки и техники, о правовых условиях хозяйственной деятельности и т. п.

Информация лежит в основе современной маркетинговой концепции управления производственной и сбытовой деятельностью и используется как для формирования и уточнения производственной программы, так и для проверки ее обоснованности рынком в процессе реализации произведенных товаров. Современный этап развития экономики требует наличия специального механизма, обеспечивающего связь между предложением и спросом, ориентирующегося на конкретные потребности рынка и подчинение этим потребностям различных сторон производства и сбыта. Важной частью такого механизма выступает информационная деятельность.

Электронные сделки (операции) включают системы резервирования билетов и мест в гостиницах, заказа товаров и услуг, банковских и расчетных операций. Они формируются:

- системами заказа билетов на железнодорожном и авиационном транспорте (системы резервирования авиабилетов "Сирена" и железнодорожных билетов "Экспресс");

- системами электронных банковских операций;

- так называемыми, электронными биржами.

Электронные коммуникации реализуются современными средствами связи и человеческого общения. Это сети передачи данных, системы электронной почты, компьютерные сети, телеконференции, электронные сетевые доски объявлений и бюллетени, центры общедоступного программного обеспечения и т. п.

Информационные структуры и инфраструктура

Информационными структурами принято считать учреждения и организации, профессионально производящие, обрабатывающие и «переупаковывающие» информацию разного рода, обеспечивающие ее хранение и доступ к ней пользователям. Это информационные агентства, службы, центры и институты информации, библиотеки и другие подобного рода учреждения и организации.

Наиболее упорядоченными информационными структурами являются системы научно-технической информации, существующие фактически в любой стране. Особенности построения каждой такой системы зависят от многих причин: географического положения страны, размеров ее территории, численности населения, используемых в ней языков, ее истории, уровня экономического развития, политического строя и др. Такие системы могут быть централизованными, децентрализованными и смешанного типа.

В настоящее время, когда информация все больше становится экономическим фактором, правительства всех стран принимают доступные им меры для укрепления и развития таких систем в своих странах. Эти меры составляют содержание *информационной политики*, которая разрабатывается и осуществляется в каждой стране.

В странах, ставших политически независимыми после второй мировой войны и/или не имевших исторически сложившихся научно-информационных систем, они стали создаваться по централизованной схеме. Она была рекомендована ЮНЕСКО и предусматривала созда-

ние национальных центров и находящихся под их методическим руководством сетей специальных центров и научно-технических библиотек. Такие национальные центры были созданы в Аргентине, Бразилии, Египте, Израиле, Индии, Индонезии, Иране, Испании, Китае, Пакистане, Португалии, Турции, ЮАР, Южной Корее, Японии, в странах СЭВ и многих других странах.

Для общего руководства работой национальной информационной системы и главных направлений ее развития в большинстве стран созданы *межведомственные комиссии* или *комитеты*, которые функционируют при соответствующем органе государственной власти (парламенте, премьер-министре, министерстве или госкомитете). Такие межведомственные комиссии (комитеты) выполняют следующие основные функции:

- организуют разработку прогнозов развития информационных систем и проводят их экспертную оценку;
- определяют цели и разрабатывают основные направления государственной информационной политики;
- разрабатывают среднесрочные (до 5 лет) и долгосрочные (10 и более лет) программы развития национальной информационной системы;
- осуществляют научную экспертизу проектов, направленных на развитие национальной системы;
- организуют разработку законодательных актов и других нормативных документов, относящихся к деятельности национальной системы.

Основными средствами, которые применяются для решения задач, одобренных межведомственными комиссиями или комитетами, являются:

- законодательные акты;
- прямые распоряжения органов государственного управления административно подчиненным организациям и учреждениям;
- прямые государственные заказы на выполнение конкретных заданий;
- государственное субсидирование проектов, способствующих достижению поставленных комиссией (комитетом) целей.

Научно-методическое руководство национальной информационной системой (сетью) той или иной страны осуществляется либо специально созданным для этой цели национальным институтом (цен-

тром) научной и технической информации, либо национальной библиотекой.

В странах, которые имеют давно сложившиеся механизмы и традиции децентрализованного управления научными исследованиями и их информационного обеспечения, они основаны на принципах координации и кооперации. Эти страны раньше других осознали большое значение научных исследований и разработок для развития экономики и культуры. Поэтому в таких странах были созданы эффективно действующие сети научно-технических библиотек, которые преимущественно и осуществляли научно-информационное обслуживание исследователей (т. е. выполняли функции информационных центров) вплоть до начала второй мировой войны. К числу таких стран относятся США, Великобритания, Германия, Франция, Италия и скандинавские страны.

Важным средством повышения эффективности национальной системы (сети) информационных органов, особенно при наличии в стране частного и государственного секторов экономики, является развитие и укрепление *научных обществ, научно-исследовательских ассоциаций* в отраслях промышленности и *профессиональных ассоциаций*. Эти общества и ассоциации обеспечивают неформальное взаимодействие между частным и государственным секторами экономики, в том числе и в области информационного ее сектора. Они решают следующие основные задачи:

- выпускают научно-техническую литературу, особенно периодические издания, по своим отраслям;
- организуют научные совещания ученых и специалистов;
- создают центры и системы информационного обеспечения своих членов;
- организуют и/или проводят исследования в интересах всех своих членов;
- разрабатывают общие методические и нормативные материалы, необходимые для взаимодействия между государственными и частными организациями и учреждениями;
- разрабатывают общие профессиональные требования к специалистам собственной сферы, организуют и контролируют их подготовку и повышение квалификации;
- развивают координацию и кооперацию между однотипными организациями и учреждениями;

– информируют общественность о целях, задачах и достижениях своих членов.

В области научно-технической информации такими обществами и профессиональными ассоциациями в США являются: Американская библиотечная ассоциация (1876–, св. 47 тыс. чл.), Ассоциация специальных библиотек (1909–, 4 тыс. чл.), Ассоциация научных библиотек (1932–, ок. 120 чл.), Ассоциация медицинских библиотек (1898–, 5 тыс. чл.), Ассоциация по образованию в области библиотековедения и информатики (1915–, ок. 700 чл.) и еще не менее 30 других библиотечных ассоциаций; Библиографическое общество Америки (1904–, ок. 1,4 тыс. чл.), Американское общество по информатике (1937–, ок. 4,3 тыс. чл.), Ассоциация информационной промышленности (1968–, ок. 600 чл.), Американское общество индексаторов (1968–, св. 750 чл.), Ассоциация американских издателей (1970–, св. 250 чл.), Общество научного книгоиздания (1979–, ок. 600 чл.), Американская ассоциация книготорговцев (1900–, 7,3 чл.) и многие другие.

В Великобритании важнейшими обществами и профессиональными ассоциациями в области библиотечного дела и информатики являются: Библиотечная ассоциация (1877–), Ассоциация по информационному менеджменту (прежде Ассоциация специальных библиотек и информационных бюро, 1926–, ок. 2 тыс. чл.), Институт информационных специалистов (1958–, св. 2 тыс. чл.), Общество индексаторов (1957).

Во Франции такими организациями являются: Ассоциация французских библиотекарей, Французская ассоциация документалистов и работников специальных библиотек, Французский союз органов документации (1932–). В Германии насчитывается несколько десятков обществ и профессиональных ассоциаций в области библиотечного дела и информатики, из которых крупнейшими являются: Объединение немецких библиотекарей (1900–, 1350 чл.), Немецкий библиотечный союз (1949–, 2,3 тыс. чл.), Объединение дипломированных библиотекарей научных библиотек (1948–, ок. 3 тыс. чл.), Объединение немецких документалистов.

В настоящее время стал модным термин *инфраструктура*, который неправильно употребляют в значении внутренней структуры. Понятие инфраструктуры на самом деле определяется как *совокупность отраслей хозяйства, которые организационно не входят в данную отрасль, но без участия и развития которых невозможно эф-*

фективное функционирование последней. Для информационной отрасли инфраструктурными отраслями можно считать издательское дело и полиграфическую промышленность, книжную торговлю, библиотечное дело, почтовую и телеграфную связь, телевидение и некоторые другие отрасли.

В информатике это понятие связано с понятиями национальной и глобальной инфраструктуры и впервые возникло в США. В речах о политике в области науки и технике 17 и 22 февраля 1993 г. президент США Б. Клинтон объявил о намерении своей администрации создать в стране Национальную информационную инфраструктуру и супермагистраль (National Information Infrastructure – НИИ)¹.

Идея создания инфраструктуры и ее станового хребта – *информационной супермагистрали* (information superhighway) или *инфобана* (Infobahn) была выдвинута в 1989 г. А. Гором, тогда еще сенатором, а затем вице-президентом США. Им же введен в обращение в 1978 г. и сам термин “информационная супермагистраль”. Осуществление этой идеи – по замыслу ее авторов – должно было стать национальной задачей США, вдохновляющей и сплачивающей исследователей, инженеров, бизнесменов и других граждан этой страны, т. е. примерно такой, какой была в 60-е годы программа “Аполлон” по выполнению полетов американских астронавтов на Луну².

За всю 219-летнюю историю США по этой стране прокатились лишь две обновляющие экономические волны, которые радикально изменили образ жизни ее населения. Первую волну породило строительство железных дорог в середине XIX века, а вторую – строительство скоростных автострад, особенно между штатами, в 50-х годах XX века.³ Третью волну должно было породить строительство информационной супермагистрали, что могло бы оказать на жизнь американцев

¹ Lowe S. J. Data communications // IEEE Spectrum. – 1995. – Vol. 32, N 1. – P. 26–29.

² Doty P., Bishop A. P. The National Information Infrastructure and electronic publishing: A reflective essay // Journal of American Society for Information Science. – 1994. – Vol. 45, N 10. – P. 785–799; Goddard S. B. The information superhighways: Crisis and opportunity // Library Journal. – 1994. – Vol. 119, N 12. – P. 56; Williams L. Libraries without walls: Vision for a National Information Infrastructure // Journal of Agriculture and Food Information. – 1994. – Vol. 2, N 2. – P. 3–9.

³ Инициатором программы строительства скоростных автострад между штатами был сенатор А. Гор, отец бывшего вице-президента.

и людей всего мира не меньшее воздействие, чем в XV веке оказало на жизнь народов Европы изобретение книгопечатания Гутенбергом¹.

В документе “План действий по созданию Национальной информационной инфраструктуры”, который был обнародован Белым Домом 15 сентября 1993 г., эта инфраструктура видится как “бесшовное сплетение коммуникационных сетей, компьютеров, баз данных и устройств бытовой электроники, которое будет доставлять непосредственным потребителям огромные объемы информации. Создание национальной информационной инфраструктуры может способствовать развязыванию информационной революции, которая навсегда изменит образ жизни людей, то, как они работают и взаимодействуют друг с другом”.

В другом документе Национальная информационная инфраструктура трактуется как программа, направленная “на повышение эффективности работы правительства”, на “сохранение Соединенными Штатами мирового лидерства в науке, технике и технологии”, на то, чтобы “все американцы получали нужную им информацию тогда, когда она нужна им, когда они хотят ее получить – и за приемлемую плату”².

Как следует из названия этой программы, в ней предусматривалось создание информационной *инфраструктуры*. Это означало разработку средств, позволяющих быстро передавать и получать информацию, организационно не входящих в существующие информационные структуры (например, как почта, которая лишь обеспечивает быструю доставку корреспонденции адресатам, но не создает и не использует ее сама). Поэтому, чтобы собственно информационные структуры – центры информации, библиотеки и т. п. – могли эффективно пользоваться этой инфраструктурой (т. е. сверхбыстродействующей сетью передачи данных), они должны быть переоснащены новейшим информационным оборудованием, создавать новые виды информационной продукции, внедрять новые методы информационного обслуживания.

Основой для создания информационной супермагистрали должна служить сначала сеть Internet, а затем – сеть NREN, которая созда-

¹ Maddox I. The great multimedia revolution // Nature. – 1994. – Vol. 367, N 6460. – P. 213.

² Gordon M. L., McKenzie D. J. P. From county roads to superhighways: Keeping pace with the new business and legal turns on the international superhighway // Illinois Libraries. – 1994. – Vol. 76, N 3. – P. 124.

ется с 1989 г. тоже по инициативе А. Гора. С созданием инфраструктуры тесно связывают программы оцифровки рукописных и печатных фондов библиотек и появление так называемых “виртуальных библиотек”.

Информационный поиск

Предыстория и сущность

"Знание бывает двух видов. Мы знаем предмет по существу, или же мы знаем, где можно найти информацию о нем". В этой простой мысли давно известного высказывания английского писателя XVIII в. Сэмюэля Джонсона содержится главный признак, по которому деятельность информационная отделяется от научно-исследовательской. В ходе научного исследования возникает новое знание, а в сфере информации происходит отчуждение этого знания от его творцов и превращение в общее достояние. Однако простота этого разграничения мнимая, потому что в науке грань, отделяющая информационную деятельность от исследовательской, непостоянна. Развитие информационной технологии все время сдвигает эту грань, поскольку информационной деятельности становятся подвластны все более сложные процессы переработки знаний. То, что вчера еще делали сами исследователи, сегодня оказывается целесообразным передать информационным работникам.

Эти соображения, уже высказанные в лекции об информационной деятельности, уместно повторить, начиная разговор об информационном поиске, поскольку он является основным процессом этой деятельности и на протяжении нескольких десятилетий – центральной проблемой информатики. Новая информационная технология меняет подход к этой извечной проблеме и во многом определяет сегодня развитие информационных систем. Но информационный поиск как процесс и проблема известен с давних пор (наиболее ранние из дошедших до нас информационно-поисковых систем насчитывают тысячелетий) и продолжает волновать ученых и специалистов-практиков.

Само понятие информационного поиска появилось только в середине нашего века. Оно объединило такие, казалось бы, разные виды деятельности, как составление библиотечных каталогов и библиогра-

фических указателей, организация библиотек и справочно-информационного обслуживания, архивное дело, создание словарей, справочников, энциклопедий, вспомогательных указателей к монографиям и сборникам.

В основе этого понятия лежит представление о том, что поиск необходимой информации в любом собрании документов практически невозможен путем прочтения или даже беглого просмотра текстов всех документов данного собрания. Поэтому уже с незапамятных времен для поиска информации применяют ряд логических процедур, которые в совокупности и составляют процесс информационного поиска. Прочтение полного текста документа заменили просмотром заглавий, аннотаций, рефератов. Однако и эта процедура в многотысячных собраниях документов оказалась слишком трудоемкой. Документы пришлось систематизировать по содержанию, которое условно стали обозначать индексами, т. е. буквами и/или цифрами. Систематизация по разделам наук (классам) – один из самых первых способов раскрытия содержания научно-технических документов, моделирующий работу человеческого сознания и восходящий к глубокой древности.

По мере увеличения количества письменных и печатных документов и объема наших знаний о мире их классификация усложнялась. Эти классификации получили название иерархических. Многотомные схемы классификации конца прошлого – начала нашего века насчитывали десятки тысяч классов, подклассов, отдельных рубрик. Специалистам смежных областей знания и особенно массовому читателю библиотек стало трудно ориентироваться в схемах классификации и определять в их иерархии место той рубрики, по которой необходимо получить информацию.

Да и сами рубрики, строго ориентированные на узкие разделы наук, подвергающихся непрерывному процессу дифференциации, перестали удовлетворять специалистов-практиков, которым нужна была все более комплексная, предметная информация. Это привело к созданию в 70-х годах XIX в. предметной или точнее алфавитно-предметной классификации. На долгие годы она стала господствующей при составлении энциклопедий, вспомогательных указателей к трудам, систематически излагающим проблему или раздел науки, а в США, где она была создана, при организации каталогов.

Стремительный рост объемов литературы значительно усложнил также задачу идентификации каждого произведения печати. Биб-

лиотеки первыми столкнулись с необходимостью создать инструмент, при помощи которого можно было бы быстро и надежно устанавливать наличие определенного произведения в их фондах. Таким инструментом стал в XIX в. авторский, именной указатель (алфавитный каталог, по библиотечной терминологии), который однозначно идентифицировал произведение по именам лиц, принимавших участие в его создании или же связанных с его содержанием. Таким образом, до середины XX в. возможности содержательного поиска информации по справочникам или документам, содержащих нужную информацию, в библиотеках ограничивались тремя способами: систематическим, предметным и алфавитным.

Традиционной технологией реализации этих способов были списки, перечни книг и статей, содержавших необходимую информацию. С 70-х годов XIX в. эти сведения стали записываться на дискретных носителях – библиотечных карточках из плотного картона формата 75x125 мм (размер сложенной пополам американской почтовой карточки). Следует отдать должное этой традиционной технологии. Она успешно обеспечивала культурный прогресс на протяжении целого столетия вплоть до нынешнего этапа научно-технической революции, позволила накапливать и использовать многомиллионные собрания документов, обслуживать тематические потребности ученых и специалистов в необходимой им информации. На ней и сегодня еще в значительной степени зиждется деятельность всей мировой библиотечной системы – этого краеугольного камня человеческой культуры, важными составными частями которой является наука и техника.

Однако недостаточность, ограниченность этой технологии стала все более остро ощущаться уже в первой четверти XX в. В науке первыми почувствовали это химики из-за быстрого роста числа синтезируемых ими веществ. В настоящее время каждые три года появляется свыше миллиона таких веществ. Обычные методы оповещения – библиографические указатели, библиотечные каталоги, справочники типа «Гмелина» для неорганической химии и «Бельштейна» для органической – начали значительно отставать по времени от успехов исследователей и перестали охватывать их результаты в полном объеме. Революции в физике и электронике, характеризующие середину нашего столетия, усугубили трудности информационной коммуникации.

Процедуры и понятия

Научное сообщество осознало необходимость организационного оформления информационной деятельности, которая в течение нескольких десятилетий подспудно созревала в недрах науки и техники. Большая наука индустриального типа, пришедшая на смену "малой" науке университетского типа, выдвинула задачу создания систем научно-технической информации. Именно в это время, в конце 40-х – начале 50-х годов были сформулированы понятия информационного поиска, информационно-поисковой системы, информационно-поискового языка, была выдвинута задача механизации, а затем и автоматизации информационного поиска. Не случайно именно в это время В. Буш писал о необходимости новых форм справочных материалов, которые учитывали бы ассоциативные связи и были пригодны для механизации.

К этому времени стало ясно, что *информационный поиск – это совокупность логических процедур, в результате которых в ответ на информационный запрос выдается либо необходимая информация, либо документы, в которых она может содержаться, либо библиографические адреса этих документов.* В первом случае поиск получил название фактографического, во – втором документального, в третьем – библиографического. Эти процедуры сводятся к следующему.

Каждый вновь появляющийся документ подвергается анализу, в результате которого определяется его смысловое содержание. Этот анализ осуществляется интеллектом человека, возможность его формализации остается пока неясной. У автора документа и различных его читателей может быть разное представление о содержании документа. Затем это абстрактное представление о содержании (считается, что оно должно совпадать с авторским) выражается на некотором информационно-поисковом языке, т. е. синтезируется в виде библиографического описания и индекса.

Индекс образуется путем мысленного сопоставления основного смыслового содержания с потенциальными запросами потребителей информации. Эти запросы как бы зафиксированы в схемах классификации и обозначены индексами. Сама процедура выражения основного смыслового содержания документов и информационных запросов на информационно-поисковом языке получила название индексирования и составляет существенную часть аналитико-синтетической обработки документов. Информационный поиск, таким образом, заключается в замене содержательного прочтения полного текста документов фор-

мальным сличением (сравнением на соответствие) их поисковых образов с запросами на языке индексов.

Понятно, что такая замена значительно упрощает и убыстряет нахождение нужной информации, делает возможной автоматизацию процедуры сравнения. Но за это приходится платить неполнотой и неточностью поиска. Описанные выше логические процедуры допускают субъективизм осуществляющих их лиц, а используемые информационно-поисковые языки несовершенны и неспособны адекватно передавать содержание документов и смысл запросов. Следовательно, информационные потери и шум – неизбежные условия информационного поиска. Когда говорят, что поиск осуществлен со 100 %-ной полнотой, имеют в виду, что информационного поиска не производилось, а был осуществлен полный перебор всех текстов (современная технология в некоторых случаях предоставляет такую возможность).

Информационный поиск реализуется при помощи *информационно-поисковой системы*, которая в абстрактном виде должна состоять из информационно-поискового языка, правил перевода на этот язык и критерия смыслового соответствия, определяющего объем выдачи документов или информации (критерий выдачи). Конкретная система включает также средства реализации (перечень, картотека, механический селектор, компьютер), информационный массив и обслуживающий персонал.

Функционирование простейшей документальной информационно-поисковой системы можно проследить по ее блок-схеме на рис. 9. В системе имеется два входа (для документов и запросов) и один выход (для выдачи документов по запросам). На входах имеются преобразователи для индексирования документов и запросов. Поисковые образы документов вместе с адресами их хранения (номера) направляются в активное запоминающее устройство (ЗУакт), а сами документы – в пассивное (ЗУпас). Индексы каждого запроса сравниваются с индексами всех документов в решающем устройстве (РУ), которое в случае их соответствия (полного или предусмотренного критерием выдачи) дает в хранилище (ЗУпас) команду на выдачу документа. Это хранилище составляет как бы второй контур системы (сами документы), которого нет у библиографических (одноконтурных) систем.



Рис. 9. Блок-схема информационно-поисковой системы

Даже названия элементов на блок-схеме говорят о возможности автоматизации информационно-поисковой системы. Однако блок-схема верно обрисовывает работу любой системы, включая и наиболее традиционные. Это легко видеть на примере библиотеки. Преобразователи на входах соответствуют отделам обработки и справочно-библиографическому, ЗУакт – каталогам, ЗУпас – фондам. Нет в библиотеке только РУ – оно моделируется интеллектом читателя, который (хотя часто он и не осознает этого) вырабатывает собственный критерий выдачи и собственную стратегию поиска.

Не случайно именно эта интеллектуальная часть функционирования информационно-поисковой системы представила наибольшие трудности для автоматизации, именно она больше всего сдерживала развитие этих систем. Камнем преткновения явились, прежде всего, традиционные информационно-поисковые языки, ограничивающие возможности содержательного поиска информации. Расхожее мнение о том, что эти языки трудно поддаются автоматизации, неверно. Но они рассчитаны на ручную реализацию, и поэтому использование их в компьютерах удорожает поиск, ограничивает число пользователей и не дает никаких выигрышей, т. е. не снимает ограничений, присущих этим языкам.

А ограничения эти стали особенно ощутимыми на нынешнем этапе научно-технической революции. Прежде всего, традиционная технология поиска рассчитана на стабильный, медленно меняющийся состав запросов. В схемах классификации и перечнях предметных рубрик уже заранее как бы скоординированы все понятия, по которым можно извлекать информацию из документов и затем производить по ним поиск (такие языки поэтому и получили название *предкоординатных*). Это приводит к тому, что при возникновении новой проблемы или направления исследований, по которым имеется полученная прежде информация, система не обеспечивает ее поиска. Ведь эта тематика раньше не была сформулирована и не нашла места в схемах классификации и списках предметных рубрик, а значит и индексирование по ней не производилось.

Другими словами, традиционная технология поиска не позволяет искать информацию по любому, заранее не предвиденному сочетанию признаков. При этом, как уже говорилось, субъективизм индексатора при извлечении основного содержания документа увеличивает информационный шум и потери, предопределенные характером традиционных поисковых языков. Нельзя не отметить также, что основанные на них системы ручного поиска, даже фактографические, не пред-

назначены для манипулирования полученными из них данными. Они не имеют логического аппарата для содержательной переработки этих данных. Подобная задача всегда решалась самими потребителями без помощи информационных систем.

Координатное индексирование

Новая технология пришла в информационный поиск в виде метода координатного индексирования, разработанного в США в 50-е годы математическим логиком М. Таубе и работником службы химической информации К. Муэрсом. Этот метод основан на предположении, что основное смысловое содержание любого документа и информационного запроса можно выразить при помощи набора терминов, по большей части содержащихся в самом индексируемом документе. Эти термины получили название ключевых слов. Если, к примеру, нужно заиндексировать документ, в котором говорится о защите от коррозии лопаток газовых турбин, то совокупность терминов "турбина", "газ", "лопатки", "коррозия", "защита" и будет служить поисковым образом документа. Эти ключевые слова образуют для данного документа как бы координатную сетку, по которой в дальнейшем ведется информационный поиск по соответствующему запросу.

Преимущества данного метода очевидны. Прежде всего, информационные работники и потребители информации освобождаются от жестких рамок классификационных схем и перечней предметных рубрик. Индексирование новых документов ведется без оглядки на отраженные в них потенциальные и часто уже устаревшие запросы специалистов. С другой стороны, индексирование освобождается от субъективизма – ключевые слова выбираются формально. Эту работу, в принципе, можно поручить автомату. Во многих современных информационно-поисковых системах оператор вводит в машину библиографические данные документа, его реферат (аннотацию, резюме), а иногда и наиболее информативные части текста (например, первый и последний абзацы статьи, содержащие наибольшее число терминов, относящихся к ее содержанию). При помощи "запретительного" списка служебных и общезначимых слов, введенных в компьютер, осуществляется автоматический отбор ключевых слов, которые программно приводятся к нормальному виду (единственное число именительного падежа существительных и прилагательных, инфинитив глаголов). Это существенный шаг к автоматизации ввода информации в информационно-поисковую систему.

При поиске необходимой информации специалист может формулировать свой запрос в виде цепочки терминов, на пересечении которых и окажется большинство документов, содержащих необходимую информацию. При этом потребитель может произвольно менять стратегию поиска в зависимости от оценки его промежуточных результатов. Если документов по запросу мало или нет в системе, можно снять из запроса какие-либо ключевые слова (в приведенном выше примере "газ" и "защита"). Тогда система выдаст документы более широкого содержания о коррозии лопаток турбин, в которых все же может содержаться нужная информация. В случае, если документов по запросу слишком много, можно добавить ключевые слова, ограничивающие поиск, например, определенным классом турбин или же конкретными методами защиты их лопаток от коррозии. В этом уже заключен важный элемент возможности диалога с системой при помощи слов естественного языка.

Основные достоинства этого принципиально нового подхода к раскрытию содержания документов и поиску информации заключаются в том, что он позволяет находить информацию по любому, заранее не предвиденному сочетанию признаков. Кроме того, при появлении совершенно новых направлений исследований можно вести поиск во всем массиве документов, ранее заиндексированных по этому методу. Традиционные методы таких возможностей не предоставляли.

Было бы несправедливо умолчать о том, что достоинства нового метода приходится оплачивать преодолением дополнительных трудностей. Прежде всего, поиск с использованием естественного языка ограничивает его рамками знакомых пользователю языков. Чтобы расширить этот круг, приходится прибегать к словарям. Затем, каждый естественный язык отличается богатством своего словарного состава – слова, одинаковые по написанию, могут иметь разный смысл (многозначность, омонимия), а одно и то же понятие может выражаться разными терминами (синонимия). Запросив информацию о косах, вы получите сведения не только о сельскохозяйственных орудиях, но и о географических объектах, а может быть и о прическах. Желая получить документы о транзисторах, следует помнить, что они могут также называться полупроводниками.

Термины находятся в сложных взаимоотношениях между собой, выражают более узкие или более широкие понятия, могут быть связанными по сходству, по контрасту или по другим ассоциациям. Чтобы иметь возможность учитывать это при поиске, приходится составлять

на каждом языке специальные понятийные справочники (*тезаурусы*). В них для каждого понятия (класса условной эквивалентности) выбирается один термин – *дескриптор*, а для остальных слов указывается их связь с дескриптором. Тезаурусы иногда называют дескрипторными словарями, а сам поиск с их использованием дескрипторным. Кроме словарей для поиска по ключевым словам и дескрипторам часто создают специальную грамматику. Необходимость в ней вызывается возникновением ложной координации терминов, ошибочным их сочетанием. В ответ на запрос "трубы" x "медь" x "свинец" x "покрытие" можно получить не только необходимую информацию о покрытии медных труб свинцом, но и о покрытии свинцовых труб медью.

Курьеза ради, следует упомянуть, что метод координатного индексирования для поиска информации, явившийся принципиальным шагом к новой информационной технологии, на самом деле новшеством не был. В 1915 г. он был реализован на перфокартах американским орнитологом Т. Тейлором при составлении определителя птиц, а у Б. Веккери возникло предположение, что шумерские врачи еще в III тысячелетии до н. э. пользовались диагностическими устройствами, работавшими по этому принципу. На глиняных клинописных плитках записывались симптомы болезней, а под каждым из них – названия болезней, при которых эти симптомы встречаются. Совокупность симптомов составляла координатную сетку, а совпадающие для всех симптомов названия болезней – наиболее вероятные недуги больного.

Из этого понятно, что информационно-поисковые системы, основанные на принципе координатного индексирования, могут быть реализованы простейшими средствами ручного обращения. Система "унитерм-карт" самого М. Таубе представляла собой особым образом организованную картотеку, позволявшую легко сличать номера документов, чтобы выявить совпадающие номера для заданных терминов ("унитермов"). Первые информационно-поисковые системы такого типа часто создавались на просветных перфокартах. Однако подлинный размах создание координатных, по большей части дескрипторных систем получило, когда они стали использовать компьютеры второго поколения. В 60-е –70-е годы на базе крупнейших в мире реферативных служб были созданы мощные автоматизированные информационные системы, которые предназначались для ускорения выпуска информационных изданий и расширения спектра информационных услуг, а затем стали основными генераторами документальных баз данных на магнитных лентах.

Цитирование, библиографическое сочетание, социтирование

Принцип цитирования¹ был использован Институтом научной информации США, основанным в 1958 г. Ю. Гарфилдом, для создания принципиально нового вида информационного обслуживания. При поиске информации он взял в качестве индексов библиографические ссылки в документах. В выпускаемых им указателях цитированной литературы, называемых также "индексами цитирования", эти ссылки располагаются по алфавиту фамилий авторов цитированных работ с указанием сведений о документах, в которых они упоминаются. Произведения, использованные при написании статьи, составляют как бы координатную сетку для ее поиска. Если статья написана по совсем новой проблеме, не нашедшей рубрики в классификации наук, с еще не устоявшейся и малоизвестной терминологией, найти ее в потоке мировой литературы другими методами очень трудно. Указатель цитированной литературы можно представить себе как многоуровневую систему библиографических описаний документов, находящихся в обратной связи друг с другом.

Указатели цитированной литературы позволяют искать информацию по совершенно новым межотраслевым или комплексным проблемам под фамилиями пионеров и наиболее известных специалистов каждой из таких проблем. Например, для поиска литературы по цитированию достаточно знать фамилию Ю. Гарфилда, так как почти в каждой работе по этой проблеме есть упоминание о нем и его статьях. Фамилии авторов найденных работ могут в свою очередь служить входами в указатель, и за 1–5 таких итераций (последовательных поисков) все сведения об отраженной в указателе литературе по проблеме оказываются найденными.

Индексы цитирования предоставляют уникальную возможность проследить за всеми случаями применения какой-либо идеи или метода, за их критикой и обсуждением, оценить информационный вклад того или иного ученого или научной школы, степень и динамику популярности их работ. Известны случаи, когда по этим указателям

¹ Английское слово *citation* означает упоминание, ссылку и не соответствует русскому слову *цитирование*, означающему дословное повторение «чужого» текста. Однако в данном случае в качестве термина привилась прямая калька с английского языка, поскольку речь идет о библиографических ссылках. Для цитирования в русском значении употребляется английское слово *quotation*!

предсказывали нобелевских лауреатов. Институт научной информации США выпускает указатели цитированной литературы по точным, естественным и прикладным наукам (с 1964 г.), по общественным наукам (с 1969 г.), по искусству и гуманитарным наукам (с 1976 г.), для чего просматривается около 6 тыс. научных журналов и ежегодно до 1,5 тыс. названий книг. Большинство указателей распространяется не только в обычном (бумажном), но и в машиночитаемом виде (на магнитной ленте, дискетах, оптических дисках). Нужно ли говорить о том, что осуществление принципа, положенного в основу этих изданий, стало возможным лишь благодаря компьютерам. Ведь речь идет о ежегодном библиографировании почти 10 млн ссылок.

В двух статьях *E* и *F*, например, (см. рис. 10) имеются библиографические ссылки, которые устанавливают прямую библиографическую связь между ними (цитирующими документами) и статьями *A*, *B*, *C* и *D*, которые в них упоминаются (цитируемыми документами). В указателе цитированной литературы эти ссылки, по алфавиту которых упорядочивается его массив ("цитации" по терминологии Г. Я. Узи-левского, которому принадлежит данный пример), обозначают цитируемые документы, а под ними располагаются "библиограммы", т. е. описания цитирующих документов.

<i>A</i>	<i>C</i>
<i>E</i>	<i>E</i>
<i>B</i>	<i>F</i>
<i>E</i>	<i>D</i>
<i>F</i>	<i>F</i>

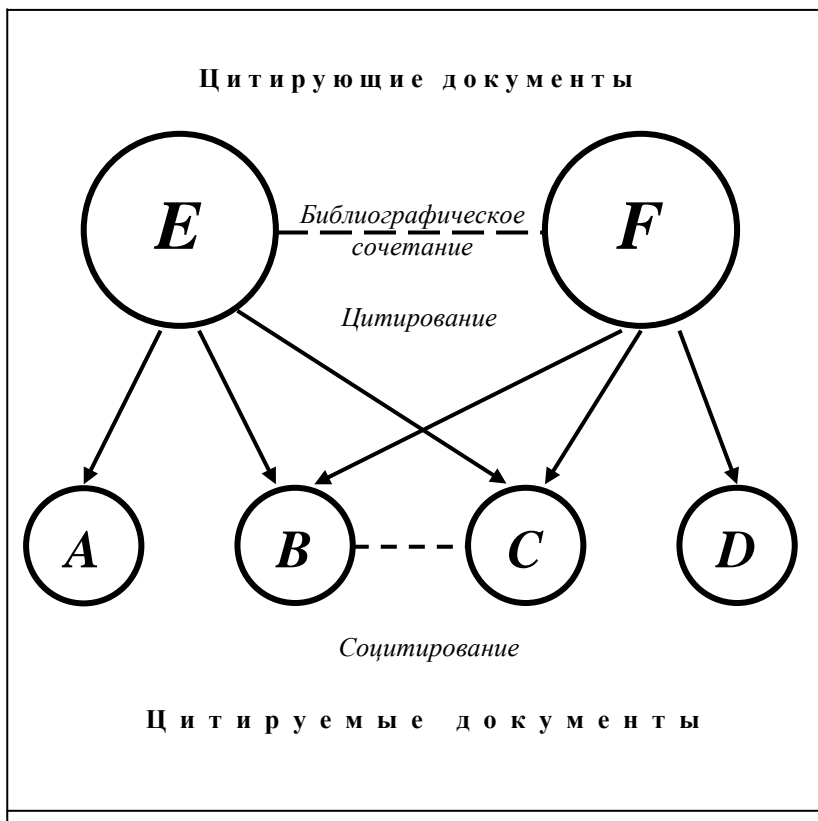


Рис. 10. Схема цитирования, социтирования и библиографического сочетания документов

Понимание потенциальных возможностей комплексирования документов по признаку общих ссылок и стремление максимально использовать накопленный массив в машиночитаемой форме повели к поискам новых путей применения метода цитирования. Еще в 1963 г. М. Кесслер в Массачусетском технологическом институте предложил считать связанными по смыслу документы, авторы которых ссылаются на одни и те же работы, а числом совпадающих ссылок измерять степень такой связанности. Этот метод, который он назвал *библиографическим сочетанием* документов, долгое время не имел широкого практического применения, но в 1968 г. Ю. Гарфилд использовал его для создания ретроспективной поисковой системы на компакт-дисках.

По-другому подошли к этой проблеме сотрудник Института научной информации США Г. Смолл и тогдашняя аспирантка ВИНТИ И. В. Маршакова. Они одновременно и независимо друг от друга в 1972 г. предложили считать связанными по смыслу и тематике работы, на которые совместно ссылаются авторы нескольких документов. Этот метод, чаще всего называемый *социтированием*, имеет другую коммуникационную основу. В каждой исследовательской области имеется некоторый набор важных работ, отражающих познавательную основу этой области. Данные работы цитируются многими исследователями и поэтому принадлежат к числу высоко цитируемых. Больше того, они часто цитируются вместе, образуя таким образом социтирование. Другими словами, социтированием принято называть одновременное упоминание любых двух или большего числа публикаций в какой-либо последующей.

Для лучшего представления разницы в этих методах выше приведена схема, на которой *E* и *F* являются цитирующими документами текущего года, а *A*, *B*, *C* и *D* – цитируемыми документами более ранних годов. Сплошными стрелками показаны связи по цитированию, т. е. *E* цитирует *A*, *B* и *C*, а *F* цитирует *B*, *C* и *D*. Тогда между цитирующими работами *E* и *F* образуется *библиографическое сочетание*, а между цитируемыми работами *B* и *C* – *социтирование* (обозначено пунктиром). Для простоты и наглядности степень связанности на схеме минимальная, хотя на практике она значительно больше (т. е. для признания библиографического сочетания между двумя работами или кластера социтирования в каждом отдельном случае устанавливается определенный минимальный порог, который тем выше, чем интенсивнее цитирование).

Между характером этих методов установления связи и областью их применения имеется существенная разница. *Библиографическое сочетание* – это однократно произошедший факт, поскольку *E* и *F* были однажды опубликованы со своими ссылками, и с ними в дальнейшем ничего уже произойти не может. Именно поэтому данный метод применяется для ретроспективного поиска документов, связанных между собой единством тематики, исследовавшей их авторами.

Совсем по-другому обстоит дело с *социтированием*, так как связь между цитируемыми *B* и *C* может сохраняться (увеличиваться или уменьшаться) в последующие годы в зависимости от того, насколько часто они будут попарно цитироваться в новых работах. Частое социтирование указывает на их концептуальную близость, поскольку они используются как единый комплекс. Между этими работами как бы возникают невидимые связи, которые после наглядного их выражения образуют смысловые сгустки (кластеры). Совокупность таких кластеров ключевых работ, отражающих исследовательские области, представляют собой как бы карту определенной научной области, а совокупность карт – атлас науки на данный момент.

При регулярном выпуске подобных атласов (выходили атласы по биологии, биохимии, геологии, математике, вычислительной технике) появляется возможность регулярно следить за динамикой развития научных дисциплин, школ, направлений, коллективов, а, следовательно, и целенаправленно воздействовать на это развитие, т. е. управлять им. Методы библиографического сочетания и кластеризации социтирования моделируют содержательные отношения между документами, используя практику цитирования, сложившуюся при публикации научных работ. Но это не единственный возможный подход к установлению таких связей между документами, заложенных в их библиографических элементах и фрагментах текстов. В следующей лекции об информационных системах вы познакомитесь с методом логико-смыслового моделирования, а теперь мы перейдем к более традиционным видам информационно-поисковых языков.

Иерархические и фасетные классификации

Было бы неверно думать, что будущее только за цитированием, за информационно-поисковыми языками координатного индексирования, которые вытеснят традиционные языки. В этом случае, как и во всей системе коммуникации, действует закон развития, по которому новые средства не заменяют полностью прежних, а лишь перераспре-

деляют функции между ними. Это в полной мере относится и к такому древнему средству информационной технологии, как иерархические классификации, наиболее распространенной представительницей которых выступает Универсальная десятичная классификация (УДК). Создание УДК явилось переломным моментом в развитии ИПЯ: она завершила тысячелетнюю историю линейных классификаций перечислительного типа и открыла пути к построению фасетных классификаций.

Библиотечные классификации – самый ранний из известных нам типов ИПЯ. Если проследить за их развитием от классификации вавилонских библиотек вплоть до библиотечных классификаций второй половины XIX в., становится ясно, что принципы их построения почти не изменялись. Следуя за наиболее известными системами классификации наук, библиотечные классификации строились на основе иерархического "древа знаний" с выделением специальных разделов и подразделов для систематизации особых видов книг. До нас дошло очень немного сведений о классификациях, применявшихся в древних библиотеках. Вероятно, в античные времена такие классификации были разработаны достаточно детально. Об этом можно судить по обширности библиотек, высокому уровню классификации наук того времени, а также по некоторым косвенным свидетельствам, содержащимся в литературных памятниках.

На классификации средневековья известное влияние оказала распространенная в то время система "семи свободных искусств". Она состояла из двух комплексов наук, изучавшихся в тогдашней школе: "тривиума" (грамматики, диалектики и риторики) и "квадривиума" (арифметики, геометрии, музыки и астрономии). К концу XV в. в университетских библиотеках начали применять группировку книг по содержанию в соответствии с существовавшими в большинстве университетов четырем факультетам: философским, медицинским, юридическим и богословским. Это послужило толчком к возникновению так называемых факультетских систем классификации, пользовавшихся популярностью на протяжении столетий вплоть до XIX в. Они оказали влияние на выдающиеся для своего времени классификации швейцарского ученого и библиографа К. Геснера (1548) и немецкого философа Г. Лейбница (1700).

Дальнейшее развитие библиотечно-библиографических классификаций проходило под воздействием идей английского философа Ф. Бэкона (1561-1626). Созданная им в начале XVII в. классификация наук группировала знания в соответствии с идеалистической традици-

ей по "способностям человеческого духа". "Память" определяла возникновение истории, "воображение" - поэзии, "разум" - философии, или собственно науки. При всей условности такого деления классификация Ф. Бэкона включала новые отрасли знания и представляла собой значительное событие для науки того времени.

В XIX в. библиотечно-библиографические классификации стали широко разрабатываться в России. Большой интерес представляет схема ученого-натуралиста П. Г. Демидова, составленная им для каталога личной библиотеки. В этой схеме наиболее детализированными были разделы естественных наук и технологии. Заслуживает внимания схема, опубликованная в 1809 г. А. Н. Олениным и предназначенная для императорской СПб Публичной библиотеки, а также оригинальная схема К. Ф. Рейса, предложенная им для библиотеки Московского университета в 1826 г. и основанная на дихотомическом принципе деления. Самостоятельную схему классификации, в основу которой было положено условное деление наук "по потребностям человека", создал для библиотеки Казанского университета К. К. Фойгт в 1843 г.

Одной из лучших в первой половине XIX в. заслуженно считается классификация выдающегося русского натуралиста К. Э. Бэра, которую он разработал для иностранного отделения библиотеки Академии наук в Санкт-Петербурге в 1841 г. В этой классификации была предпринята одна из первых попыток расположить науки в последовательности, отражающей историю развития мира: науки о неорганической природе, науки об органической природе, науки о человеке и обществе.

В библиотеках Западной Европы в этот период особенно широко применялись схемы классификации Ж. Ш. Брюне (1810) и А. Э. Шлейермахера (1847). Первая из них представляла собой одну из поздних разновидностей так называемой французской системы, которая в течение почти двух столетий использовалась в библиографии и книготорговле Франции. Вторая, особенно популярная в библиотеках немецких университетов, была создана на основе старой "факультетской системы". Эти классификации создавались для расположения книг на полках и для систематизации их описаний в каталогах и указателях. Такое их назначение обусловило необходимость линейной последовательности их рубрик и строгого подчинения между классами и подклассами, всегда связанными в этих классификациях родовидовыми отношениями. Одна из наиболее сильных сторон этих языков заключается в том, что классификация по родовидовым признакам все-

гда служила важным инструментом познания и привычным методом определения понятий.

Иерархические классификации обеспечивают высокую эффективность информационного поиска по широким тематическим запросам, сформулированным в определенном аспекте, который был предусмотрен заранее, при составлении схемы и при индексировании по ней документов. Для реализации подобного поиска наиболее оптимальным техническим средством служит просто перечень (каталог). Эти особенности объясняют прочное положение иерархических классификаций как единственного на протяжении тысячелетий средства поиска документов по их содержанию. Лишь во второй половине XIX в. появилась необходимость в другом типе языка, который упрощал бы для массового читателя разыскание нужных ему рубрик и облегчал бы введение в систему новых понятий без коренной ее перестройки. Таким языком стала алфавитно-предметная классификация, теорию которой в 70–80-х годах прошлого века разработал Ч. Э. Кеттер.

Крупнейшим достижением в области систематизации явилось создание в 1876 г. видным американским библиотечным деятелем М. Дьюи "десятичной классификации". Сам он видел свою основную заслугу в том, что применил в своей схеме десятичную индексацию: "Дело шло о достижении абсолютной простоты путем использования самых простых и известных символов, арабских цифр в виде десятичных дробей в качестве индексов классификации всех человеческих знаний в печатных произведениях"¹. Однако теперь, по прошествии столетия, значение созданного М. Дьюи нам представляется в другом. Он теоретически обосновал и практически внедрил стандартизацию типовых делений (литературной формы, вида издания и т. п.) в различных разделах схемы и частично лингвистических, этнических и географических делений, использовав прием факультативного превращения в постоянные подразделения окончаний индексов разделов "Филологии" и "Истории".

Другой его важной заслугой было введение в систему алфавитно-предметного указателя, который позволил разыскивать книги по любому предмету независимо от его места в схеме. Идея такого вспомогательного указателя, правда, не была новостью, такие указатели уже применялись в энциклопедиях XIII в., в изданиях эпохи Воз-

¹ Шамурин Е. И. Очерки по истории библиотечно-библиографических классификаций. – Т. 2. – М., 1959. – С. 500.

рождения, у К. Геснера, в таблицах А. Э. Шлейермахера. Однако только теперь подобные указатели стали неотъемлемой частью классификационных таблиц и систематических каталогов библиотек. Таким образом, впервые была предпринята попытка избавиться от жесткости линейной схемы иерархической классификации и расширить число входов в ее схему. Однако этот шаг был еще очень робким, и принцип проводился не очень последовательно.

Эта непоследовательность была преодолена в "брюссельском варианте" десятичной классификации, которая получила широкое распространение после 1905–1907 гг. как "Универсальная десятичная классификация". Ее создатели выдающиеся бельгийские документалисты П. Отле и А. Лафонтен использовали преимущества десятичной системы, индексация которой понятна людям, говорящим на разных языках, и развили заложенные в ней возможности более гибкого использования иерархической классификации. Идею стандартизации делений схемы они довели до логического конца, создав вспомогательные таблицы типовых делений – общих и специальных (аналитических) определителей. Общие определители (языка, формы документа, места, времени, народности, точки зрения) используются во всех отделах схемы с одним и тем же значением. Специальные определители предназначены для использования только в нескольких отделах одной отрасли знания для их деления по одним и тем же признакам.

Важным достижением УДК явилось также введение принципа комбинации индексов, разработка приемов их присоединения, распространения, отношения и объединения (синтеза), кроме того, было предусмотрено использование индексов подразделений одного раздела в других и введение параллельных (альтернативных) делений для отражения классифицируемых объектов в разных аспектах. Таким образом, на смену прежним "перечислительным" схемам с заранее установленными рубриками и готовыми индексами пришла подвижная схема, в которой нужные рубрики могут создаваться в процессе классификации путем сочетания индексов с определителями или соединения их друг с другом.

К основным достоинствам УДК как иерархической классификации относятся следующие ее характерные черты:

- универсальность, заключающаяся в охвате всех отраслей знания,
- логическая ступенчатая индексация, позволяющая неограниченно делить подклассы без нарушения основной структуры классификации,

- международная применимость благодаря использованию только цифровых десятичных индексов, всем понятных и легко запоминаемых,

- развитая система определителей и комбинационного построения индексов, обеспечивающих относительную гибкость при отражении достаточно узких и сложных понятий,

- устойчивый и четко организованный международный механизм поддержания классификации на уровне новых достижений науки.

В качестве недостатков УДК часто называют ее естественные ограничения, присущие всем иерархическим классификациям. Они не могут удовлетворительно отражать процессы интеграции и взаимопроникновения наук, и в них трудно находить место для направлений и понятий, возникающих на стыке наук. Далеко не все явления в природе и понятия в науке можно связать родовидовыми отношениями. Это особенно ярко проявляется в технике, медицине, в других прикладных, а также в комплексных дисциплинах, таких, например, как кибернетика, информатика, семиотика. Эта ограниченность УДК, в которой отдельные науки жестко разделены в соответствии с формальными логическими правилами, противоречит тенденции синтетического развития науки. Нельзя сказать, чтобы это были недостатки, скорее это внутренние свойства иерархических классификаций, обеспечивающие их эффективность в условиях широкого тематического поиска.

Следующий шаг в развитии комбинационного принципа в классификации был сделан выдающимся индийским библиотековедом Ш. Р. Ранганатаном в созданной им в 1933 г. "Классификации с двоеточием", которая явилась родоначальницей *фасетных* классификаций. Об их функциях английский информатик Б. Виккери писал: "Потребители хотят иметь возможность отыскать документ, посвященный сложной специальной теме, не только тогда, когда именно она является непосредственным объектом поиска, но также тогда, когда поиск ведется по любому термину или группе терминов, входящих в сложное понятие. Для удовлетворения этих требований необходимо, чтобы не только понятия могли входить в неограниченное количество сочетаний, но также, чтобы в структуре системы были отражены родовые связи понятий и связи между разделами"¹.

Фасетная классификация вместо единого ряда делений в каждом основном классе имеет несколько "фасетов", соответствующих ас-

¹ Фасетная классификация. – М.: ГПНТБ, 1970. – С. 8.

пектам классифицируемого понятия или предмета. Все существенные термины данного класса распределяются по фасетам и образуют их "фокусы". При индексировании документов их содержание выражается цепочкой фокусов, последовательность которых определяется специальной "фасетной формулой". Примером построения индексов по системе Ш. Ранганатана могут служить следующие фасеты и фокусы из области медицины:

Ф А С Е Т Ы		
Органы тела	Проблемы медицины	Уход и лечение
1 Органы в целом	1 Общие проблемы	1 Питание
2 Органы пищеварения	2 Морфология	2 Этиология
23 Пищевод	3 Физиология	3 Диагностика
24 Желудок	4 Болезни	4 Лечение
25 Кишечник	42 Инфекционные	
3 Кровеносная система	421 Туберкулез	
4 Органы дыхания		
45 Легкие		

В этой схеме индекс документа по диагностике инфекционных заболеваний кишечника – 25:42:3, по лечению туберкулеза легких – 45:421:4.

Преимущества этого вида классификаций в том, что они облегчают многоаспектное индексирование документов, позволяя собирать в одном месте все аспекты рассмотрения какого-либо предмета или темы, они легче поддаются изменениям при введении новых понятий, допускают большую глубину индексирования при более коротких индексах. Их применение особенно эффективно при поиске в небольших по объему узкоспециализированных собраниях документов. Видный английский информатик Д. Фоскетт так обосновал достоинства фасетных классификаций: "От схемы не требуется более, чтобы она указывала "место" для каждого документа, включая любой термин или набор терминов в явном виде в классификационные таблицы по каждой предметной области. Эти схемы могут задать набор правил, или рабочих процедур, с помощью которых такие контексты можно, по мере надобности, формулировать на основе тех же самых схем"¹.

¹ Теоретические проблемы, информатики. – М.: ВИНТИ, 1968. – С. 67.

Рубрикаторы информационных изданий

На большинстве европейских языков рубрикатором называли переписчика рукописей, который в скрипториях средневековья и Возрождения размечал красной краской первые буквы смысловых фрагментов текста, получивших название рубрик. Это название сохранилось и до наших дней, хотя в нынешних произведениях печати рубрики отмечаются абзацными отступами или отделяются друг от друга пробелами. В журналистике рубриками принято также называть постоянные разделы в журналах и газетах, а в библиотековедении – структурные подразделения систематического и предметного каталогов.

В 50-е годы в информатике рубрикаторами стали называть перечни рубрик реферативных журналов и других информационных изданий. В данном случае рубрика выступает как содержательный фрагмент такого издания и состоит из индекса и заголовка раздела, а также библиографических записей (с аннотациями или рефератами) произведений печати, которые по своему содержанию относятся к данной рубрике. По мере роста числа и увеличения объемов реферативных журналов их рубрикация стала усложняться. Появилась необходимость в создании такого перечня рубрик, который отвечал бы определенным требованиям и мог бы служить средством систематизации библиографических записей вместе с рефератами. Поскольку библиотечно-библиографические классификации оказались непригодными для этого, реферативные службы стали создавать собственные рубрикаторы.

Рубрикатор – это особым образом организованный перечень рубрик иерархической классификации, предназначенный для отражения сведений о текущих публикациях в информационных изданиях или системах информационного обслуживания. К его характерным особенностям относятся сравнительно небольшая глубина индексации, ориентированность на межотраслевые, междисциплинарные, комплексные проблемы, простота и линейность структуры, достаточная гибкость, частая и безболезненная изменяемость формулировки рубрик. Любой рубрикатор создается под влиянием двух противоречивых факторов, отражает два взаимосвязанных, но разных информационных потока: документального и запросов потребителей. Первый оказывает преимущественное влияние на структуру рубрикатора, второй – на формулировку заголовков рубрик, причем изменение структуры документального потока несколько отстает от быстро меняющегося харак-

тера информационных запросов.

Возникает вопрос, почему же все-таки для создания рубрикаторов не использовались существующие классификации? Можно указать на несколько обстоятельств, которые ведут к серьезным различиям в схемах иерархических классификаций, используемых для библиотечных каталогов и для построения рубрикаторов. Первые, рассчитанные, в первую очередь, на систематические каталоги и картотеки библиотек, отражают структуру универсального потока документов: книг, брошюр, периодических и продолжающихся изданий. Рубрикаторы реферативных журналов ориентированы преимущественно на журнальные статьи и другие публикации из научной периодики, которые имеют другую содержательную структуру, более дробную и гибкую. В реферативных журналах подчас приходится открывать рубрики для таких вопросов, которые в библиотечном каталоге могут стать необходимыми лишь через десять лет.

Систематические каталоги библиотек ориентированы на дисциплинарную структуру, т. е. на выделение основных классов в соответствии с научной классификацией. В рубрикаторе наряду с дисциплинарными характеристиками необходимо учитывать комплексные междисциплинарные проблемы и отрасли народного хозяйства. Это нарушает строгую логику иерархической классификации, но придает рубрикатору особую гибкость. Библиотечная классификация предназначена для ретроспективного поиска, для накопления записей за много лет, это требует сложной структуры, ее стабильности, устойчивости, медлительности в изменениях. Для рубрикатора частые изменения являются правилом, формулировка заголовков рубрик, публикуемых в каждом номере издания, играет сравнительно большую роль, а форма индексов, выполняющих служебную роль, менее значима. Рубрикатор легко обозрим, имеет небольшую глубину и простой служебный аппарат (систему ссылок и вспомогательных делений, способы сочетания рубрик).

По рубрикаторам классифицируются самые мощные потоки научных публикаций – во всем мире ежегодно не менее 5 млн несопадающих документов (из них только в ВИНТИ около 1 млн). Если ориентировочное число публикуемых ежегодно научных документов принять близким к 10 млн, то половина из них систематизируется по различным рубрикаторам. Это на порядок больше, чем приходится на долю классификаций, применяемых ежегодно для описания входных потоков всеми библиотеками мира. Поэтому рубрикаторы приобрели

большое значение в научно-информационной деятельности. Во многих информационных центрах избирательное распространение информации, сигнальная информация и даже справочно-библиографическое обслуживание осуществляются при помощи рубрикаторов. Чтобы они могли справиться с такими несвойственными им функциями, приходится оснащать их различными вспомогательными средствами, которые приближают их к библиотечным классификациям, но затрудняют их использование по прямому назначению. Как и во всех подобных случаях, здесь приходится прибегать к разного рода компромиссам, но это неизбежно там, где мы не пользуемся новой информационной технологией.

Разные типы информационно-поисковых языков

Информатика, заявившая о себе в середине нашего века, принесла с собой не только новую и получившую распространение терминологию ("дескрипторы", "тезаурусы", "индексирование") и не только удовлетворила нашу обычную потребность в противопоставлении нового традиционному. Новым, действительно новым, оказался более широкий подход к явлениям и принципам. Понятие, например, информационно-поискового языка (ИПЯ) позволило рассматривать предметизацию, систематизацию, книгописание, координатное индексирование как процессы, использующие искусственные языки, семантическую силу которых можно измерять по сравнению с возможностями естественного языка. Понятие информационно-поисковой системы (ИПС), как уже говорилось, объединило многие предметы, которые прежде рассматривались изолированно, например, библиотечные фонды и каталоги, различного вида самостоятельные и вспомогательные указатели, справочники, энциклопедии, автоматизированные поисковые системы. Это дало возможность выявить общие принципы их построения, найти общие критерии их эффективности и другие общие параметры.

Мы установили, что дескрипторные информационно-поисковые системы открыли принципиально новую возможность поиска необходимых документов и содержащейся в них информации по любому сочетанию заранее не предвиденных признаков. Однако за реализацию этой возможности приходится платить не только интеллектуальными потерями, но и материальными ресурсами. Эти системы приходится ориентировать на дорогостоящие компьютеры и программы, что предполагает более трудоемкий ввод информации и более строгие ограни-

чения на число одновременных пользователей. Вполне естественно в такой ситуации попытаться сочетать уже имеющиеся поисковые средства со вновь создаваемыми. Отсюда вытекает и желание найти общие черты в этих разных системах и лежащих в их основе ИПЯ: языке предметных рубрик и дескрипторном языке, что обычно сочетается с поисками путей их совместимости.

Всегда можно найти такую удаленную позицию, такое основание деления, при которых эти языки попадут в один общий класс. В ряду искусственных языков они принадлежат к классу информационных, в ряду информационных – к подклассу информационно-поисковых. В них используются в качестве индексов слова естественного языка. При построении этих языков применяются внешне схожие приемы: перечень предметных заголовков и словарная часть тезауруса упорядочиваются в алфавите слов. Тем не менее, учитывая эти общие и сходные черты, нельзя забывать и о принципиальных различиях данных языков. Основной словарный состав языка предметных рубрик это имена сложных классов, построенных до индексирования документов, поскольку этот язык принадлежит к типу *предкоординируемых*. Дескрипторный же язык является *посткоординируемым*, т. е. строится из имен простых классов, которые образуют необходимые понятия при их пересечении (логическом умножении) в момент индексирования и/или поиска документов.

Для того, например, чтобы индексировать статью о *производстве и экспорте вычислительных и пишущих машин в США, Японии и Великобритании* достаточно дескрипторов ПРОИЗВОДСТВО, ЭКСПОРТ, КОМПЬЮТЕР, ПИШУЩАЯ МАШИНА и названий трех этих стран. Тогда при любой комбинации признаков при запросе (а таких комбинаций может быть $1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 = 5040$, т. е. число перестановок из семи признаков) этот документ будет найден. Если же пользоваться языком предметных заголовков, то в зависимости от их заранее составленного перечня потребуется значительное число готовых рубрик. Оно, конечно, меньше указанного выше, но ровно настолько будет больше потерь при поиске.

Координатное индексирование в том и состоит, что для характеристики содержания документа или запроса перечисляются такие ключевые слова или дескрипторы, пересечение (логическое умножение) которых выражает основное смысловое содержание (главную тему, предмет) этого документа или запроса, тогда как в предметизации для

данной цели используются заранее сформулированные заголовки и подзаголовки.

При индексировании, т. е. выражении основного смыслового содержания документа в терминах ИПЯ, процессы информационного анализа и синтеза совершаются в два этапа. Первый этап является общим для всех языков. Содержание документа анализируется как с позиций того, какие идеи и факты заложены в него автором, так и с позиций научных и практических интересов большинства его потенциальных читателей. (Если не иметь в виду узкоспециальных интересов, то обе точки зрения чаще всего совпадают). Результаты этого анализа синтезируются в виде субъективного представления индексатора об основном содержании документа.

Второй этап зависит от языка индексирования. Если это *предкоординированный* алфавитно-предметный язык, то свое представление о содержании документа индексатор сверяет с потенциальными запросами читателей, отраженными в перечне предметных заголовков. Для *посткоординируемого* дескрипторного языка аналогичному анализу подвергается тезаурус (не связанный непосредственно с потенциальными запросами потребителей) и сам текст индексированного документа. Синтез в данном случае выражается в выборе соответствующих предметных заголовков или дескрипторов (ключевых слов).

Другими словами, при всей внешней схожести процедур индексирования посредством этих разных типов ИПЯ, характер их использования различен. В одном случае мы пользуемся готовыми продуктами в виде заголовков и подзаголовков, обозначающих класс документов определенного содержания. В другом случае это лишь исходный материал, дескрипторы и ключевые слова, при перемножении которых образуется класс, соответствующий данному содержанию. Вот почему перечень предметных заголовков и словарная часть тезауруса, при всей их внешней схожести, при том, что определенная часть слов в них может совпадать, на самом деле являются совершенно отличными друг от друга списками, слова для которых отбираются на основе разных критериев и играют различную роль.

Разные типы ИПЯ имеют свои достоинства и ограничения, которые делают их особо пригодными для решения разных поисковых задач. Возможности дескрипторного языка эффективно реализуются при узко тематическом поиске по произвольной комбинации признаков. Широкий тематический поиск по традиционным отраслям знаний и поиск по конкретным предметам, дисциплинам и их разделам в фон-

дах документов за многие годы и в условиях одновременного обращения к ним большого числа читателей по-прежнему хорошо обеспечиваются библиотечными каталогами, основанными на иерархических и алфавитно-предметных классификациях. Выпуск информационных изданий требует разработки специальных рубрикаторов с небольшим числом уровней иерархии и подвижной, быстро меняющейся рубрикой.

Базы и банки данных

С самого начала в теории информационного поиска предполагалась возможность построения не только информационно-поисковых, но и информационно-логических систем, которые осуществляли бы автоматическую переработку информации, а также извлечение из научных текстов неявно содержащейся в ней информации. Эту идею много лет назад высказал один из пионеров этой теории в нашей стране В. А. Успенский¹. В то время такая возможность связывалась с дальнейшим совершенствованием электронной вычислительной техники, главным образом, с увеличением емкости оперативной памяти компьютеров и их быстродействием, что было вполне понятно. Но подобный ход мысли характерен и в наше время для специалистов в области вычислительной техники. Недаром девизом пятого поколения вычислительных машин служил переход от переработки данных и информации к переработке знаний.

Информатики же давно поняли, что переработка знаний связана не только и даже не столько с совершенствованием компьютеров, сколько с организацией самих этих знаний. В этой сфере до последнего времени господствовали представления, связанные с традиционной структурой научного знания, которое фиксируется в статьях и монографиях и отражается в библиографических бюллетенях и реферативных журналах. Но задача заключается в том, чтобы понять внутреннюю структуру знаний, взаимосвязь данных, фактов, гипотез и теорий.

За последние десятилетия значительное развитие получили исследования в области неклассических логик, баз данных и баз знаний, формализованного представления содержания текстов. Они опираются на достижения математической логики, логической семантики, структурной лингвистики и ряда других фундаментальных и прикладных

¹ Успенский В. А. К проблеме построения машинного языка для информационной машины // Проблемы кибернетики – 1959. – Вып. 2. – С. 39–50.

дисциплин, Результаты этих теоретических исследований находят все большее применение в автоматизации информационных процессов и построении информационных систем различных типов, которые рассматриваются как средство переработки данных и знаний.

В среде информационных работников стало привычным говорить о базах и банках данных (БД и БнД). Само по себе это свидетельствует о том, что профессионалы эффективно используют компьютеры в информационной деятельности. Хуже то, что смысл и значение этих терминов, пришедших из области программирования, понимаются недостаточно глубоко. Это напоминает библиотекарей, которые с появлением теории информационного поиска стали называть каталоги информационно-поисковыми системами, что, в общем-то, верно, но ничего не изменило в традиционной организации каталогов. Так и теперь информационные издания (бюллетени сигнальной информации, реферативные журналы) на магнитной ленте или дискете любят называть базами данных, что тоже не совсем неправильно, но не отражает принципиального смысла этого понятия.

Понятие базы (или банка) данных возникло в конце 60-х гг. в связи с необходимостью освободить программы от данных, которыми они оперируют, сделать их независимыми. До этих пор данные организовывались под нужды каждой конкретной программы, что создавало массу неудобств и затруднений, особенно при изменении данных или самих программ. "База данных это реализованная средствами вычислительной техники специальная система для хранения данных о некотором фрагменте действительности. Главные идеи, лежащие в основе такой системы, – это объединение в одном месте данных, нужных для решения многих задач (может быть, даже тех, которые еще не начинали программироваться) и обеспечение независимости данных от их обработки"¹. В процессе развития этого понятия базой данных стала называться и сама совокупность данных, организованных по определенным правилам их описания, хранения и манипулирования ими независимо от прикладных программ.

Появление баз и банков данных оказалось существенным шагом, приблизившим возможность автоматического решения многих интеллектуальных задач. Некоторые специалисты даже сравнивают это достижение с изобретением книгопечатания. Информационные работники со временем стали различать термины "база" и "банк" данных,

¹ Борщев В. Б. Банки и базы данных // Природа. – 1982. – № 3. – С. 64.

которые первоначально употреблялись как синонимы. Теперь они понимают под банком данных систему программных, языковых, организационных и технических средств, предназначенных для централизованного накопления и коллективного использования данных, а также сами данные, хранимые в виде баз данных. Существует и более ограниченное понимание БД как одних только программных средств: баз данных, их справочника, системы управления ими (СУБД) и библиотеки запросов и прикладных программ.

Для автоматизированных информационных систем создание банков и баз данных открывает возможность осуществлять информационный поиск не только документов, но и заключенных в них фрагментов – идей и фактов, а также манипулировать ими. Появляется реальная перспектива обогащать собственные наблюдения и результаты исследований ученых всем мировым опытом науки, содержащимся в научно-технической литературе. Значительная часть трудоемкой работы по извлечению и упорядочиванию имеющихся в документах данных, производимая прежде каждым исследователем, в принципе может быть переложена на информационную систему. В концепции банка данных реализуется давняя мечта информационных работников о создании не только автоматизированной информационно-поисковой системы, но и информационно-логической системы, позволяющей осуществлять анализ и синтез научной информации.

Ученый-исследователь и экспериментатор, инженер-разработчик и проектировщик получают возможность оперировать большим, чем до сих пор количеством данных, быстро меняя их организацию. Это можно проиллюстрировать на примере научной работы врачей. Целенаправленное наблюдение и лечение больных, страдающих определенным недугом, получает отражение в историях их болезни. Обычно несколько десятков историй болезни, сопоставленных с данными нескольких десятков литературных источников, служат материалом для написания статьи, и, в конечном счете, кандидатской диссертации. На эту работу уходят годы труда. Несколько сотен историй болезни и литературных источников могут привести к созданию монографии и докторской диссертации, на что приходилось затрачивать значительную часть жизни.

Непосредственный доступ к банку данных позволяет выполнить существенную часть этой работы значительно быстрее. Банк данных облегчает перестройку всевозможных сведений, приведение их к необходимому единообразию, получение статистической информации, по-

иск зависимостей между параллельными рядами данных. Другими словами, работа с базами данных на компьютере в оперативном режиме устраняет противопоставление поиска информации ее творческой переработке, стирает грань между ними. Происходит как бы диалектический возврат к слиянию этих процессов в деятельности ученого на новом витке развития по спирали.

Было бы неверно думать, что информационные системы уже сегодня готовы к повсеместному переходу на описанный режим работы, хотя и в нашей стране и особенно за рубежом создаются и функционируют автоматизированные системы такого рода. Прежде всего, теория и практика баз и банков данных еще очень молоды и быстро развиваются. Даже наиболее распространенная реляционная ее модель, имеющая ряд преимуществ перед иерархической и сетевой организацией банков данных, как математическая структура реальной действительности далека от совершенства. Именно поэтому разрабатываемые в математической логике модели информационных систем пользуются в информатике таким вниманием. Многообразные задачи, поставленные перед новыми типами информационных систем, называемых теперь интеллектуальными, требуют адекватных средств реализации.

Информационные системы

Информационно-поисковые системы

За последние десятилетия в области информационного поиска произошли важные изменения, которые требуют принципиально новых подходов и решений. В крупных банках данных накоплено огромное количество источников информации. Насчитываются десятки тысяч общедоступных базы данных, в которых содержатся миллиарды записей. В них ежегодно проводятся десятки миллионов информационных поисков. Почти на каждый из них информационно-поисковые системы теперь быстро выдают сотни и тысячи источников, т. е. слишком много, чтобы потребитель был в состоянии все прочитать или хотя бы просмотреть. Поэтому потребитель вынужден ужесточать свои требования к выдаваемым источникам. Возникла возрастающая потребность в значительном усилении критериев выдачи релевантных документов, более строгим их отборе, для чего необходима разработка и применение других принципов, переход на моделирование – пусть вначале даже самое грубое – процессов информационного поиска в памяти человека. Ясно, что поиск информации в памяти человека осуществляется совершенно иначе, чем посредством современных информационно-поисковых систем.

Другой важный фактор, все сильнее воздействующий на сферу информационного поиска, заключается в появлении и росте численности полнотекстовых БД, представляющих собой электронные аналоги печатных изданий и документов – энциклопедий, словарей, справочников, книг, журналов и т. п. Это обусловлено расширяющимся внедрением современных средств вычислительной техники в издательское дело и полиграфию.

Опыт информационного поиска в полнотекстовых БД, когда поиск проводится не по поисковым образам документов, а по их полным текстам, показывает, что использование ключевых слов, встречающихся в полных текстах, не дает тех результатов, которые получаются в

первом случае. Хотя использование полных текстов документов создаст новые, дополнительные возможности для повышения эффективности поиска (например, благодаря использованию библиографических ссылок как поисковых признаков), стало ясно, что необходима разработка существенно новых принципов информационного поиска по полным текстам документов, основанных на результатах исследования механизмов человеческого мышления и на использовании баз знаний и опыта, накопленного при разработке и эксплуатации экспертных систем, систем машинного перевода и других интеллектуальных информационных систем.

Проблема информационного поиска весьма существенно усложнилась еще в одном отношении. Если раньше объектом поиска были источники информации, то теперь возрастает потребность в поиске самой информации, которая при этом не всегда имеется в поисковом массиве в явном виде. Для получения такой информации необходимо выполнять над текстами источников или записями фактов те или иные логические операции, требующие привлечения определенных знаний о мире, которые непосредственно не содержатся в этих текстах и записях. Такого рода знания отображаются в *базах знаний*, простейшими примерами которых могут служить информационно-поисковые тезаурусы и таблицы классификации. Кроме того, необходимо учить компьютеры «понимать» тексты и факты, оценивать их, рассуждать, делать логические выводы, формулировать гипотезы, т. е. выполнять многие интеллектуальные операции, присущие человеческому мозгу. А это становится все более возможным по мере того, как мы узнаем все больше о механизмах человеческого мышления и учимся их моделировать. Таким образом, возрастает потребность в глубоком изучении процессов языкового и логического мышления, в создании и использовании информационно-логических, или *интеллектуальных* систем, которые сами становятся мощнейшим инструментом исследования этих процессов.

Интеллектуальные информационные системы

В настоящее время зрелость информатики как науки характеризуется тем, что в ней взаимодействуют теория вычислений, алгоритмических языков и архитектуры компьютеров, а также искусственный интеллект, понимаемый как дисциплина об имитации и усилении рассуждений, и о восприятии и переработке информации посредством компьютера.

В результате развития этих разделов информатики стало возможно создание нового информационного продукта – *интеллектуальных систем*¹. Интеллектуальные системы реализуют взаимодействие «человек – компьютерная система» таким образом, что они, являясь человеко-машинными системами, образуют симбиоз (человек, компьютерная система). Компьютер в диалоговом режиме усиливает комбинаторное мышление и логические возможности человека. С этим фактом связано возникновение новой информационной технологии, реализующей функционирование интеллектуальных систем по следующей схеме:

Интеллектуальная система =
рассуждающая система + поисковая система + интеллектуальный интерфейс.

Интеллектуальная система – открытая система, принимающая решение с использованием новой поступающей информации, если она релевантна цели рассуждения. Таким образом, поисковая система оказывается средством принятия решения в интеллектуальной системе. Очевидно, что поисковая система может использоваться в двух режимах: в автоматическом – для подбора информации, близкой решаемой задаче, в диалоговом – для отбора информации, релевантной цели рассуждения, которая задана пользователем на некотором этапе работы системы.

Информационно-поисковая система как подсистема интеллектуальной системы должна обладать как механизмом поиска фактов, так и механизмом поиска документов. Высокоразвитая информационно-поисковая система (ее можно назвать интеллектуальной информационно-поисковой системой) должна обладать процедурами извлечения фактов, пополняющих базы данных из текстов на естественных языках. Это делает возможным полуавтоматическое (с использованием диалога пользователя и системы) расширение *базы знаний*, которая пополняется индуктивными обобщениями. Интеллектуальные системы являются средством компьютерной обработки и анализа данных и знаний высокого уровня: они не только имитируют рассуждения квалифицированного эксперта, но и усиливают их.

Из сказанного следует, что охарактеризованные нами интеллектуальные системы являются системами поддержки и усиления интеллектуальной активности человека в том смысле, который деклариро-

¹ Гергей Т., Финн В. К. Об интеллектуальных системах // Экспертные системы: состояние и перспективы. – М.: Наука, 1989. – С. 9–20.

вался авторами известного японского проекта компьютерных систем пятого поколения¹. Для отечественных условий, в которых создаются интеллектуальные системы, разумеется, специфичны трудности реализации конструктивных идей, воплощенных в современных по замыслам, логическим и программным средствам системах, так как для задач большой комбинаторной сложности требуется применение суперкомпьютеров и рабочих станций. В настоящее же время многие отечественные интеллектуальные системы созданы для персональных компьютеров. В связи с этим актуальной является задача создания сетей с использованием больших ЭВМ.

Интеллектуальные системы как инструмент новой информационной технологии обладают некоторыми новыми (по сравнению с информационными системами предшествующих поколений) возможностями. Например, при прогнозировании биологических активностей химических соединений интеллектуальные системы могут содержать как информацию о химических соединениях (физико-химические и стереохимические данные), так и информацию о путях их воздействия на организм (биохимические данные) и о противопоказаниях лекарственных соединений (медицинские и экологические данные).

В ВИНТИ разрабатываются интеллектуальные системы типа ДСМ, названные так по имени английского философа Джона Стюарта Милля. Эти системы применяются для прогнозирования свойств структурированных объектов в базах данных с неполной информацией для задач фармакологии, медицины и технической диагностики. Они могут быть применены и в других областях науки (например, в социологии), где знания слабо формализованы, данные хорошо структурированы, а в базах данных содержатся как положительные, так и отрицательные примеры некоторых эффектов.

Гипертекстовые системы

Возникновение и развитие идеи гипертекста

Гипертекст – это форма организации текстового материала, при которой его смысловые единицы (фразы, абзацы, разделы) представлены не в линейной последовательности, а как система явно указанных возможных переходов, связей между ними. Следуя этим свя-

¹ Симонс Дж. ЭВМ пятого поколения: компьютеры 90-х годов. – М.: Финансы и статистика, 1985.

ням, можно читать материал в любом порядке, образуя разные линейные тексты. Если речь идет о достаточно обширном материале с большим количеством связей, то возникает весьма сложное гипертекстовое пространство (сеть). Формирование и просмотр такой сети текстовых единиц возможны только при помощи компьютера¹.

Компьютерная гипертекстовая технология в самой общей форме понимается как "поддержка связей", т. е. обеспечение максимальной комфортности для пользователя при формировании и обработке сети связей. Имеется в виду, прежде всего, предоставление пользователю возможности легко добавлять в базу данных новые текстовые единицы, указывая их связи с уже имеющимися (было бы оптимально, если бы эти связи устанавливались автоматически на основе учета значения служебных слов). Не менее важна для пользователя и простота перемещения по образованной сети, т. е. возможность "читать" гипертекст в любом задуманном порядке².

Широкое внимание научной общественности к этой идее было привлечено несколько лет назад, когда на рынке компьютерных программ стали появляться системы, предназначенные для необычной интеллектуальной деятельности составления текста, имеющего нетрадиционную, "нелинейную" форму. Смысловые элементы этого текста могут читаться в разной последовательности, в соответствии с "разрешенными" смысловыми переходами, которые так или иначе указаны автором.

Теперь становится ясно, что успехи в развитии вычислительной техники и программирования позволили реализовать идеи, давно разрабатывавшиеся в недрах информатики. Возможности и тенденции развития информационной технологии в данном направлении были угаданы и верно предсказаны пионерами информатики. П. Отле, имя которого большинству специалистов известно лишь в связи с созданной им в 1905 г. Универсальной десятичной классификацией (УДК), уже в начале нашего века понимал необходимость упорядочения всемирной системы научной коммуникации.

¹ Conklin J. Hypertext: An introduction and survey // Computer. – 1987. – N 9. – P. 17–41.

² Гиляревский Р. С., Субботин М. М. Гипертекстовая информационная технология: Общая характеристика, применение, перспективы // Всесоюзная конференция «Актуальные проблемы развития и внедрения новой информационной технологии», Таллинн, 21–23 марта 1989 г.: Тезисы докладов и сообщений. – Ч. 1. – М., 1989; Субботин М. М. Гипертекст: Новая форма письменной коммуникации // Итоги науки и техники. Сер. Информатика. – Т. 18. – М.: ВИНТИ, 1994. – С. 1–158.

В его докладе на Международном конгрессе по библиографии и документации (Брюссель, 1908) была высказана мысль, содержащая зерно гипертекстовой технологии: "Средствами организации научной работы является книга и особенно ее нынешняя форма – журнал. Развитие науки шагнуло так далеко, что единственно правильным, соответствующим действительности подходом будет рассматривать все книги, все журнальные статьи, все официальные отчеты как тома, главы, параграфы одной великой книги, универсальной книги, исполинской энциклопедии, составленной из всего того, что было напечатано..."¹.

Следует, конечно, учитывать, что эта мысль была высказана в начале нашего века и ориентирована на технические возможности того времени. И хотя по теперешним понятиям они были весьма ограниченными, П. Отле предвидел современные достижения вплоть до систем теледоступа к банкам данных. В 1934 г. в "Трактате о документации" он писал: "Любой человек сможет прочесть издали спроецированный на его персональный экран отрывок, расширенный или суженный до объема необходимого предмета. Тем самым, сидя в своем кресле, каждый сможет созерцать весь мир или отдельные его части"².

Статья другого видного предтечи информатики В. Буша (1890–1974) "Возможный способ нашего мышления", опубликованная в 1945 г., получила всемирную известность в свое время и до сих пор считается наиболее значимым прогнозом развития информатики³. На нее ссылаются и почти все пишущие о гипертексте, так как в ней впервые было ясно показано, что неизбежная специализация научных интересов и ассоциативный характер мышления ученых приходят во все большее противоречие с традиционной информационной технологией. Человеческий мозг, по мнению В. Буша, работает совсем не так, как традиционные информационно-поисковые системы, – он мыслит ассоциативно. Получив информацию, ученый моментально испытывает потребность в другой информации, причем эта потребность возникает по ассоциации мысли, в соответствии с сетью связей между клетками

¹ Otlet P. La documentation en matière administrative // Actes de la Conference internationale de bibliographie et de documentation. – Bruxelles, 1908. – P. 147–154.

² Otlet P. Traité de documentation. – Bruxelles, 1934.

³ Bush V. As we may think // Atlantic Monthly. – 1945. – Vol. 176, N 1. – P. 101–108. Во время второй мировой войны В. Буш был научным советником президента Ф. Д. Рузвельта по военной технологии. См.: Zachary G. P. Vannevar Bush on the engineer's role // IEEE Spectrum. – 1995. – Vol. 32, N 7. – P. 65–69.

мозга. Желая имитировать этот мыслительный процесс техническими средствами, В. Буш предложил создать «расширитель памяти» – *Memex*, который хотя и не был построен, послужил прототипом микрофильмовых селекторов и других поисковых устройств. В начале 60-х годов эти идеи были использованы Д. Энгельбартом¹ и Т. Нельсоном, которые независимо друг от друга работали над созданием автоматизированных систем информационного поиска.

Как известно, Т. Нельсон является создателем термина "гипертекст", впервые приведенного им в докладе на конференции, а затем в статье 1967 г.; он привел этот термин в следующем контексте: "Современные информационно-поисковые системы как документального, так и фактографического типа не всегда могут удовлетворить запросы специалистов. Применение совершенных методов хранения и отображения текста в цифровой форме обеспечивает потенциальную возможность построения массива информации по крайней мере одного нового мощного вида: гипертекста или нелинейного текста. Ему будут свойственны отличительные черты книги и фильма... Гипертекст может отличаться от обычного текста порядком следования материала (его элементы могут размещаться в виде иерархического дерева или сети, он может иметь несколько уровней краткости изложения и детализации материала), способом его представления (воспроизведение движущихся и преобразуемых иллюстраций) и т. д."².

В то время и доклад и основанная на нем статья Т. Нельсона прошли незамеченными или в лучшем случае были восприняты как очередная компьютерная фантазия. И лишь спустя десятилетия идея гипертекста получила практическое воплощение, а сфера ее применения быстро расширяется. Со времени этих первых работ 60-х годов и до середины 80-х годов идея гипертекста переживала "инкубационный" период, когда многочисленные разработки, развивающие отдельные стороны этой идеи, велись разрозненно в рамках разных научных направлений.

¹ Engelbart D., English W. A research center for augmenting human intellect // AFIPS Conference Proceedings. – 1968. – Vol. 33. – P. 1.

² Nelson T. The information systems in the future // Information retrieval: A critical view / Ed. by G. Schechter, based on 3rd Annual Colloquium on Information Retrieval, May 12-13, 1966, Philadelphia, Pa. – Washington-London: Thompson, 1967; Нельсон Т. Информационные системы будущего // Информационный поиск: Сборник материалов / Сокр. пер. с англ. под ред. К. Н. Трофимова. - М.: Воениздат, 1970. - С. 217-228.

Логико-смысловой граф и логика связности

Гипертекст всегда представляет собой некоторую сеть или граф, отображающие систему связей между смысловыми единицами текста. Свойства гипертекста, его функциональные возможности в значительной степени зависят от структурных характеристик гипертекстовой сети. Она может иметь разную степень сложности, быть иерархической или циклической, члениться на обособленные части, быть "стройной" или "хаотичной". Чем более сложной, запутанной, насыщенной циклами является структура гипертекста, тем труднее его освоение как в функции чтения, так и в функции подготовки текста. Почти все авторы отмечают, что в гипертексте можно "заблудиться", потерять ориентацию, не найти удобных путей чтения и письма.

Можно трактовать семантические графы не как отображения текстов, а как представление знаний о предметных областях. Это явилось промежуточным этапом на пути к смысловым сетям, которые стали служить самостоятельным авторским средством изложения своих знаний и представлений о соответствующей предметной области¹. При формировании смысловых графов автор должен пользоваться определенными критериями и процедурами, чтобы отличать прямую смысловую связь от косвенной. Смежными по смыслу считались лишь те понятия и утверждения, которые можно объединить при помощи логических связей (типа "есть", "является причиной", "поэтому", "в этих целях" и т. п.). После того как был принят этот своеобразный критерий связи, направление стало называться "логико-смысловым моделированием", а логико-смысловые графы вплотную приблизились к гипертекстам. Их можно было читать, вставляя при переходе к смежному узлу соответствующую логическую связку.

Логико-смысловые графы, действительно, можно рассматривать как одну из версий гипертекста. Но введение критерия связи повело и к другим важным следствиям. Стал применяться принцип полноты связей, т. е. связь стала фиксироваться для всех пар высказываний, которые могли быть соединены связкой. Для каждого нового высказывания, вводимого в логико-смысловой граф, нужно было указать все его связи с высказываниями, уже имеющимися в этом графе. Это, конечно, создавало технологические трудности подбора "кандидатов на связь".

¹ Субботин М. М. Использование ЭВМ при построении содержательных рассуждений // Научно-техническая информация. – 1986. – Сер. 2, № 11. – С. 1–7.

Но именно принцип полноты связей открыл путь к исследованию структурных характеристик смысловой сети.

Стали значимыми такие характеристики, как число связей высказывания (степень соответствующей вершины графа), показатель центральности (сумма расстояний от данной вершины до всех других), наличие и число путей между какими-либо вершинами. При этом структурные характеристики получают содержательное толкование. Например, число связей высказывания можно рассматривать как признак его относительной значимости в рамках данной системы суждений. Появилась возможность представления предметной области с разной степенью детализации путем построения укрупненных графов, включающих лишь те смысловые единицы, у которых число связей превышает определенный порог. Но и для самого гипертекста здесь таятся интересные возможности: ведь таким способом в его смысловой сети можно автоматически отыскивать предпочтительные пути.

Логико-смысловой граф оказывается адекватным средством для анализа связности, для исследования систем, у которых ценится высокая связность. Так могут интерпретироваться социальные позиции, системы взглядов, научные концепции, новые идеи, которые должны обладать единством, целостностью. Все элементы такого смыслового образования – принципы, утверждения, аргументы – должны быть хорошо связаны между собой, а не являться набором разрозненных высказываний. Взаимосвязанность положений концепции обычно ощущается непосредственно, интуитивно. Однако довольно часто возникает необходимость представить эти связи эксплицитно, особенно если концепция претендует на практическое воплощение и затрагивает интересы многих людей. Описываемый метод позволяет установить, насколько тесно конечные выводы связаны с тем материалом, на который они опираются.

Гипертекст как развитие функций чтения и письма

Известны гипертекстовые системы, в которых на первый план выступает функция чтения. Таковы, например, учебные и справочные системы, в которых читатель сам выбирает, как ему двигаться при освоении материала в сети связанных по смыслу текстовых фрагментов, причем система подсказывает ему возможные варианты такого движения. В других случаях в качестве основной выступает функция письма, авторской работы по составлению текста. Здесь многовариантное

представление фрагментов текста в виде сети возможных переходов используется для нахождения хорошей последовательности изложения.

В соответствии с этим в применении гипертекстовых систем сложилось несколько основных направлений. Одно из них – "электронная книга" – обеспечивает освоение материала с большим количеством ссылок и смысловых пересечений. В качестве объектов могут выступать справочные и учебные материалы, проектная и программная документация. Каждый предъявляемый пользователю текстовый фрагмент снабжается указанием всех его ссылок и возможных смысловых переходов к другим фрагментам. Другое направление применения гипертекстовых систем – компоновка крупных текстовых материалов из фрагментов, которые первоначально представлены в форме сети с указанием их взаимных смысловых связей. Третье направление – представление в форме единого гипертекста идей, аргументов и предложений, вносимых участниками коллективной работы, рассмотрение и анализ взаимосвязи этих идей и аргументов.

Коммерческие гипертекстовые системы выпускаются с 1987 г. Наибольшую известность получили американские системы Guide, Hypercard и французская – Hyperdoc. Первая из них принадлежит американско-шотландской фирме Owl Technologies и привлекает своей простотой. Фактически Guide является развитой системой обработки текста с оригинальной концепцией "кнопки". Любое слово или словосочетание в тексте может быть определено как "кнопка". В этом случае постановка над ним курсора активизирует определенное действие: замену слова другим, вызов комментариев, установление связей слова с другими текстами, формирование на экране окон.

Система Hypercard фирмы Apple для персональных компьютеров Macintosh представляет собой своеобразную электронную картотеку. Карточки могут содержать не только текст, но и изображения. Определенные места карточки являются "кнопками", нажатие на которые вызывает новые карточки, которые тоже имеют "кнопки". Это позволяет охватывать любое число карточек, связанных между собой иерархически, тематически, ассоциативно. Массивы могут обрабатываться независимыми программами, написанными на специальном языке Hupertalk. Это позволяет сделать изображения движущимися, проводить сложные виды поиска, генерировать музыкальное сопровождение, создать систему обработки текста и изображений. Можно считать, что эта система явилась полной реализацией *Мемекса*, предложенного В. Бушем.

Еще одной коммерческой системой является Hyperdoc, созданная французской фирмой GECI. Основное ее достоинство – независимость от характера данных и от аппаратных средств. Hyperdoc оперирует с текстами, чертежами, диаграммами, логическими схемами, реализуется на компьютерах IBM PC, Macintosh, Atari, Vax. Информационные массивы хранятся на оптических дисках. Любая зона экрана может быть ассоциирована с любой группой данных при помощи устройства "мышь". Наряду с межуровневыми связями каждый уровень может иметь свои связи. Hyperdoc легко сопрягается с системами обработки текста, электронной графики, издательскими пакетами, СУБД dBASE-3.

Интересным применением гипертекстовой технологии является представление в форме гипертекста Оксфордского словаря английского языка (322 тыс. статей, 56,3 млн слов, 2,4 млн ссылок)¹. Это позволяет просматривать все связи между словами и их толкования с помощью программ, повышает эффективность справочного аппарата. Словарь служит основой установления гипертекстовых связей для других документов. Их тексты смогут связываться отношениями цитирования и тематической близости. Гипертекст будет использоваться и как средство совершенствования и редактирования словаря в рамках систем компьютерной лексикографии.

Отечественные гипертекстовые системы ГИПЕРЛОГ и СЕМПРО

Системы разработаны на основе многолетних исследований М. М. Субботина, о которых говорилось выше. Они реализуют функции и процедуры, позволяющие:

- формировать из элементов гипертекстовой базы данных связанные, упорядоченные тексты на задаваемые пользователем темы;
- контролировать качество формируемого текста, выявлять в нем логические и смысловые разрывы;
- структурировать гипертекстовую базу данных, выявлять в ней комплексы тесно взаимосвязанных идей, понятий, проблем.

Эти функции и процедуры реализуются на основе оригинальных идей так называемой *логизированной* версии гипертекста. Системы предоставляют пользователю возможность при вводе фрагментов текста в базу данных устанавливать между ними связи (ссылочные, смы-

¹ Raymond D. R., Tompa F. Hypertext and the Oxford English Dictionary // Communication ACM. – 1988. – Vol. 31, N 7. – P. 871.

словые, логические, ассоциативные и другие), обеспечивают компьютерную поддержку этих связей и перемещение по ним. Таким образом, гипертекст формируется как совокупность взаимосвязанных фрагментов текста. Эти фрагменты могут представлять собой как целые документы, так и отдельные высказывания, формулировки идей, проблем, предложений, мероприятий, фактов. Система обеспечивает максимальную открытость гипертекста, возможность его пополнения, изменения структуры и содержания на любом этапе работы. Она не навязывает пользователю готовые схемы и ограничения на структуру представления информации.

Данные системы предназначены для использования в таких областях деятельности, как анализ проблем, изучение прецедентов, прогнозирование социальных явлений, обоснование управленческих решений, подготовка различных документов: обзоров, аналитических материалов, пояснительных записок, докладов и т. п. В ряду гипертекстовых систем общего назначения они выделяются тем, что позволяют посредством анализа структурных характеристик гипертекста увидеть в обозримой форме укрупненные комплексы проблем, узловые вопросы и аспекты. При формировании текстовых документов рассматриваемые системы позволяют увидеть в текстах логические пробелы, для устранения которых требуется дополнительная информация. Первая система реализована на основе СУБД Revelation, вторая – на специально разработанной СУБД, которая имеет более дружественный интерфейс и предоставляет большие возможности пользователю. Обе они могут функционировать на IBM-совместимых компьютерах в среде MS-DOS.

Системы гипермедиа как развитие гипертекста

Новые подходы к манипулированию информацией, хранящейся в ретроспективном фонде, открывают перспективы качественно иного, более эффективного использования постоянно возрастающего объема документальных источников информации. Принципиальной особенностью *гипермедиа* (их называют еще «гиперсредствами» или «системами гиперзаписи») является распространение идеи гипертекста, т. е. ассоциативно связанной текстовой информации, на изобразительную и звуковую информацию, хранящуюся в цифровой форме.

Информационные системы, обеспечивающие функционирование гипермедиа, должны иметь особые технические, программные и

телекоммуникационные средства. Разумеется, эти средства создают лишь необходимые предпосылки для реализации систем гипермедиа, основу же их функционирования составляют алгоритмы и программы. Можно указать на некоторые из них:

- указатель к гиперБД, содержащий аннотированный перечень характеристик всего массива;
- карта связей гиперБД, отражающая в графической форме ее структуру и методы доступа к информации;
- средства передвижения пользователя в гиперБД и возможности создания им своих способов манипулирования данными;
- средства аудио- и видеоконтроля, обеспечивающие доступ к изобразительной и звуковой информации.

Поскольку гипермедиа не имеют пока точного определения, есть тенденция понимать их слишком широко. В одном из ранних определений говорилось: «Системы гипермедиа относятся к типу систем, базирующихся на использовании наиболее передовых технологий и технических средств и предназначенных для повышения эффективности и интенсификации процессов взаимодействия человека и всей среды, относящейся к знаниям»¹. Данное определение подчеркивает основное функциональное назначение гипермедиа – обеспечивать эффективную коммуникацию между человеком и источниками знания, а также их связь с новыми для нашего времени технологиями, но оно, разумеется, носит слишком общий характер.

Не претендуя на свое определение гипермедиа, хотел бы сказать, что из всех возможных трактовок, предпочтительной является та, которая связывает эти системы с интеллектуальными информационными системами. Верно, что гипермедиа интегрирует цифровую запись текстовой, изобразительной и звуковой информации, но это чисто прикладная особенность данных систем, как и то, что они используют все существующие виды носителей оцифрованной информации. Все же основные функциональные характеристики этих систем связаны с решением принципиального вопроса о формализации представления и структурирования информации и алгоритмизации процессов ее обработки.

Системы гипермедиа, как и гипертекстовые, могут рассматриваться в разных аспектах. Один из подходов, близкий программистам,

¹ Gains B. R., Vickers J. N. Design considerations for hypermedia systems // Microcomputers Information Management. – 1988. – Vol. 5, N 1. – P. 1-27.

заключается в том, чтобы сравнить методы доступа к информации в гипертексте с соответствующими методами в СУБД. Эти методы различны: в гипертексте они опираются на ассоциативные связи между понятиями, а в СУБД – на структурные свойства данных. В соответствии с этим гипертекст можно рассматривать как систему ассоциативной организации и поиска информации. Между системами гипертекста и гипермедиа нет четкой границы. Следует иметь в виду, что в последнее время термин «гипермедиа» используется все реже, так как заменяется термином «мультимедиа», который первоначально означал систему совместного использования цифровой и аналоговой записи информации (например, компьютерного текста и видеоизображений).

Эти системы представляют собой этапное достижение в развитии информационной технологии, ориентированной в первую очередь на обработку знаний. Новые возможности интерактивного доступа человека к неограниченным объемам накопленных знаний, обусловленные широким использованием вычислительной техники и интеграцией различных носителей информации, создают предпосылки для повышения творческой активности человека.

Многие специалисты высоко оценивают перспективы технологий гипертекста и гипермедиа, считая, что эти технологии вышли на уровень стратегических ресурсов компьютерных корпораций¹.

Системы машинного перевода

В современную эпоху научно-технической революции и информатизации общества возросла интенсивность общения между народами и странами. Однако этот процесс в значительной мере тормозится языковыми барьерами. Обучение иностранным языкам и переводческая деятельность в какой-то мере смягчают остроту проблемы, но полностью ее не решают. Более радикальным решением является создание систем автоматического перевода текстов с одних естественных языков на другие. Такие системы создаются во многих развитых странах мира, однако качество автоматического перевода оставляет желать лучшего.

¹ Smith K. E. Hypertext – linking to the future // Online. – 1988. – Vol. 12, N 2. – P. 32–40; Гиляревский Р. С., Субботин М. М. О возможностях оценки перспективности новых информационных технологий (на примере гипертекстовой технологии) // Научно-техническая информация. – 1988. – Сер. 2, № 12. – С. 2–5.

Многие выдающиеся лингвисты вообще ставили под сомнение – и не без основания – возможность адекватного перевода текстов с одного естественного языка на другой, как это ни парадоксально звучит в эпоху интенсивной переводческой деятельности. Для получения на практике адекватного перевода необходимо использование экстралингвистической информации, т. е. такой, которая не содержится в переводимом тексте, но существует в виде накопленного общественного знания. Это и служит основным препятствием для полностью автоматического (т. е. осуществляемого без участия человека) перевода с одного языка на другой. Поэтому, говоря о машинном переводе, мы подразумеваем лишь частично автоматизируемую деятельность, в которой на разных ее этапах участвует человек. Поскольку перевод специальных текстов при помощи компьютера может быть значительно облегчен и ускорен, системы машинного перевода стали полезным инструментом в работе переводчика и важным фактором снижения затрат в этой области.

Человеческий перевод текстов с одних естественных языков на другие – это сложный мыслительный процесс. Он осуществляется на основе восприятия исходного текста и последующей передачи его смысла средствами выходного языка. При этом переводятся не слова и их последовательности, а понятия и мыслительные образы, порождаемые в сознании переводчика под их воздействием. Системы машинного перевода текстов предназначены для моделирования работы человека-переводчика. Но если моделировать эту работу в полном объеме пока не представляется возможным, то нужно, по крайней мере, стремиться при машинном переводе оперировать теми единицами языка и речи, которые позволяют наиболее точно передавать содержание текста, написанного на одном языке, средствами другого языка. Такими единицами являются, прежде всего, фразеологические обороты и терминологические словосочетания и, во вторую очередь, отдельные слова. Поэтому перспективные системы машинного перевода должны опираться на фразеологическое богатство естественных языков. Они должны быть *системами фразеологического перевода*.

Концепция фразеологического машинного перевода была впервые четко сформулирована профессором Г. Г. Белоноговым в 1975 г. Далее она была развита и в настоящее время реализована в ВИНТИ в виде двух систем: системы русско-английского перевода (RETRANS) и

систем англо-русского перевода (ERTRANS)¹. Если в других системах перевода в качестве основной минимальной единицы смысла, представляемой в машинных словарях, рассматривается слово и их можно охарактеризовать как системы преимущественно пословного семантико-синтаксического перевода, то в системах фразеологического перевода в качестве основной единицы смысла считаются фразеологические словосочетания, выражающие понятия, отношения между понятиями и ситуации. Это позволяет точнее передавать смысл переводимых текстов.

Как уже было указано, система RETRANS предназначена для перевода текстов с русского языка на английский. Тематика переводимых текстов включает широкий спектр предметных областей: экономику, коммерческую деятельность, машиностроение, электротехнику, энергетику, транспорт, авионавтику, космонавтику, биологию, медицину, экологию, сельское хозяйство, математику, физику, химию, автоматику и радиоэлектронику, вычислительную технику, информатику, астрономию, геофизику, геологию, горное дело, металлургию, политику, законодательство и другие дисциплины. Словарь системы содержит около миллиона словарных статей и обеспечивает покрытие политематических текстов на 97–99 %. Это самый большой в мире русско-английский машинный словарь. Доля словосочетаний и фразеологических оборотов в словаре – около 80 %.

Система реализована на персональных компьютерах типа IBM PC/AT. Скорость перевода текстов в автоматическом режиме – не менее 10–30 слов/сек. и зависит от быстродействия машины. Предусмотрена возможность работы в интерактивном режиме (с целью повышения качества перевода). Есть также возможность дополнительной настройки системы на конкретного пользователя. Для функционирования системы необходим объем оперативной памяти не менее 600 Кбайт и объем дисковой памяти не менее 20 Мбайт. Система работала под управлением операционной системы MS DOS 6.0 и выше. Теперь она работает под Windows 2000, встраивается в Word и доступна в Интернете на сайте ВИНТИ. Система англо-русского перевода (ERTRANS) имеет характеристики, аналогичные системе RETRANS.

¹ Белоногов Г. Г. и др. Интерактивная система русско-английского и англо-русского машинного перевода политематических научно-технических текстов // Научно-техническая информация. – 1993. – Сер. 2, № 3. – С. 20–26; Белоногов Г. Г., Зеленков Ю. Г. Еще раз о принципе аналогии в морфологии // Научно-техническая информация. – 1995. – Сер. 2, № 3. – С. 29–32.

Одной из важнейших проблем, стоящих перед переводом, является частое и не всегда сразу заметное изменение значений слов. Словари не всегда успевают отразить эти изменения в научно-технической терминологии. В одном из докладов на международной конференции переводчиков приводились интересные примеры из вычислительной техники. Слово «компьютер» во времена Шекспира обозначало человека, выполняющего арифметические вычисления. В наше время подобное изменение претерпело слово «редактор», которым все больше обозначают программу обработки текста. Английские же слова *word processor*, первоначально употреблявшиеся в значении компьютера для обработки текста, а затем – и соответствующей программы, претерпело обратное изменение: теперь они часто применяются к людям и называют на специалистов, поддерживающих работу этих программ.

С точки зрения пользователя системы машинного перевода могут подразделяться на три основных типа:

Информативные, предназначенные для помощи тем, кому нужен доступ к информации на иностранном языке и кто готов пользоваться «грубым», но достаточно понятным переводом. Такие системы, как правило, имеют словари большого объема, но не опираются на новейшие достижения в лингвистике и программировании.

Профессиональные, которые дают лишь черновые наброски перевода для профессиональных переводчиков и тем освобождают их от черновой работы. Такие системы теперь используются все реже, – как правило, при большом объеме текущей переводческой работы, выполняемой одновременно многими специалистами в одной предметной области. Чаще в этих ситуациях переводчиков снабжают автоматическими словарями, тезаурусами с интерактивным доступом или системами, получившими название «памяти переводчика».

Персональные – для авторов, желающих перевести свои статьи на иностранный язык, которым они не вполне владеют. Такие системы обычно работают в диалоге с пользователем и могут давать удовлетворительный перевод (качество которого все же зависит от того, насколько автор владеет выходным языком).

По применяемым лингвистическим методам системы машинного перевода можно разделить также на три типа:

Системы прямого перевода – наиболее многочисленные, поскольку начали создаваться еще в 50–60-е годы для фиксированных пар языков. В этих системах словарь и синтаксис входного языка анализируются лишь в той мере, в какой это необходимо для идентифика-

ции правильных выражений выходного языка и порядка слов. В начале своего развития эти системы выдавали пословные переводы и лишь позднее – переводы, основанные на анализе предложений входного языка.

Системы перевода с использованием языка-посредника, служащего для отображения «смысла» входного текста, который преобразуется в семантические и синтаксические представления, общие для нескольких выходных языков. Этот метод применяется обычно при необходимости перевода исходного текста на несколько языков (в переводческих центрах Европейского сообщества, например).

Системы перевода с трансфером более сложны, нежели предыдущие типы, поскольку языки-посредники применяются дважды – первый раз при переводе с входного языка, второй – при переводе на выходной язык. В этом случае становится необходимым дополнительный этап перевода – с языка-посредника входного языка на язык-посредник выходного языка. За этот счет достигается более глубокий лингвистический анализ и синтез.

В последние годы все большее применение в машинном переводе находят *методы искусственного интеллекта*, которые при переводе учитывают семантику текста. Это означает, что они опираются не столько на грамматические, сколько на семантико-синтаксические категории. Обычные для лингвистических методов многочисленные неоднозначности и неясности устраняются за счет внеязыковой базы данных. Это означает, что система пытается «понять» текст на входном языке до его перевода. Однако и методы искусственного интеллекта пока не дают всей информации, необходимой для полноценного машинного перевода. В частности, проблемы возникают при переводе с английского языка на японский. «Понимание» английского текста не дает достаточной информации о состоянии пишущего и читающего, необходимой для адекватного перевода на японский язык.

Несмотря на все оговорки, связанные с несовершенством систем машинного перевода, существуют уже сотни достаточно широко используемых систем такого рода.

Информационная технология

О понятии информационной технологии

Слово "технология" имеет в русском языке два значения. В соответствии с толковым словарем оно означает совокупность процессов обработки или переработки материалов в определенной отрасли, а также научное описание способов производства. В каком-то смысле оно противопоставляется слову "техника", которое означает совокупность средств труда и приемов, служащих для создания материальных ценностей, а также употребляется собирательно вместо слов "машины", "орудия", "устройства". В английском языке слово *technology* означает технические науки и часто употребляется в значении, которое на русский язык должно переводиться словом "техника", хотя обычно калькируется как "технология".

Вот почему термин "информационная технология", пришедший к нам из английского языка, часто трактуется расширительно и охватывает не только процессы и методы обработки информации, но и технические средства их осуществления. Это обстоятельство наложило некоторый отпечаток на содержание данной лекции, так как трудно говорить о собственно технологии без учета быстрого развития технических средств. Однако сведения об основной технике – вычислительных машинах и их программировании вынесены в отдельные лекции, чтобы здесь акцентировать внимание на электронных информационных технологиях, их применении и социальных последствиях их внедрения.

Информационные технологии не являются самоцелью: их разработка и внедрение служат основой информатизации общества как один из главных факторов научно-технической революции. Новые и перспективные информационные технологии, в свою очередь, создаются на основе новейших технических средств, высокопроизводительных вычислительных машин, внешних запоминающих устройств

сверхбольшой емкости, информационно-вычислительных сетей, электронных средств коммуникации и печати.

Информационная технология нередко включается в более широкое понятие информационной сферы, которая представляет собой совокупность общенациональных, отраслевых и региональных информационных структур, в составе которых находятся и библиотеки. Информационные технологии удовлетворяют их потребности не только при помощи технических средств, но и социальных институтов и действующих норм. Хотя термин "технология" и трактуется нами расширительно и охватывает не только и не столько процессы и методы обработки информации, сколько технические средства их осуществления, однако нельзя сводить информационную технологию к технике, только к вычислительным машинам и их периферийным устройствам. Она определяется, в первую очередь, видом перерабатываемой информации, производимым продуктом или предоставляемой услугой, а также информационной структурой, которая использует данную технологию.

Примерами информационных технологий могут служить:

- автоматизированное проектирование и производство,
- телеобработка данных,
- автоматическая обработка текстов и изображений,
- поиск информации в базах данных,
- системы мониторинга окружающей среды,
- системы технической диагностики и контроля,
- экспертные, обучающие и роботизированные системы,
- гибкие автоматизированные производства,
- видеотекст и телетекст,
- электронная полиграфия,
- моделирование сложных научно-технических процессов в реальном масштабе времени и многое другое.

Обо всем этом следовало сказать, чтобы не создалось обманчивое представление о том, что в сфере науки используется весь современный арсенал электронной технологии, вся совокупность ее методов и средств. Пока мы освоили лишь незначительную ее часть, связанную, главным образом, с обработкой информации и ее поиском.

Западные специалисты считают, что можно выделить три стадии усвоения обществом той или иной технологии:

- улучшение и ускорение привычно выполняемой работы,

- появление под влиянием технологических изменений новых проблем, задач и целей, ранее не возникавших,
- изменения в самом обществе, его институтах и образе жизни его членов в связи с технологическими достижениями.

В развитых странах уже не первое десятилетие решается вопрос информатизации общества, что соответствует третьей стадии усвоения электронной информационной технологии.

Традиционная информационная технология тесно связана с процессами письма и чтения, редактирования и издания, которые осуществляются "естественным" интеллектом человека и на протяжении нынешнего тысячелетия предполагали использование бумаги. Поэтому процессы автоматического представления и сканирования текста, автоматизированной переработки информации, осуществляемые при помощи "искусственного" интеллекта в процессоре ЭВМ, часто называют "безбумажными".

Сначала специалисты по обработке данных писали о безбумажных информационных системах, о безбумажной информатике, а затем социологи начали говорить о безбумажном обществе. Концепция безбумажного общества есть не что иное, как представление о воздействии новой информационной технологии на те сферы жизни людей, которые ближе всего соприкасаются с научной или, шире, с семантической информацией.

В нашу специальную литературу термин "безбумажная информатика" вошел как синоним новой информационной технологии вместе с книгой В. М. Глушкова (1923–1981) "Основы безбумажной информатики", вышедшей в 1982 г. Во введении к ней значение электронной технологии обработки информации обосновывалось следующим образом: "Задача накопления, обработки и распространения (обмена) информации стояла перед человечеством на всех этапах его развития. В течение долгого времени основными инструментами для ее решения были мозг, язык и слух человека... Поскольку в эпоху книгопечатания основным носителем накапливаемой информации стала бумага, технологию накопления и распространения информации естественно называть бумажной информатикой. Следует подчеркнуть, что революция в информатике, связанная со становлением письменности и книгопечатания, практически не затронула область *переработки* информации. Положение в корне изменилось с появлением электронных вычислительных машин (ЭВМ). Подобно тому, как изобретение механического двигателя открыло эру комплексной механизации и автома-

тизации физического труда, изобретение ЭВМ сделало то же самое в отношении труда умственного"¹.

Тенденции развития информационных технологий

Если посмотреть на процесс внедрения новой информационной технологии в информационное дело не с узко технологических позиций, а в более широком социологическом аспекте, то нельзя абстрагироваться от общих тенденций развития информационной коммуникации и ее инфраструктуры. К этой инфраструктуре помимо библиотек относятся издательства и типографии, система книготорговли и распространения литературы, средства связи (почта, телеграф, телефон, радио и телевидение, телефакс, видеотекс, сети передачи данных), специализированные информационные центры.

В обобщенном виде эти тенденции таковы. Увеличение числа полнотекстовых баз данных и расширение производства оптических дисков углубляет процесс сращивания библиотечно-информационной деятельности с редакционно-издательским делом и полиграфией. Распространение микроЭВМ и персональных компьютеров ведет к созданию в центрах информации и у пользователей на дому автоматизированных рабочих мест, связанных в локальные информационные сети с выходом в глобальные. По этим сетям пользователи могут выходить в мощные национальные и международные центры поиска информации и копирования первоисточников (в некоторых системах копии пересылаются по каналам электросвязи и даже через искусственные спутники Земли).

Таким образом, специалисты могут формировать собственные базы данных, получая исходную информацию из разных источников, включая и приобретение компактных оптических дисков с текстами монографий, статей, библиографических указателей, каталогов, справочников. Все шире распространяются программные системы, позволяющие ученым и специалистам осуществлять поиск и переработку информации самостоятельно, без посредников (экспертные системы, гипертекстовые, системы обработки текста и изображений, электронного редактирования, автоматического индексирования, машинного перевода и т. п.). Все более широкое применение находят настольные издательские системы (персональный компьютер + лазерный принтер), электронная почта, телеконференции. Наконец, самое главное, вся эта

¹ Глушков В. М. Основы безбумажной информатики. – М.: Наука, 1982. – С. 9.

деятельность осуществляется на прибыльной коммерческой основе, так как необходимая для нее совершенная техника, хотя она довольно быстро дешевеет, требует затраты значительных материальных средств.

Возможно, что именно последней из названных тенденций объясняется то, что библиотеки пока еще не играют заметной роли в обеспечении своих читателей электронной информацией. Чтобы не возвращаться к этому принципиальному моменту, требующему специального обсуждения, хочу сразу высказать свое к нему отношение. Бесплатность библиотечного обслуживания явилась исторически важным этическим принципом. Но он давно уже на практике оказался размытым. Ведь предоставление копий во всем мире является платным, хотя в большинстве стран – на бесприбыльной основе. Все электронные информационные услуги как в информационных центрах, так и в библиотеках тоже предоставляются за плату. Так как же быть с принципом бесплатности?

Считаю, что общество не может снять с себя ответственность за осуществление права каждого гражданина быть информированным на некотором базовом уровне. Этот уровень предполагает, что любой потенциальный читатель должен иметь доступ ко всей опубликованной информации при некотором минимуме усилий и в разумные сроки.

Поэтому традиционное библиотечно-библиографическое обслуживание должно оставаться бесплатным, а, скажем, цены на библиографические указатели, реферативные журналы, другие информационные издания, а также на копии первоисточников – доступными (на уровне подписных цен на массовые издания). Весь же информационный сервис на базе новой информационной технологии неизбежно является и останется платным. При этом по мере возрастания интеллектуальной сложности информационных услуг цены на них будут также расти.

Любопытно проследить и за судьбой давней дискуссии по не менее принципиальному для данной темы вопросу о том, прекратился ли информационный взрыв и сохраняются ли прежние темпы прироста информации, по крайней мере, научно-технической. В 1973 г. профессор Сорбонны Ж. Андерла в исследовании, озаглавленном "Информация в 1985 году", утверждал, что к этому времени прирост недублируемой информации составит 12 млн. единиц в год, увеличение темпов роста 12–13 %, доля машиночитаемой информации в автоматизированных системах 25 %, а темпы ее роста 35–40 % в год. Ему возражали

ведущие американские специалисты В. Коч (Институт физиков, Вашингтон), К. Херринг ("Белл Телефон") и Д. Прайс (Йельский университет) на семинаре Национального научного Фонда США в 1974 г. Они проявили скептицизм в отношении оценок количественного роста информации и высказались за то, что в течение полувека ежегодный информации сохранится на уровне 7–8 %.

В 1985 г. Ж. Андерла проанализировал, в какой мере сбылись его прогнозы. Он отметил, что не предвидел микропроцессорной революции и распространения персональных компьютеров, влияния европейских правительств на развитие информационной индустрии и поэтому предсказанные им изменения в информационной технологии наступили на пятилетие раньше предсказанного срока, хотя в начале 70-х годов были восприняты как утопические. До настоящего времени учет единиц машиночитаемой информации не налажен в той мере, в какой он ведется для традиционно издаваемой литературы. Поэтому точно подсчитать, насколько сбылись прогнозы роста информации, пока невозможно. Но приблизительные оценки, основанные на учете баз данных М. Вильямс, показывают, что в споре с американскими учеными правота была на стороне Ж. Андерла. Электронная информационная технология не сдерживает, а, наоборот, стимулирует рост информации, в том числе и научно-технической. Только происходит это в нетрадиционных формах, которые мы не очень умеем и не торопимся учитывать.

В начале 90-х гг. проводилось много исследований с целью оценить степень распространения информационной технологии в библиотеках и информационных службах в 1995–2010 гг. Эксперты не ошиблись в том, что в этот период от половины до трех четвертей потребителей будут иметь собственные телефаксы и широко использовать электронную почту. Видеотекс во французском варианте может сыграть определенную роль в межбиблиотечных связях, но его широкое распространение маловероятно из-за конкуренции компьютерных сетей. До 90 % научно-технической информации может издаваться в электронной форме, при этом в цифровом виде будет храниться значительная часть изобразительной и звуковой информации.

Использование экспертных систем, считали эксперты, позволит улучшить результаты информационного поиска. Произойдут изменения в системе МБА: тексты и изображения будут передаваться по каналам связи и выводиться на терминалы потребителей. Каталоги будут во все большей степени становиться сводными и, возможно, наряду с

библиографической информацией, классификационными индексами и шифрами хранения содержать изображения титульных листов.

Как в большинстве прогнозных исследований последних полутора десятилетий эксперты подчеркнули, что внедрение новой информационной технологии зависит от экономических факторов не в меньшей степени, чем от технических достижений. Хотя период, на который были рассчитаны эти прогнозы, еще не закончился, можно констатировать, что они реализуются с опережением. При этом прогнозы развития техники и технологии оказываются слишком осторожными, а прогнозы достижений в фундаментальных науках – слишком оптимистичными.

Влияние информационных технологий на развитие науки

Важной тенденцией в развитии информационных технологий было названо слияние редакционно-издательской и информационной деятельности. Технически она связана с появлением настольных издательских систем, которые получили большое распространение. Обычно такие системы состоят из персонального компьютера с достаточно большой внешней памятью и снабжены быстродействующим лазерным принтером и сканером (вводным читающим устройством). По некоторым подсчетам экономия от их применения в подготовке деловой документации и научно-технической информации составляет до 50 %. Однако для науки значение этой тенденции значительно глубже, нежели просто совершенствование издательского процесса.

Наиболее сильно влияют на развитие науки следующие изменения в информационной технологии:

а) рост производительности компьютеров при снижении стоимости обработки информации (их быстродействие и объем оперативной памяти продолжают расти на порядок каждый год без увеличения общей стоимости компьютеров),

б) возможность хранения в машиночитаемой форме больших объемов информации. Появление компактных оптических дисков с их большой емкостью (ныне равной объему нескольких тысяч книг) ознаменовало относительно невысокую стоимость хранения информации (стоимость компакт-диска сравнима с ценой граммофонной пластинки) и универсальность применения (текстовые, графические, звуковые, видео файлы, исполняемые и другие системные программы),

в) увеличение пропускной способности каналов связи и снижение стоимости передачи данных (волоконная оптика, спутниковые сис-

темы: стоимость передачи 1 бита информации снижается на порядок за несколько лет),

г) возможность "персонализации" процесса обработки информации за счет широкого доступа конечного пользователя к электронным информационным средствам.

Развитие технологии обработки информации привело в сфере научного обеспечения к возникновению новых информационных структур, меняющих методiku научного исследования, стандарты точности, сроки выполнения научных работ. Возможности этих технологий используются пока недостаточно, хотя эти структуры существенно влияют на технологию получения, передачи, хранения, и использования информации. Речь идет о следующих элементах информационной технологии.

Автоматизированные системы сбора данных приводят к стиранию границ между научным прибором и компьютером, между процессами сбора данных и их обработки, становится несущественной пространственная разобщенность исследователя и прибора. Машинные модели, позволяющие исследовать природные явления и технические системы в реальном масштабе времени, начинают конкурировать по затратам и точности с натурными экспериментами в физике и биологии. Сети передачи данных (локальные и удаленные), создаваемые как системы межмашинного обмена, во все большей степени становятся средствами межличностного общения ученых. Наиболее популярной сетевой услугой теперь является электронная почта и ее производные (электронные журналы, телеконференции). Наблюдается усиленное использование компьютеров для подготовки обычных научных изданий (уже в середине 80-х гг. 60 % всех документов по науке и технике на Западе подготавливались при помощи вычислительной техники).

Характерно широкое применение конечным пользователем коммерческих текстовых и цифровых баз данных на основе совершенствующихся систем человеко-машинного интерфейса. Базы данных научного характера составляют половину всех текстовых баз и обслуживают треть всех запросов. Наиболее перспективными считаются полнотекстовые базы (быстрое обновление, поиск по всему тексту, а не только по поисковому образу, немедленная выдача всего текста). Рост числа цифровых общедоступных баз данных сдерживается недоверием к "недокументированной" информации, отчужденной от своего создателя, а также недостаточной стандартизацией данных на всех уровнях

(от методов их получения до формата представления). При этом растет число цифровых баз "для домашнего пользования".

Информатизация в науке сопровождается некоторыми последствиями, которые оказываются неожиданными для инициаторов:

– значительную и быстро растущую долю вычислительных ресурсов, выделяемых науке и образованию, реально используют административные службы этих сфер. Вообще бизнес и управление более эффективно используют информационные технологии, чем наука и образование;

– хотя информатизация и повысила уровень соревнования в науке, но в значительной мере оно стало соревнованием за доступ к информационным и вычислительным ресурсам.

Одним из популярных видов современной информационной технологии являются экспертные системы. Они обеспечили возможность новых подходов к раскрытию знания, зафиксированного не только в научных документах, но и в интеллекте ученых. Экспертные системы пришли на помощь информационно-поисковым системам координатного индексирования. Принципиальное отличие этого нового средства видно из сопоставления обычного информационного поиска и работы с экспертной системой.

Если в первом случае упор делается на то, какая информация необходима, то во втором оказываются более важными ответы на вопросы, почему и для чего она нужна. Информационно-поисковая система в качестве запроса требует указания предмета или области интереса пользователя и выдает ему релевантные сведения или документы, в которых они могут содержаться. Экспертная система нуждается в описании проблемы или ситуации, а в ответ выдает гипотезы, рекомендации и объяснения, сопровождаемые показом хода рассуждений.

Но и экспертные системы уже не последнее слово информационной технологии. Даже самые "интеллектуальные" из них не могут удовлетворить потребности специалистов в манипулировании знаниями и в использовании возможностей компьютера при работе с текстом. Были проведены эксперименты, которые показали, что при большом корпусе текстов экспертные системы не справляются со своими задачами. Они часто решают не ту проблему, которая стоит перед специалистом, и ограничивают его активную роль в поиске решения.

Дело в том, что модуль объяснения в этих системах часто основан на ограниченных логических средствах и не позволяет включать в процесс поиска богатство человеческой интуиции и непредсказуемых

ассоциаций. Самая же существенная ограниченность экспертных систем состоит в том, что мы не умеем извлекать для них экспертное знание ни из специалистов, ни из множества текстов, в которых это знание может быть заключено.

В последнее время сделаны значительные успехи в создании систем управления базами данных (СУБД), систем обработки текста, включая текстовые редакторы и процессоры. Они позволили реализовать такие программы, которые служат путеводителями по большим собраниям текстов в компьютере. Эти системы также имитируют человеческий интеллект при работе с текстом, но значительно превосходят человеческие возможности при охвате гигантских массивов информации.

Социальные последствия новой технологии

Автоматизированные информационные системы воздействуют на общественное развитие, меняя социальный облик труда, в первую очередь, творческого. Но и сами эти системы появились в результате научно-технического прогресса как ответ на общественную потребность в новой технологии переработки информации.

В начале нашего века в структуре занятого населения США люди, занятые переработкой информации (включая образование, науку, культуру, организационное управление, связь), составляли единичные проценты, а к началу 80-х гг. – более половины. Между тем, их техническая оснащенность намного уступала и еще уступает вооруженности рабочих в промышленности и фермеров в сельском хозяйстве. Поскольку в условиях бумажной информатики рутинные интеллектуальные процессы мало производительны, падает и общая производительность труда. Это, в конечном счете, стало основным стимулом информатизации общества.

Возникнув и быстро совершенствуясь, электронные информационные технологии проникают во все поры социальных механизмов и оказывают сильное обратное воздействие на общественное производство. При этом совершенно не обязательно представлять себе функционирование автоматизированных систем лишь как работу их абонентов за дисплеем удаленного от банка данных терминала, при которой они получают необходимую информацию по каналу связи на компьютерный дисплей. Эти системы позволяют выпускать такие информационные издания, которые были невозможны при традиционной технологии. С их помощью можно заглянуть в скрытые от нас до сих

пор глубины социальных процессов и влиять на действие их объективных законов.

Не следует забывать урока, который несколько десятилетий назад преподал нам М. Маклюэн, считавший средства коммуникации определяющим фактором общественного развития. Мы хорошо понимаем, что как ни важны средства коммуникации, они все же играют подчиненную роль. И в информационной сфере смена технологии и технических средств predetermined общественной потребностью. Но это видно тогда, когда мы достаточно удалены от нее. С позиций же сегодняшнего дня, находясь на разломе, будучи участниками этой психологически тяжелой смены, мы склонны фетишизировать технические средства, выступающие для нас зримым символом происходящих перемен.

Для нас, работников умственного труда, наиболее зримым предвестником безбумажного общества служит внедрение компьютеров в издательское дело. На протяжении последних десятилетий электронный набор текста используется для подготовки различного рода изданий: сначала научных и справочных, а теперь и многих массовых, в первую очередь журналов и других периодических изданий. Вначале электронный набор применялся для ускорения их выпуска, а затем дискеты и компакт-диски с машиночитаемой записью текста стали выпускаться как побочный продукт информационной деятельности.

В настоящее время распространение некоторых электронных изданий, например, зарубежных реферативных журналов, превышает тиражи их обычных изданий на бумаге. Увеличивается и число таких электронных изданий, которые не имеют бумажных аналогов. Чаще всего это техническая документация к компьютерным и программным системам, материалы крупных конгрессов и конференций, регулярно пополняемые фактографические базы данных, но все больше и научно-технические журналы. Переход от бумажных изданий к электронным вызывается не только экологическими причинами, хотя они играют существенную роль. Рост информационных потоков и необходимость увеличения производства бумаги ведет к дальнейшей вырубке лесов, дающих атмосфере кислород, и к загрязнению водоемов.

Смысл такого перехода заключается и в том, что рост числа дорогостоящих, неудобных для хранения и использования бумажных изданий создает трудности для передачи информации во времени и в пространстве. Уже сейчас проще найти необходимую специалисту информацию в автоматизированной информационной системе или в Ин-

тернете, нежели во многотомном своде справочного издания или в нескольких сотнях названий специальных журналов. Но ведь мы находимся лишь на пороге новой информационной технологии, и по ее начальным шагам не можем еще представить ее будущих достижений.

Как это обычно бывает при смене информационных технологий, новые средства поначалу копируют старые формы. Известно, что первопечатную книгу старались сделать похожей на рукописную. Современное электронное издание тоже пока сохраняет облик полиграфического, хотя и обладает многими новыми возможностями. В учебнике или монографии можно не только описать какое-либо явление, но и продемонстрировать его динамическую модель. Электронную книгу можно не только читать, но и активно работать с ней, добавляя в нее собственные наблюдения, размышления и результаты экспериментов. В ней могут содержаться видео- и звукозаписи необходимых иллюстраций со стереоскопическими и стереозвуковыми эффектами.

Принципиально новая черта электронных изданий заключается в предоставляемой ими возможности не только получать информацию, но и подвергать ее обработке на месте. Мы говорим, что информацией электронного издания можно манипулировать. Персональный компьютер становится привычным инструментом для представителей многих профессий, но прежде других его оценили ученые. Уже было сказано, что компьютер дал им возможность расширить рамки камерального эксперимента, просматривать под определенным углом зрения огромные массивы информации.

Это меняет не только практику информационной деятельности научного работника, но и социальные условия его труда. Ведь конечным продуктом труда ученого является новая научная информация, получаемая им в процессе переработки результатов исследований, проводимых им и его коллегами. Новая информационная технология сокращает время и затраты труда на получение и обработку данных, позволяет меньшими силами достигать больших результатов, меняет пропорции квалифицированного и вспомогательного персонала в науке.

Компьютер служит современному ученому записной книжкой, дневником и календарем, позволяет составлять отчеты, получать корреспонденцию, проводить заочные конференции с коллегами. Через Интернет он пользуется услугами информационных центров, их базами данных, читает на своем экране полные тексты публикаций статей,

описаний изобретений, стандартов, законов и т. п., а также поддерживает постоянную связь с коллегами в любой стране.

Даже эти немногие примеры, которые в дальнейшем будут пополнены, показывают, что новая информационная технология выходит далеко за рамки научно-технической области, в которой она возникла. Научная коммуникация не является замкнутой социальной системой. Она всегда была тесно связана с такими общественными институтами, как библиотеки, книжные издательства и редакции журналов, радио, телевидение и другие средства массовой информации. Их функционирование с внедрением электроники становится более эффективным.

Но эта проблема имеет и более широкий аспект. Научная и массовая коммуникации взаимосвязаны, они обмениваются своими достижениями. Первые печатные книги были богослужебными, а первые журналы научными, что не мешает им в течение столетий трудиться на общей ниве. Принципы информационного поиска, сформулированные для научных публикаций, успешно используются в газетных банках данных. Нельзя не предвидеть, что электронная информационная технология довольно скоро сделает всю социальную информацию доступной каждому члену общества. Это поведет к важным переменам в образе жизни и мышлении людей. Мы еще не знаем, к каким именно, как не знали этого поколения, выросшие до распространения телевидения. Но теперь мы понимаем, что социальные последствия новой информационной технологии заслуживают внимания и серьезного изучения.

Компьютерные средства коммуникации

Электронные вычислительные машины

Новую информационную технологию справедливо связывают с поистине ошеломляющими успехами в развитии электроники и вычислительной техники. Действительно, эти успехи значительно превосходят и опережают самые оптимистические прогнозы прошлых лет. Именно они позволили существенно продвинуться в автоматизации интеллектуальных процессов. Однако было бы неверно отождествлять развитие информационной технологии только с прогрессом электроники.

Теоретически устройства для переработки информации могут быть реализованы и с применением иных принципов. С другой стороны, далеко не все тенденции развития электронных вычислительных машин совпадают с теми потребностями информационной технологии, о которых мы ведем речь. Дальнейшее уменьшение габаритов компьютеров, увеличение их быстродействия и объемов оперативной памяти вызывается иными сферами их применения, связанными, в частности, с необходимостью управления быстродвижущимися объектами.

Наше внимание, напротив, привлекают далеко еще не освоенные возможности компьютеров в работе с текстами на естественных языках, в переработке данных с целью получения информации и знаний. Поэтому, отдав необходимую дань тому явлению, которое чаще всего называют компьютерной (или, точнее, микропроцессорной) революцией, и определив свое отношение к перспективам развития вычислительных машин, мы сосредоточимся на феноменах "персональных вычислений" и обработки "деловой прозы". Они тесно связаны с широким развитием персональных компьютеров, автоматизированных рабочих мест на их основе, спецификой их программного обеспечения и обработки текстов, в особенности их электронного редактирования.

Быстрое развитие и внедрение компьютеров в различные сферы жизни происходит на протяжении нескольких последних десятилетий.

Мое поколение еще хорошо помнит то время, когда автоматическая переработка информации обсуждалась в фантастических романах. Но тридцать лет это достаточный срок, чтобы привыкнуть к тому, что автоматизация стала необходимой во всех областях письменной коммуникации. Уже сейчас во многих наших учреждениях значительная часть информационной продукции выпускается не только в традиционной, но и в машиночитаемой форме.

Однако большинства людей электронная информационная технология непосредственно почти не коснулась. Они по-прежнему пишут пером или на пишущей машинке, читают книги, журналы и газеты в их привычном бумажном виде, обмениваются информацией по почте, телеграфу или телефону, на экран телевизора смотрят в часы досуга. Хотя, разумеется, компьютер уже не представляется громоздким и дорогим устройством, занимающим обширные залы, стоящим сотни тысяч, а то и миллионы рублей, требующим обслуживания большими коллективами электронных инженеров и программистов.

Это представление уже не подтверждается нашей повседневной практикой ни в одной из информационных сфер, для которых современная информатика служит основным теоретическим базисом. Реальное положение дел в мире коренным образом изменилось, и на сегодняшний день эти представления не соответствуют действительности. Что же касается перспективы, даже самой ближайшей, даже на несколько лет вперед, то она предвещает такие изменения, которых мы пока не можем предвидеть.

Электронная вычислительная машина стала дешевым настольным инструментом, доступным в обращении всем, кто занят переработкой данных. Это оказалось возможным благодаря невиданному по темпам в истории техники развитию электроники, вычислительной техники и программирования. Стало обычным проследивать этапы этого развития по поколениям компьютеров, которые сменялись каждое десятилетие, а теперь почти ежегодно.

Почти каждый этап заслуживает рассмотрения, так как указывает на принципиальные изменения при переходе от одного поколения к другому. Важно обратить внимание на то, что изменения основных параметров компьютеров, их быстродействия и объема оперативной памяти от поколения к поколению менялось на один, а то и на два порядка. Для нас, заботящихся о переработке не только и не столько числовой, сколько текстовой информации, особое значение имеют последние изменения. Переход от третьего поколения машин к после-

дующим сопровождался использованием языков программирования ультравысокого уровня, диалоговым режимом использования компьютера и удобным интерфейсом пользователя: в четвертом поколении – цветным дисплеем с графопостроителем и звуковыми сигналами машины, а в пятом – возможностью устного общения с ней.

Это означает, что пользователь может забыть об устройстве машины и думать лишь о содержании и структуре тех проблем, которые он решает при ее помощи. Если прибегнуть к весьма условному, но часто встречающемуся сравнению с управлением автомобилем, то водителю как бы не нужно думать о порядке вспышек в цилиндрах двигателя, опережении зажигания на форсированных режимах его работы, он может сосредоточиться на маршруте и особенностях дорожной обстановки.

Продолжая эту аналогию, можно сказать, что как в автомобиле стал необязателен шофер-профессионал, так и с компьютером можно управляться без операторов и программистов. По-видимому, нынешний период овладения вычислительными машинами и их развития можно уподобить двадцатым годам автомобилизма, когда автомобиль приобретал современный облик, а промышленность переходила к его массовому выпуску. Только темпы развития и распространения компьютеров намного выше.

Разумеется, деление компьютеров на поколения проводится очень обобщенно и не отражает многих процессов в информационной технологии. Оно удобно лишь для выделения некоторых принципиальных моментов, как в их устройстве, так и в использовании, а главное, в ведущих тенденциях их совершенствования.

Что касается персональных компьютеров, то для нас важно понимать границу между теми, которые были рассчитаны на работу только с дисковой оперативной памятью (ДОС-совместимыми, их последняя версия – АТ 286), и предназначенными для работы с графическими оболочками (Windows-совместимыми, начиная с компьютеров, оснащенных 386 процессором). Наиболее впечатляющим прогнозом развития компьютеров стал японский проект пятого поколения ЭВМ, суливший невиданные ранее возможности обработки данных. Нам уже известен девиз этого проекта: "От обработки данных и информации к обработке знаний". Обсуждали мы и технократическую ограниченность представления о чисто аппаратном решении проблемы доступа к накопленным человечеством знаниям.

Здесь хотелось бы подчеркнуть, что при переходе к пятому по-

колению компьютеров впервые изменился тип их архитектуры, который оставался неизменным на протяжении четырех десятилетий. За это время на четыре порядка (т. е. в 10 тыс. раз) выросли быстродействие и объем оперативной памяти ЭВМ, несколько раз принципиально менялась их элементная база, тип и режим использования машин. Но их архитектура оставалась постоянной – однопроцессорная ЭВМ с последовательным принципом вычислений, восходящая к модели Джона фон Неймана, американского математика венгерского происхождения. Он предложил вводить в машину данные вместе с программой, кодируя и то и другое на одном языке двоичной системы счисления.

Принцип работы такой ЭВМ с одним процессором и одной оперативной памятью показан на рис.11. Необходимо сложить четыре числа. Они помещены в память вместе с текущим результатом, который сейчас равен нулю, и программой действий процессора. За первый такт процессор складывает первое число (3) с текущим результатом (0) и получает новый результат (3), который помещается в память. За второй такт процессор складывает второе число (15) с текущим результатом (3) и получает новый результат (18), который помещается в память. За третий такт процессор получает новый текущий результат ($22 = 4+18$), который снова помещается в память. За четвертый и заключительный такт процессор получает новый результат (43), который является ответом. Если этот процессор должен сложить тысячу чисел, необходима тысяча тактов ЭВМ.

Новые поколения компьютеров

В пятом поколении ЭВМ предполагалось применить многопроцессорную машину параллельной архитектуры. Принцип работы одного из вариантов суперкомпьютера такой архитектуры – системы потока данных – иллюстрируется на схеме, составленной директором лаборатории вычислительной техники Массачусетского технологического института М.Л.Дертузосом (рис.12).

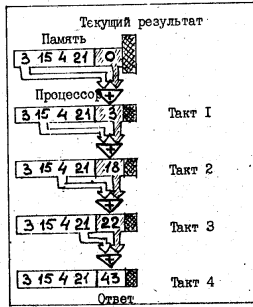


Рис. 11. Модель Дж.фон Неймана: однопроцессорный компьютер.

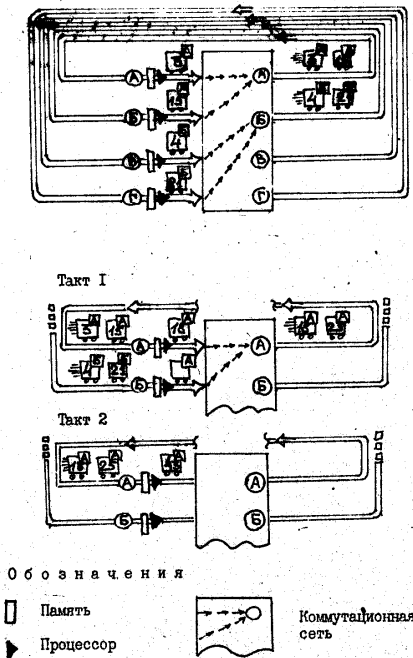


Рис. 12. Многопроцессорный компьютер параллельной архитектуры.

Особенностью этой системы является наличие нескольких комбинаций процессора и оперативной памяти, каналов "потока данных" и коммутационной сети между ними. Появляющиеся возможности технической реализации такой системы обеспечивает одновременность выполнения операций и ускорение процесса вычисления. Для наглядности схема воспроизводит тот же простой пример сложения четырех чисел, каждое из которых как бы помещено в тележку, движущуюся по рельсам.

В системе потока данных на схеме каждая из четырех комбинаций процессора и памяти присваивает числу (внутри тележки) направление (на флажке сбоку). Тележки движутся по коммутационной сети как на железнодорожных стрелках. Сеть начинает операции, направляя 3 и 15 в А, а 4 и 21 в Б, где они и помещаются в соответствующие оперативные памяти. (Таким образом, комбинации памяти и процессора В и Г в вычислениях участвовать не будут). За первый такт программы в памяти А и Б предписывают соответствующим процессорам сложить числа и получить результаты, 18 и 25. Эти числа направляются коммутационной сетью в А. (Память, процессор и канал потока данных Б больше не потребуются). За второй такт, попав снова в процессор А, числа 18 и 25 складываются, образуя ответ 43.

Таким образом, имея 4 процессора, можно одновременно сложить 8 чисел за 3 такта или, другими словами, скорость вычислений в такой системе теоретически возрастает как экспонента к числу процессоров. ЭВМ Крэй-2 (США) имел четыре процессора, Крэй-3 – шестнадцать, а в Манчестерском университете проектировалась вычислительная машина с 256 процессорами. При этом, разумеется, изменение типа архитектуры было не единственным способом совершенствования компьютеров на пути к пятому поколению, а описанная модель их новой архитектуры являлся лишь одним из возможных вариантов.

Впервые идею о построении многопроцессорной ЭВМ высказали советские ученые в середине 70-х годов. Академик В. М. Глушков в одной из последних своих бесед, будучи уже тяжело больным, предложил принцип вычислений на ЭВМ, который он назвал "макроконвейерным". Он уподобил работу обычного однопроцессорного компьютера заводскому конвейеру. Макроконвейер может быть организован так, что несколько заводов кооперируются, изготавливая на своих конвейерах разные детали, которые затем собираются в готовое изделие. Точно так же, по его мнению, можно решать сложную вычислительную задачу параллельно на нескольких процессорах, если один из них

обеспечивает управление остальными. Эта идея сейчас успешно реализуется.

Характеризуя пятое поколение компьютеров, обычно говорят об их быстрой скорости в сотни миллиардов операций в секунду и объеме памяти в миллиарды байт, о том, что они воспринимают и выдают информацию в форме устной речи, распознают и отождествляют трехмерные цветные изображения, моделируют рассуждения специалистов в узких предметных областях. Какое же конкретное применение находят эти машины в жизни?

Большинство зарубежных специалистов считает, что самыми распространенными сферами использования компьютеров нового поколения являются промышленное производство и делопроизводство, наука и техника, конструирование и программирование вычислительных машин, авиация и космонавтика, военное дело, сфера торговли и услуг, образование, здравоохранение, искусство и культура. Значительная часть необходимых технических и программных средств уже создана и работает в промышленном режиме. Для массового их производства и внедрения требуются экономичные технологии и серьезные мероприятия по подготовке пользователей.

В промышленности компьютеры нового поколения позволяют создать полностью автоматизированные производства, управление которыми централизовано и осуществляется на уровне заданий по ассортименту, количеству и качеству изделий. В делопроизводстве интегрируется хранение, поиск и распространение служебной документации с организационным управлением и средствами коммуникации. Получают распространение телеконференции.

В науке, технике, медицине, авиации, торговле и сфере услуг повышается эффективность справочно-информационных систем, которые позволяют абонентам на рабочем месте перерабатывать в нужном аспекте полученную информацию. Экспертные и особенно диагностические системы достигают в этих сферах высокой степени интеллектуализации и занимают важное место в структуре трудовой деятельности ученых, инженеров, врачей и специалистов многих других отраслей народного хозяйства.

Образование и игры составляют особую сферу применения компьютеров, особенно персональных, на которую в настоящее время падает значительная доля их сбыта. Важную роль здесь играет совмещение цветного и объемного изображения со звуком и возможностью манипулировать текстом. Учебный процесс становится активным и

динамичным. Каждый может преобразовать учебник под свои нужды. Сведения сообщаются не только в виде текста, но могут иллюстрироваться изображениями процессов и сопровождающих их звуков. Обучение сочетается с творчеством, отдых с просвещением. Вы можете, не выходя из дома, совершать путешествия, посещать музеи, обучаться вождению автомобиля, самолета, космического корабля и т. п.

Персональный компьютер и персональные вычисления

Одна из статей о персональных компьютерах в научно-популярном журнале начиналась со следующей аналогии. Если бы за последние четверть века самолетостроение развивалось в том же темпе, что и производство вычислительных машин, то пассажирский лайнер был бы в цене телевизора, и на нем можно было бы облететь земной шар за полчаса с одной канистрой горючего. Эта аналогия, хотя и неточная, как всякая другая, хорошо подчеркивает фантастичность возможностей, открываемых ныне вычислительной техникой для личных надобностей каждого человека в переработке информации.

Появление персонального компьютера уже обросло типичной американской легендой. В ней рассказывается о том, как в начале 70-х годов два молодых калифорнийца, инженер Стивен Джобс и программист Стефен Возняк, запершись в своем гараже, сконструировали простой в употреблении домашний компьютер для хозяйственных и деловых нужд. Это и положило начало знаменитой фирме "Эппл" с миллиардными доходами и собственной линией развития персональных компьютеров.

Разумеется, в действительности все было не так просто. Однако, глядя на фотографию их популярной модели "Лайза II", трудно отделаться от впечатления, что они изобрели портативную пишущую машинку, которая подключается к бытовому телевизору и магнитофону и позволяет писать на экране вместо бумаги. Возможно, что это чисто внешнее "обытовление" компьютера, незаметность самого микропроцессора, встроенного внутрь привычных в домашнем обиходе устройств, сыграло определенную роль в преодолении психологического барьера, отделявшего прежде компьютер от его пользователей.

Что же представляет собой персональный компьютер? Академик А. П. Ершов определял его как массово выпускаемую организованную совокупность средств ввода, обработки, хранения, передачи и воспроизведения информации, находящуюся в полном распоряжении своего пользователя. Эта совокупность включает микропроцессор, ос-

новную память (постоянную и оперативную), внешнюю память (накопители на жестком или гибких магнитных дисках), дисплей (монитор), клавиатуру, а также устройства для печати текстов и изображений (принтер) и для связи с другими компьютерами по телефонным каналам (модем). В дешевых компьютерах дисплей и дисковод прежде могли заменяться бытовым телевизором и кассетным магнитофоном, а принтер – электрической пишущей машинкой.

Стоимость персонального компьютера непрерывно снижается и соразмерна со стоимостью такого бытового прибора, как телевизор. Поскольку дальше дается краткая характеристика каждого из основных устройств персонального компьютера, хочу оговориться, что к приводимым мною данным нужно отнестись критически. Они быстро стареют, поскольку электронная техника развивается стремительно, а издательские процессы в нашей стране пока еще неторопливы.

Микропроцессор, управляющий работой всех остальных компонентов, является центральным (хотя и малозаметным) устройством компьютера. Его вычислительная мощность характеризуется разрядностью, т. е. размером информационного слова, как бы задающим ширину тракта передачи данных, и частотой тактового генератора, обеспечивающего скорость каждого шага выполняемых машиной операций. Первые персональные компьютеры имели 8-разрядный процессор, т. е. работали со "словом", равным 8 битам (или 1 байту). Другими словами, они передавали за один такт одну букву или две цифры. Сейчас распространены 16-, 32-, 64-разрядные микропроцессоры, а тактовая частота их работы за несколько лет возросла от 12 Гц до 2 МГц (млрд. периодов в сек.).

Основная память состоит из постоянного запоминающего устройства и оперативной памяти (с произвольной выборкой). В постоянной памяти навсегда записаны самые основные программы, которые работают сразу после включения компьютера, преобразуют в коды команды, связанные с нажатием определенных клавиш. По мере увеличения объема основной памяти в ее постоянную часть стремятся записать все большее число системных программ.

Запоминающее устройство с произвольной выборкой (оперативная память) служит для записи в основную память операционной системы, транслятора и прикладных программ, необходимых в данный момент для выполнения работы. Они считываются из внешней памяти накопителей на магнитных дисках. Минимальный объем основной памяти в персональном компьютере – 64 Кбайт (что составляет

64x1024 байт, или 65536 алфавитных знаков, включая пробелы, или 36,4 машинописных страниц, или 1,6 авторского листа). Стандартные объемы оперативной памяти современного персонального компьютера варьируют от 16 до 512 Мбайт.

Внешняя память персональных компьютеров первоначально выполнялась на стандартной магнитофонной ленте в кассетах. Теперь она реализуется накопителями на магнитных дисках, которые подразделяются на жесткие и гибкие. Жесткие (твердые) диски обычно встраиваются в один корпус с процессором, являются несъемными и обеспечивают высокую плотность записи и объем хранимой информации от нескольких сотен Мбайт до нескольких десятков Гбайт. Их обычное название "винчестер" объясняют тем, что их первоначальное устройство – два диска по 80 Кбайт – напоминало английскую двустволку такого же калибра, а по другой версии, что технология их изготовления была разработана в г. Винчестере.

Гибкие получили название дискет (флоппи-дисков), представляют собой сменные магнитные диски размером 3,5 дюйма в пластмассовом корпусе. Они могут хранить файлы общим объемом до 1,4 Мбайт. Имеются специальные накопители на сменных дисках большего объема – от 100 Мбайт до нескольких гигабайт. Однако они быстро вытесняются более дешевыми компакт-дисками CD-ROM с возможностью записи (R) и перезаписи (RW) объемом до 700 Мбайт.

В настоящее время в качестве внешней памяти широко используются именно эти оптические компакт-диски, позволяющие хранить десятки Гбайт информации. Они представляют собой две склеенных стеклянных пластины диаметром 30 см., на внутренние поверхности которых нанесен тончайший слой теллурического сплава. Лучом лазера на это покрытие наносится свыше 30 тыс. спиральных дорожек, разделенных на сектора для облегчения к ним доступа. На этих дорожках затем выплавляются углубления изменяемой величины (около микрона) в зависимости от интенсивности лазерного луча, которая модулируется сигналами компьютера. Так производится первоначальная запись данных.

Считываются эти углубления лучом меньшей мощности, отражение которого воспринимается фотоэлементом, преобразующим световые импульсы в электрические сигналы. Обычные компакт-диски для компьютеров штампуются как аудио- или видеодиски, так что выведенную на них информацию изменить нельзя. Записываемые и перезаписываемые на компьютере диски устроены более сложно и их

тиражирование значительно дороже.

Дисплей (экран телевизора или монитора) – основное устройство отображения информации, выводимой во время компьютера. Дисплеи бывают монохромными и цветными, отличаются по размерам, форматам (числу строк и символов в строке), разрешающей способности (числу светящихся точек по горизонтали и вертикали), числу уровней яркости или цветов. Обычные дисплеи позволяют одновременно видеть на экране лишь часть страницы текста. В монохромных дисплеях оптимальным является адаптер "Hercules", в цветных последовательное повышение качества реализуют CGA, EGA, VGA и SVGA (Color-, Extended-, Video-, Super Video-Graphic Adapter).

В настоящее время все большее распространение получают жидкокристалльные дисплеи, которые сначала встраивались в ноутбуки, а теперь применяются вместе с настольными компьютерами. Они занимают меньше места, потребляют меньше энергии, лучше используют поверхность экрана, у них меньше излучение, нет не замечаемого, но вредного для зрения подрагивания изображения. Но пока они вдвое дороже мониторов с лучевой трубкой.

Клавиатура служит средством ввода информации и управления работой компьютера путем нажатия клавиш, которые подразделяются на алфавитно-цифровые, функциональные, редактирующие и управляющие. Алфавитно-цифровые клавиши располагаются так же, как на пишущей машинке (по прежнему стандарту буквы латинского алфавита располагались по транслитерационному принципу, т. е. А-А, В-В, С-С и т. д.). В отличие от терминалов больших вычислительных машин клавиатура персональных компьютеров передает в микропроцессор не код символа, а порядковый номер и интервал длительности нажатия клавиши.

Функциональные клавиши могут менять значение других, управляют системными программами, вызывают на экран стандартные элементы графики. Редактирующие клавиши управляют движением курсора (движущейся по экрану световой точки, прямоугольника или мигающей черты), который обозначает позицию очередного символа. Курсором можно управлять также при помощи специальных устройств – "мыши" (полусферы, передвигаемой рукой по поверхности планшета), "джойстика" (используемого в динамических компьютерных играх) или "светового пера" (позволяющего менять изображение путем прикосновения к любой точке экрана).

Принтеры (печатающие устройства) служат для вывода инфор-

мации на бумагу. По способу действия они делятся на матричные (головка с иглами, управляемыми матрицей знаков), ромашковые (сменные лепестковые шрифтоносители), термографические (тепловое воздействие на специальную бумагу), струйные (распыление струи специальных чернил), лазерные (оптическая печать лучом лазера). Матричные принтеры первоначально получили наибольшее распространение, так как они быстрее ромашковых (до 200 зн/сек против 50 зн/сек) и позволяли печатать не только текст, но и любые изображения, неприхотливы к качеству бумаги и значительно дешевле других. В настоящее время наиболее распространены лазерные и струйные принтеры. Последние дешевле лазерных при покупке и в эксплуатации при почти таком же качестве печати и незначительно меньшей скорости.

Важной характеристикой персонального компьютера является надежность. Он предназначен для устойчивой работы в бытовых условиях, без особых требований к температуре, влажности и чистоте окружающего воздуха, к колебаниям напряжения или выключению тока в электросети. При последнем, правда, уничтожается содержание оперативной памяти (т. е. все наработанное с момента включения компьютера, если его периодически не сохранять). И при этом персональный компьютер обладает почти всеми возможностями большой вычислительной машины, для которой упомянутые требования крайне существенны.

Нельзя не упомянуть о том, что появление персональных компьютеров не было предвидено профессионалами. Фирма IBM в начале 70-х гг. заказала мозговому тресту "RAND Corporation" прогноз развития ЭВМ, который стоил десятки миллионов долларов. В нем не было упоминания о такой возможности, хотя Джобс и Возняк, вероятно, уже работали над своей "Лайзой". Идеологи отечественной информатики вынуждены были признать, что персональный компьютер вошел в вычислительное дело с черного хода и взломал сложившиеся представления и в технологии и в программировании. Однако в настоящее время ряд компьютерных фирм разрабатывает идею « сетевого компьютера », который противопоставляется персональному, поскольку будет выполнять роль удаленного терминала больших машин в сети вычислительных центров с распределенным банком данных.

Появление и широкое распространение персональных компьютеров – явление революционное, вызвавшее к жизни феномен персональных вычислений. Он заключается в том, что конечные пользователи – специалисты в самых разнообразных отраслях науки, техники,

народного хозяйства, культуры получили непосредственный доступ к машине, перестали нуждаться в профессиональных программистах. Существовавшее до сих пор положение, когда специалист, решавший свои задачи на машине, должен был объяснять их программисту, сильно сдерживало применение ЭВМ.

Как правило, существо решаемой задачи и ее особенности полностью раскрываются в процессе решения. Поэтому предварительная постановка задачи перед программистом всегда отличается неточностью и приблизительностью. Программист же, формализуя и уточняя задачу, стремится к краткости и изяществу программы, к экономии вычислительных ресурсов машины, пренебрегая деталями задачи, которые могут быть важными для ее решения. Поэтому неуклюжие и громоздкие программы, составленные конечными пользователями, часто отличаются от стройных и лаконичных профессиональных программ тем, что они работают более эффективно, лучше решают задачи специалистов.

Таким образом, феномен персональных вычислений заставляет нас по-новому взглянуть на проблему взаимоотношений машины и пользователя, которая существенна для решения таких глобальных задач, как всеобщая компьютерная грамотность или информатизация общества. Решение этих проблем, ставших в полном смысле социальными, требует не только и даже не столько овладения средствами вычислительной техники и их программирования. Они вызывают необходимость пересмотра отношения специалистов к существу своего дела и решаемых в нем задач. Ведь их решение при помощи компьютера вынуждает формализовать эти задачи, искать алгоритмы выполнения многих процессов, включая и интеллектуальные.

Это новое явление Г. Р. Громов удачно назвал автоформализацией профессиональных знаний, и оно намечает тот путь, по которому информатика проникает во многие другие области знаний¹. Особая ситуация складывается в программировании. Дело в том, что овладение основами информатики часто понимают как обучение началам программирования на одном из простых языков, обычно на Бейсике.

Это создает у пользователя иллюзию умения программировать и подчас приводит его к большим затратам времени и сил с минимальным результатом при решении вычислительных задач, для которых

¹ Громов Г. Р. Национальные информационные ресурсы. – М.: Наука, 1984. – С. 52.

имеются хорошие профессиональные программы. Поэтому компьютерная грамотность предполагает в первую очередь хорошее знание всех типов прикладных программ, умение выбрать те конкретные программы, которые адекватны решаемой задаче, и навыки в работе с ней. Тем не менее, общее представление о языках программирования и их эволюции необходимо каждому, приобщающемуся к интеллектуальной коммуникации.

Работа с текстом на компьютере

О языках программирования

Когда мы по-русски говорим "компьютер", то представляем себе некоторое устройство, которое нужно запрограммировать, чтобы оно заработало. Американцы, которые, естественно, думают на своем английском, представляют себе компьютер сразу состоящим из двух частей – "твердой" ("хардуэр") и "мягкой" ("софтуэр"), что на русский язык казенно переводится как аппаратные и программные средства вычислительной машины. Это различие в понимании отражает и некоторую реальную ситуацию: у нас первая составляющая компьютера стоит во много раз дороже второй. У них наоборот. Умственный труд они ценят много дороже промышленного, что обычно для развитых стран.

В наши дни особое внимание уделяют составлению текстовых описаний к новым программам, поскольку освоение этих программ требует большого труда и занимает много времени. Возникла насущная проблема составления таких описаний к программам, которые были бы понятны обычному или даже начинающему пользователю. Она достигла столь большой остроты, что появилась специальная профессия "технических писателей", так как программисты не могут понять, что именно затрудняет пользователей при освоении их программ.

Не будучи специалистом, трудно говорить о вычислительных машинах, но еще труднее об их программировании. Дело в том, что сама сущность программирования, кажущаяся такой простой на элементарном уровне машинных кодов, оказывается крайне сложной для формулирования. Может быть, поэтому авторы, популяризирующие искусство (или науку, как полагают некоторые) программирования, так любят прибегать к аналогиям, которыми сами почти всегда остаются недовольны.

Сравнение компьютера с автомобилем, а программирования – с управлением механистично и мало что объясняет, компьютера с теле-

визором, а программы – с передачей, которую он принимает, тоже отражает лишь внешнюю сторону дела. Пожалуй, ближе всего к существу отношений компьютера с его программой сравнение их с музыкальным инструментом и нотами. Действительно, комбинацией нескольких нотных знаков можно записать любое музыкальное произведение, а пределов для творчества композиторов и исполнителей не существует. Так и программа, написанная на одном из специальных языков, в котором используется ограниченный набор символов, служит для решения многих творческих задач, от простых вычислений до игры в шахматы. Но и здесь сходство сводится к идеальности и многообразию содержания информации и материальности и ограниченности средств ее фиксации и воспроизведения.

Любая, самая сложная программа формальна и определяется алгоритмом решения задачи. Вместе с тем, возможность или границы формализации человеческого мышления, которое стремятся моделировать при помощи компьютера, пока еще не поддаются определению. В преодолении этого противоречия, в нахождении путей его разрешения скрыт важный фактор дальнейшего развития данной проблемной области. По всей вероятности, задача такого уровня должна решаться не одними программистами и информатиками, но всем научным сообществом. Это один из серьезных стимулов достижения всеобщей компьютерной грамотности. А. Кэй, из статьи которого взяты некоторые сведения о языках программирования, писал об этом в следующих выражениях:

"Грамотность применительно к компьютерам – это проникновение в суть процессов программирования, причем достаточно глубокое, чтобы было легко и приятно производить вычисления и решать различные задачи на ЭВМ, подобно тому, как истинно грамотный человек может свободно читать и писать. Как и в любом искусстве, здесь необходимо хорошо знать и любить материал. Если мы полагаем важным овладеть литературой и искусством для развития человека и общества, то имеем ли мы право меньше усилий тратить на то, чтобы сделать вычислительные средства и методы частью нашей жизни"¹.

История современных языков программирования насчитывает всего лишь пять десятилетий, если считать от момента создания языка *Планкалькюль* немецким инженером К. Цузе в 1946 г. С тех пор появились сотни языков программирования, а с вариантами, возможно, ты-

¹ Кэй А. Программное обеспечение ЭВМ // В мире науки. – 1984. – № 11. – С. 13.

сячи. Конечно, не все они употребляются на практике, и если бы их развитие шло другим путем, их могло бы быть и меньше. Но их многообразие оправдано, так как нельзя создать язык, пригодный для всех случаев. Язык должен быть пригоден для решения данной задачи на данном компьютере и удобен для программиста.

При всех различиях языков программирования все они, в конечном счете, сводятся к высоким и низким уровням электрического напряжения, соответствующим единицам и нулям двоичного кода. Комбинации единиц и нулей могут интерпретироваться компьютером как адрес в его памяти, фрагмент обрабатываемых данных или команда выполнить определенное действие. Язык самой машины – машинный код – очевиден и однозначен для каждого вида компьютеров. Но эта последовательность нулей и единиц противоестественна для памяти человека, хотя несколько поколений программистов работали только в машинных кодах. Программа может состоять из миллионов комбинаций единиц и нулей, и ошибка лишь в одном знаке приведет к неправильной работе всей программы.

Это повело к созданию языков ассемблера, в которых двоичные последовательности были заменены символами и словами. Первым таким языком был ассемблер компьютера "Эдсак", созданный в 1949 г. английским математиком М. Уилксом. Разумеется, ассемблеры проще и mnemonicнее машинных кодов, в них можно создавать инструкции для определенных действий. Однако каждая команда должна определяться отдельно и за ее прохождением в машине нужно следить. Кроме того, ассемблер жестко связан с типом компьютера, для которого он создан.

Ограниченность ассемблеров была преодолена в языках более высокого уровня. Первым распространенным языком высокого уровня явился *Фортран*, разработанный в 1954–1957 гг. Дж. Бекусом в фирме ИВМ и до сих пор применяемый для научных расчетов. Поскольку его применение занимало много времени "тихоходных" в то время машин, для него был создан компилятор, переводивший его программы в машинный код. Это положило начало особым видам программ – т р а н с л я т о р а м, которые для каждого типа компьютеров переводят программы, написанные на языке высокого уровня, в их машинный код.

Они бывают двух видов – компиляторы и интерпретаторы. При использовании к о м п и л я т о р а весь цикл трансляции завершается до начала выполнения программы, так что прохождение этой про-

граммы требует меньше времени. Программа, составленная на интерпретаторе, записывается в оперативную память в виде команд на языке высокого уровня, каждая из которых транслируется в машинный код по мере прохождения программы. Это требует больше времени, но дает оператору возможность контролировать результат каждой операции.

Все исследования в области программирования с 1957 г. стимулировались стремлением усовершенствовать Фортран, который подвергался неоднократным переделкам. Стало появляться большое число языков, что беспокоило программистов, которые создали Международный комитет для выработки единого языка программирования (КОДАСИЛ). Результатом работы комитета явились два языка, до сих пор широко распространенные у пользователей, эксплуатирующих большие машины. Один из них, *Кобол* был разработан для выполнении однородных операций над большими массивами чисел. Он применяется в сфере бизнеса и эффективен при простых арифметических вычислениях. Две трети конторских программ на Западе все еще написаны на Коболе. Но как универсальный язык, к которому по тем временам все стремились, он был неудовлетворителен. Вторым языком стала первая версия *Алгола* (1956 г.), который хотя и не стал желанным универсальным языком, но в последующих версиях сохраняет значение многоцелевого средства.

Тогда же Дж. Маккарти из Массачусетского технологического института создал язык *Лисп*, в котором программы и данные представляются в виде списков. Его простота и сила определяются тем, что в нем применяется только один вид команд – вызов функции, а ее значением может быть другая функция. С тех пор он стал популярным в исследованиях по искусственному интеллекту и послужил моделью для многих других языков. Для персональных компьютеров чаще всего используется *Бейсик*, разработанный в 1965 г. Дж. Кемени и Т. Курцем из Дартмутского колледжа в США. Он предназначался для изучения вводного курса программирования, но затем стал популярным среди программистов-непрофессионалов. Каждая строка программы на этом языке обозначается номером, а управление прохождением программы в основном осуществляется путем указания этих номеров.

Близок к нему и язык *Паскаль*, также первоначально созданный для обучения студентов в 1970 г. Н. Виртом, преподавателем Федерального технологического института в Цюрихе. В отличие от Бей-

сика в нем необходимо объявлять каждую переменную и указывать ее тип, а процедуры и функции обозначать не номерами строк, а именами функций, что облегчает чтение программы. Решая какую-либо задачу, специалист не может оперировать двоичными кодами, регистрами и адресами машинной памяти. Он мыслит формулами своей науки, например,

"Площадь = длина X ширина",

"Прибыль = доход – затраты".

Действия, описываемые этими формулами, транслируются в машинные коды компиляторами и интерпретаторами, которые можно считать средствами реализации языка программирования. На этом и были основаны такие языки программирования, как Фортран, Паскаль, Бейсик.

Эти и подобные им языки называются *процедурными*. Каждый такой язык как бы предоставляет программисту некую виртуальную (воображаемую) машину. Составляя программу, пользователь видит решение своей задачи как процесс вычисления, осуществляемый этой виртуальной машиной. Тем самым сразу описывается и задача, и метод ее решения. Появились непроцедурные стили программирования – функциональный, логический – и соответствующие языки. Основная идея логического программирования – отделение описания задачи от процесса ее решения – была выражена в уравнении одного из его авторов Р. Ковальского:

"Алгоритм = логика + управление".

Наиболее распространенный язык логического программирования *Пролог* был создан в 1972 г. А. Колмари из Марсельского университета. Программа на Прологе состоит из описаний, задающих объекты и отношения между ними. Например, формула

"Площадь = длина X ширина"

описывается отношением *"Произведение, длина, ширина, площадь"*. Описание это статическое и никаких вычислений оно не задает. Но компьютер производит их сам по запросу, и по любым в данном случае двум величинам может вычислить третью, т. е. по площади и ширине – длину. Независимо от зарубежных работ в СССР в 70-е годы В. Б. Борщевым и М. В. Хомяковым (ВИНИТИ) также разрабатывалась версия логического программирования.

Многие современные компьютерные системы используют язык *Си*, созданный в 1972 г. Д. Риччи в лабораториях "Белл Телефон" корпорации АТТ в Мюррей-хилл (шт. Нью-Джерси). Первоначально он разрабатывался для программирования новой операционной системы Ю н и к с. Операционная система предназначена для приема, хранения и выдачи информации, а также для взаимодействия аппаратных средств компьютера с прикладными программами. *Си* стал популярным как язык так называемого среднего уровня. В нем удобство, краткость и мобильность языков высокого уровня сочетаются с возможностью непосредственного доступа к машине, что традиционно обеспечивалось ассемблером, т. е. языком низкого уровня.

Другая важная тенденция в развитии программирования связана с объектно-ориентированными языками. В них процессор машины условно делится на "объекты", которые могут программироваться индивидуально, а затем соединяться друг с другом при помощи сообщений. В числе подобных языков: *С и м у л а - 67* (О. Даль и К. Нигард из Норвежского ВЦ в Осло), *С м о л л т о к* (1970 г., А. Кей из фирмы Ксерокс, Пало-Альта), *Э у р и с к о* (1979 г., Д. Ленат из Стэнфордского университета).

Прикладные программы подготовки текстов

Среди профессиональных программистов бытует мнение, что пользователям-непрограммистам не следует увлекаться программированием, а лучше применять уже готовые программы. Это отчасти противоречит высказанной в прошлой лекции идее о необходимости формализации специальных знаний и том преимуществе, которое получают специалисты различных областей знания после овладения навыками работы с персональным компьютером. Но есть в этом мнении и рациональное зерно. Оно заключено в том, что в настоящее время разработано множество эффективных программ для решения самых разнообразных задач, и неразумно начинать по-дилетантски составлять программу, не убедившись, что она существует. Еще справедливее это в отношении обучения основам информатики, когда, начиная со школы, учат только алгоритмическому мышлению и одному из простейших языков, обычно Бейсику, и не дают представления о богатстве существующих прикладных программ.

Программное обеспечение вычислительных машин можно условно разделить на системное и прикладное. К *системному* относятся операционные системы, о которых говорилось выше, и средства диаг-

ностики и контроля. *Прикладное* – это пользовательские программы решения разнообразных задач, которые объединяются в совокупности по классам решаемых задач и называются пакетами прикладных программ. К ним примыкают также библиотеки стандартных программ, которые часто используются для вычислений, решения уравнений и операций обработки данных – их сортировки, копирования и т. п.

Для нас наиболее важными являются те пакеты, которые применяются для подготовки текстов. Они получили название *текстовых редакторов*, но часто называются также текстовыми процессорами, системами обработки (подготовки) текстов. Текстовому редактору всего два с небольшим десятилетия. Его создателем считается М. Шрейер, кинорежиссер из Нью-Йорка, который в 1975 г. реконструировал компьютер "Альтаир" и составил программу "Электрический карандаш", для подготовки руководств к своим программам. В 1978 г. нью-йоркские программисты С. Рубинштейн и Дж. Барнэби создали текстовый редактор "УордСтар", который определил стандарт таких программ для персональных компьютеров.

Ввод текста при помощи такой программы осуществляется как на пишущей машинке: буква за буквой с пробелами между словами, а в конце строки (в современных программах – абзаца) нажимается клавиша возврата каретки (VK, Ввод, Enter). Никакой каретки, конечно, в компьютере нет, а под воздействием этой клавиши производится ввод набранной строки в память, а курсор (светящаяся или мигающая точка на экране) перемещается в начало следующей строки. Если набранный текст нужно исправить или отредактировать, курсор можно вернуть в любую его точку нажатием одной из редактирующих клавиш (управляющих команд): влево/вправо на одну букву, слово или в начало/конец строки, вверх/вниз на одну строку или же в начало/конец всего текста (см. рис 13).

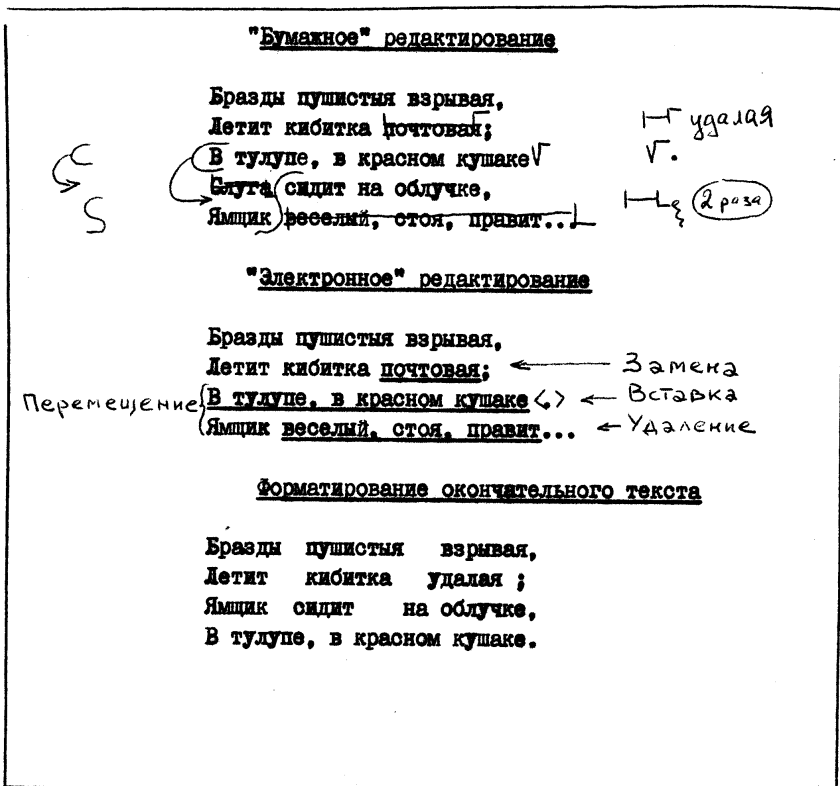


Рис. 13. Процедуры исправления текста корректурными знаками и в текстовом редакторе компьютера (пример из черновой рукописи А. С. Пушкина).

Исправление производится при помощи одной из процедур: удаления, перемещения, вставки, замены. При удалении необходимо отметить начало и конец удаляемого фрагмента, программа по команде стирает слово и убирает пробел, сдвигая оставшуюся часть строки влево. Чтобы переместить фрагмент текста, нужно тоже его отметить, установить курсор в то место, куда фрагмент перемещается, и дать команду, по которой программа изменит порядок считывания текста и соответственно перестроит его на экране.

Вставка и замена это не только процедуры исправления, но и *режимы* работы текстового редактора, в одном из которых программа работает обычно ("по умолчанию"), а другой вводится специальной клавишей. В режиме вставки часть строки, начинающаяся с курсора, при нажмие буквенной клавиши отодвигается вправо, освобождая эту позицию для вводимой буквы. В режиме замены вводимая буква забывает прежнюю. Так же работают в этих режимах и клавиши стирания букв – со смыканием текста в позиции стертой буквы или же с заменой ее пробелом.

Здесь приводятся лишь самые элементарные процедуры электронного редактирования. Возможности современных текстовых редакторов многообразны и зависят лишь от объема программы, соотносимой с размерами оперативной памяти компьютера. В их возможности входит дублирование строк и фрагментов текста, запоминание и выдача в нужный момент стандартных текстов любого объема, набор текстов в нескольких окнах с последующим их совмещением в произвольном порядке, запоминание исходного (не редактированного) текста и многие другие процедуры.

По окончании редактирования текст может быть отформатирован, т. е. выровнен по правому и левому краям, напечатан целиком или частями с отступом на любое число позиций. Для просмотра текста на экране его можно сдвигать вверх или вниз построчно или постранично. Обычно на экране размещается до 25 строк (одна из них служебная) по 64 или 80 знаков в строке. Но имеются и такие редакторы, которые позволяют "верстать" широкую полосу в две колонки по 60 знаков с пробелом, сдвигать текст на экране вправо/влево. Большинство программ осуществляют функцию поиска отдельных слов и словосочетаний по всему тексту с их заменой на другие слова и словосочетания.

Строго говоря, текстовый редактор служит для ввода текста в компьютер, а для вывода его на принтер служит другая программа системы подготовки текста - ф о р м а т и з а т о р. Эта программа предна-

значена для оформления текста в соответствии с нормами, сложившимися в полиграфии, т. е. выравнивания правого края, абзацных отступов, центровки заголовков, простановки номеров страниц, установки расстояний между строками. По мере совершенствования принтеров и форматизаторов стало возможно разнообразить шрифты, менять ширину отдельных букв, печатать подстрочные и надстрочные индексы и формулы в математических текстах, таблицы, верстать страницу с многоколонным набором и включением иллюстраций. Особая задача форматизатора – автоматически осуществлять перенос слов в соответствии с грамматическими правилами. Все это позволяет приблизить набранный на компьютере текст к полиграфическому и в каждый момент его написания и редактирования иметь его полностью готовым к печати.

Все большее распространение в системах подготовки текстов получают программы, позволяющие непосредственно производить автоматическую корректуру – а в т о к о р р е к т о р ы (спеллеры или спеллчекеры). Первоначально они строились на статистической основе и давали возможность после набора всего текста вывести на экран слова, в которых сумма номеров букв оказывалась единичной. Такие слова, один раз встретившиеся в тексте, чаще всего содержали ошибки. Этот метод, не требовавший большой оперативной памяти компьютера, не был особенно удобным и удовлетворительно работал только на сравнительно больших текстах. Поэтому с увеличением оперативной памяти он постепенно был заменен сначала полиграммным, а затем словарным методом.

Полиграммный метод основан на том, что все двух- и трехбуквенные сочетания проверяются на их допустимость в данном языке. Если в тексте встречаются слова с недопустимыми би- и триграммами, то они выводятся на экран как сомнительные. При словарном методе все набираемые слова сразу или после набора всего текста сопоставляются с находящимся в памяти машины словарем, объем которого достаточно велик (обычно от 40 до 120 тыс. слов). Понятно, что такая программа должна учитывать всю парадигму словоизменения в данном языке, что гораздо проще сделать в английском, нежели в русском языке. Автоматизация корректуры может состоять из обнаружения ошибок, предложений по их исправлению или же собственно их коррекции. Понятно, что две последние возможности необходимы людям, неуверенным в своей грамотности, и требуют усложнения и значительного увеличения программы.

Однако, стремясь сделать подготовку текста более комфортной, а сами тексты более совершенными, а также учитывая возросшую потребность в составлении текстов на неродных для пишущего языках, составители программ идут на эти усложнения. Начали появляться программные системы, в которых словари содержат не только средства нормализации слов, но и такие смысловые связи между ними, как синонимия, омонимия, полисемия, антонимия, отношения "род-вид", "часть-целое" и т. п. Справочники, содержащие подобные сведения, составляются уже давно, применяются в информационном поиске и получили название т е з а у р у с о в. Они позволяют при написании текста на компьютере вести стилистическую правку, а главное, предлагают автору выбор отдельных слов и выражений, продолжение фразы в стандартных оборотах, подсказывают устойчивые словосочетания, управление глаголов и предлогов и оказывают другие лексические услуги.

Существуют сотни всевозможных программ с редакторами текстов, пакетов прикладных программ подготовки текстов с редакторами, форматизаторами, автокорректорами. Многие из упомянутых языков программирования (Бейсик, Паскаль, Си) имеют редакторы, пригодные для ввода не только программ на этих языках, но и текстов на естественном языке. Программная среда, часто надстраиваемая над операционной системой и облегчающая использование ее средств (в персональных компьютерах ИБМ это обычно "Нортон командер"), тоже, как правило, имеет текстовый редактор. Пользователи ДОС-совместимых машин предпочитали популярные полтора-два десятилетия назад пакеты, среди которых заслуживают упоминания: из американских – "Word5", "WordStar", "WordPerfect", "ChiWriter", а из отечественных – "Лексикон" (Москва, ВЦ РАН, 1985 г.).

Каждая из этих программ имели свои особенности и сферы применения. "Word" фирмы "Microsoft" располагал таким богатством шрифтов и средств верстки, что при наличии лазерного принтера мог обеспечить настольную издательскую систему возможностями, превышающими полиграфические. "ChiWriter" был незаменим при наборе математических текстов, так как позволял легко писать формулы, под- и надстрочные индексы.

Для русскоязычных текстов получил широкое распространение "Лексикон" (автор Е. Н. Веселов) – многооконный текстовый процессор с русским знакогенератором (которым можно было пользоваться на компьютерах и принтерах западного производства без их пере-

делки). Его версия "Лексикон-Ортодок" (авторы Е. Н. Веселев и А. Б. Борковский) осуществлял перенос русских слов, располагал словарем русского языка на 100 тыс. слов (на базе известного словаря академика А. А. Зализняка). Другой русский спеллчеккер "Орфо" при помощи словаря в 120 тыс. слов не только проверял правописание, но и выявлял ошибки согласований в предложениях, находил опечатки в знаках препинания, легко "обучался" новым словам.

Все это теперь история. Головокружительный прогресс наблюдается не только в развитии вычислительной техники и ее периферии, но и в программном обеспечении. В России и во всей Европе практическим стандартом и системных и прикладных программ стала продукция фирмы "Microsoft", которая обновляется ежегодно. Это приводит к тому, что в обращении одновременно находятся сразу многие версии программ. Для системных программ – это Windows95, 98, Me, 2000, XP. Из прикладных программ большинство пользователей остановилось на Office97, хотя Office2000 и XP предоставляют значительно большие возможности, оставаясь совместимыми с программой 1997 г.

Статьи, монографии, деловые документы редко состоят из одного только текста. В них содержатся расчеты, таблицы, графики, диаграммы, указатели. Для выполнения всех этих элементов деловых бумаг недостаточно описанных возможностей текстового редактора, как бы ни были они велики. Поэтому к системе подготовки текста примыкают еще три программы: "электронная таблица", "диалоговая графика" и "база данных", которые вместе с текстовым редактором как бы образуют упряжку четырех "рабочих лошадей" для человека, использующего компьютер при подготовке разного рода документов и публикаций. В фирме "Microsoft" такой пакет и составляет Office.

Первая электронная таблица "Визикальк" была разработана Д. Бриклиным и Р.Фрэнкстоном в 1978 г. При составлении отчета, например, или сметы часто приходится пересчитывать длинные колонки взаимозависимых чисел, что делает эту работу весьма утомительной. Электронная таблица представляет собой прямоугольную матрицу, состоящую из ячеек, в каждой из которых может быть помещено значение, зависящее от значений других ячеек (например, их сумма или разность). При изменении значения одной ячейки, в остальных значения пересчитываются автоматически в зависимости от ранее заданных формул. Таким образом, любая бухгалтерская, кадровая ведомость или инженерная таблица может всегда находиться в готовом виде, незави-

симо от частоты введения в нее данных.

Таблицы всегда мало иллюстративны. Чтобы проследить динамику тех или иных изменений, можно изобразить их наглядно в виде графика или гистограммы, а статическое распределение лучше всего видно на диаграмме. Диалоговая графика позволяет перевести данные электронной таблицы в одну из этих изобразительных форм, причем они автоматически изменяются при изменении данных таблицы. Первый такой промышленный пакет программ был разработан для персональных компьютеров ИБМ М. Кэпором в 1982 г. под названием "Лотус 1-2-3". Эти программы позволяют также выполнять на компьютере любые рисунки: блок-схемы, чертежи, произвольные иллюстрации с заштриховкой или без нее, рамки, виньетки и другие украшения.

Наконец, еще одной широко распространенной программой, завершающей этот ряд, является "база данных". Она принадлежит к числу так называемых систем управления базами данных (СУБД) и позволяет создавать в машине структурированное хранилище большой емкости для информации, которую можно разыскивать по разным элементам. Создатели этого программного средства американцы У. Рэтлифф, Дж. Тейт и Х. Лашли назвали его "диБейсII, хотя никакой "первой" версии не существовало. Это был просто "ход" в рекламной кампании, развернутой в 1981 г. и увенчавшейся миллионным успехом. Фирма "Тейт" распространяет новые версии "диБейс", а программы данного типа повсеместно используются в персональных компьютерах для создания разного рода справочных и фактографических систем.

Рассказ об этой "большой четверке" самых употребительных прикладных программ нельзя закончить, не упомянув о том, что теперь они, как правило, объединяются в одну интегрированную систему. На Западе известны среди таких систем "ФреймУорк", "НолиджИн", "ЛотусДжаз", "Парадокс", у нас подобную интегрированную систему "Мастер" создал Е. Н. Веселов в 1989 г. для компьютеров с дисковой операционной системой (ДОС). Она объединяла текстовый редактор "Лексикон" с тремя другими описанными компонентами и распространялась советско-американским предприятием "Диалог" при ВЦ АН СССР. Она имела версию, работавшую под Windows, но не могла конкурировать с пакетом "Microsoft". В настоящий момент наиболее распространенным интегрированным пакетом прикладных программ, как уже сказано, является «Офис» фирмы «Майкрософт». Русифицированные версии «Office», включают текстовый редактор «Word», электронную таблицу «Excel», базу данных «Access» и редактор для презента-

ции слайдов “PowerPoint”.

Быстрый прогресс программного обеспечения подготовки текстов объясняется рядом причин. Некоторые из них мы обсуждали в прошлой лекции. Делопроизводством, составлением, редактированием и изданием текстов в мире заняты сотни миллионов людей. В США 80-х годов больше половины всего занятого населения было связано с обработкой текстов. Поэтому повышение производительности труда в этой сфере, которую справедливо называют информационной, имеет большое значение. И любая мелочь, не предусмотренная стандартами, оборачивается огромными потерями рабочего времени.

Приведу примеры. В пишущей машинке при переводе рычага совмещены две операции: возврат каретки и перевод строки. В компьютере этим операциям соответствуют два разных кода. В программах эта избыточность устранялась по-разному: либо оставлялся для обеих операций код одной из них, либо вводился для них совмещенный. Таким образом, существовали четыре варианта кодирования этой простейшей и повторяющейся на каждой строке команды, в результате чего тексты в машиночитаемой форме нельзя было прочитать в разных устройствах.

Эта неоднозначность была устранена с введением Американского стандарта кодирования символов (ASCII), который стал международным. Но для русских букв коды в нем не были предусмотрены, а при их добавлении возникли варианты. Если учесть, что для других языков, пользующихся русским письмом (болгарского, сербскохорватского), были созданы новые варианты, то проблема кодирования кириллических букв никак не могла считаться решенной.

Если добавить к этому, что существовали два отечественных стандарта для расположения латинских букв на клавиатуре (старый и новый), по которым до недавнего времени изготавливались компьютеры разных типов, то актуальность стандартизации становится очевидной. В операционной системе “Windows” был использован новый американский стандарт (ANSII), в котором буквы русского письма получили постоянное место (под не вполне верным названием «кириллица»).

Но и на этом мытарства русского письма (как его надо правиль-

но называть¹⁾ не закончились. При переходе от Office97 к Office2000 была изменена разрядность основных шрифтов (Times New Roman, Arial, Courier New, Times), и русские буквы снова сменили свои кодовые номера, что создает затруднения при переходе от более поздних версий Office к более ранним.

В заключение хотел бы подчеркнуть еще одно обстоятельство. Как мы выяснили, компьютеры берут на себя выполнение все более сложных процессов подготовки оригинал-макета для печати: перенос слов по правилам, размещение на странице заголовков, сносок, формул, таблиц, рисунков, верстку полос и т. п. Все это теперь не требует специального программирования, а создается непосредственно на экране при помощи клавиатуры, устройств "мышь" или "световой карандаш".

На языке программистов такой "дружественный" интерфейс между машиной и пользователем называется "ВИЗИВИГ" (аббревиатура английского выражения "Что вы видите, то вы получите" – What you see is what you get – WYSIWYG). Перефразируя это по-нашему, можно сказать, что электронные средства подготовки текстов позволяют "резать и клеить" тексты и рисовать на экране, одновременно отражая это в оперативной памяти компьютера.

"Понимание" текста на естественном языке

Многие процессы информационной деятельности: поиск информации, ее отбор, аналитико-синтетическая переработка, распространение – все это процессы, связанные с чтением, пониманием (извлечением смысла) и формулированием текста на естественном языке. Вот почему автоматизация этих процессов занимает важное место при разработке новых информационных технологий. Впервые информационные работники вплотную столкнулись с этой проблемой, когда в начале 50-х годов начались интенсивные эксперименты по машинному переводу. По этому поводу существуют разные мнения, о чем говорилось в лекции об информационных системах.

Моя позиция заключается в том, что адекватный перевод текстов с одного естественного языка на другой – задача, не имеющая однозначного решения. Всегда можно получить несколько разных пере-

¹ Буквы славянского алфавита скорее всего изобрел не Кирилл, а Мефодий, а начертание большинства букв, которыми мы пользуемся отличны от славянских и были выбраны собственноручно Петром I.

водов одного и того же текста, в отношении которых допустимо говорить, что они достаточно близки к оригиналу и стилистически корректны, причем степень того и другого не поддается измерению. Подтверждение этой мысли можно найти, если рассматривать эти переводы в диахронии, т. е. на протяжении некоторого времени. Оригинал художественного произведения всегда остается неизменным, а перевод быстро устаревает и нуждается в обновлении. А раз так, то и формализовать эту задачу для ее машинного решения можно лишь в зависимости от формализованности оригинального текста.

Обсуждая возможность адекватного перевода, полезно представить мысленно некоторую шкалу, на которой расположены разные типы текстов различной степени переводимости. На левом краю шкалы находятся поэтические тексты, в отношении которых термин "перевод" применяется условно, поскольку здесь речь идет о переложении поэтических образов, т. е. о сочинении новой поэзии. Продвигаясь по шкале вправо, мы последовательно встретимся с художественной прозой, научными и деловыми бумагами, личной и ведомственной перепиской. Наконец, на правом краю шкалы мы найдем некоторые типы текстов, однозначно передающихся из одного естественного языка в другой. Это различного рода юридические формулы (включая патентные), номенклатурные перечни, транскрибируемые или транслитерируемые названия и имена. Очевидно, что возможность автоматизации перевода и вообще переработки текста будет возрастать по этой шкале слева направо.

Нас, в данном случае, интересуют те типы текстов, которые занимают довольно большое пространство в центре шкалы и которые по меткому выражению покойного академика А. П. Ершова называют "деловой прозой". Он считал, что деловая проза отражает производственные отношения людей и является таким фрагментом естественного языка, который может быть "воспринят" компьютером.

Это убеждение он основывал на том, что данные отношения людей более строго регламентированы, чем другие, что деловая проза используется в модельных ситуациях, которые ведут к ее формализации. "Стихийно реализуемая, – писал он, – но властно диктуемая сутью дела потребность обеспечить быстрое и точное взаимопонимание на градила деловую прозу жесткими средствами выражения, экономичностью и другими полезными для человека и машины свойствами". Ясно, что большая часть публицистики, научных и административных документов написана деловой прозой.

В последние десятилетия успехи лингвистики и логики во многом продвинули наше понимание сложностей машинного перевода, а достижения электронной техники сделали возможными практические системы, которые работают в промышленном режиме (обычно с предварительной подготовкой оригинального текста и последующим редактированием машинного перевода человеком).

Но все же камнем преткновения автоматизированной обработки текста, которая лежит в основе диалога человека с компьютером на естественном языке, является необходимость для понимания этого текста владеть определенными знаниями экстралингвистической (т. е. не содержащейся в тексте) информацией и логическим мышлением (т. е. способностью к логическому выводу и правдоподобным рассуждениям).

Мы уже говорили, что пятое поколение вычислительных машин, с внедрением которых связывали новые революционные изменения в информационной технологии, авторы проекта представляли как компьютеры, ведущие диалог на естественном языке. Многие лингвисты сомневаются в правомерности такой формулировки. Не входя слишком глубоко в существо проблемы, попытаемся вникнуть в представления специалистов о тех видах лингвистического анализа, которые, собственно, и являются машинным "пониманием" естественного языка. Одна из первых трудностей заключается в неоднозначности многих его выражений, даже когда речь идет о языке деловой прозы.

Выделяют пять типов такой неоднозначности: лексическую, структурную, "глубинную", семантическую и прагматическую.

Лексическая неоднозначность возникает из-за полисемии большого числа слов, включая специальные термины. Нам удастся устранить эту неоднозначность на уровне человеческого интеллекта, так как, зная контекст, всегда понимаешь, идет ли речь о *ключе* гаечном, от двери или том, который бьет из-под земли. При информационном поиске нам помогает в этом тезаурус, где слова с разными значениями маркируются. Однако для различения этих значений в машине часто приходится прибегать к трудоемким (и не всегда дающим правильный результат) статистическим процедурам.

Структурная неоднозначность – это, прежде всего, возможность разного синтаксического членения предложения. Например, фразу "*Наблюдения над языком маленьких детей*" можно понять двояко: кто-то наблюдает за языком детей или дети ведут наблюдения над языком, в зависимости от того, относится ли слово "детей" к слову

"язык" или к слову "наблюдения".

Неоднозначность на уровне *глубинной структуры* содержится во фразе "*Этот текст улучшить нельзя*": либо потому, что он совершенен, либо потому, что безнадежно плох (примеры Ю. Д. Апресяна). Сравнивая две фразы "*Каша готова к обеду*" и "*Цыплята готовы к обеду*", мы усматриваем двусмысленность второй из них, поскольку знаем, что цыплят можно и кормить и есть.

Семантическая неоднозначность в речи часто возникает из-за незнания ситуации. Вам говорят: "*Купите автомобиль*", и это может означать, что вы выбираете между автомобилем и мотоциклом, а может быть просто не знаете, куда потратить деньги. "*Человек упал, разбил окно, повредил себе руку*" – все эти действия могли быть и нечаянными и намеренными. В зависимости от этого перевод фразы на другой язык и представление ее в машине будут разными.

Прагматическая неоднозначность иллюстрируется фразой "*Он уронил карандаш на стол и сломал его*". Для человеческого опыта в ней нет неясности, поскольку мы понимаем, что сломался карандаш, но машине это не очевидно. При переводе на другие языки "карандаш" и "стол" могут оказаться словами разного грамматического рода (например, в испанском и французском первое мужского рода, а второе женского), и это окажется существенным.

Для иллюстрации лингвистического и логического анализа, необходимого при машинном «понимании» языка, воспользуемся моделью американского филолога Т. Винограда (Стэнфордский университет), на идеях которого основывается наше изложение. Реализация этой модели (рис. 14) требует сложных компьютерных программ баз данных и баз знаний, содержащих различные словари и правила. Предположим, что в машину введена фраза: *Технологии будут развиваться по законам, которые мы поняли очень давно*. Если она введена с голоса, то для нее должен быть выполнен фонетический анализ, если же вводится письменный текст (с клавиатуры или сканера), то программа начинается с морфологического анализа. Целью первого является распознавание и идентификация фонем, целью второго – установление основных форм слов и их флексий. На третьем этапе проводится лексический анализ, в результате которого образуется последовательность слов, соотнесенных с частями речи и их морфологическими характеристиками (число, падеж и т. п.). На четвертом этапе осуществляется синтаксический анализ фразы – грамматический разбор предложения, – который дает синтаксическую ее структуру (на рисунке она по-

казана в виде дерева). Однако эта поверхностная структура не всегда однозначна, как мы уже убедились на примерах. Поэтому требуется еще анализ глубинной структуры (на рисунке не показанный).

Дальнейшие этапы машинного понимания текста переводят его синтаксическую структуру в логическую, которая позволяет применить процедуры логического вывода и рассуждений. Существуют различные формы семантических анализаторов для кодирования смысла языковых выражений. В данной модели используется исчисление предикатов.

После семантического анализа логическая структура предложения записывается цепочкой логических символов, которые могут быть прочитаны следующим образом:

Существуют такие x, y, z, t_0, t_1, t_2 ,

что x есть *технология*,

y есть *закон*,

z есть *произносящий фразу, который понял у в момент t_2* ,

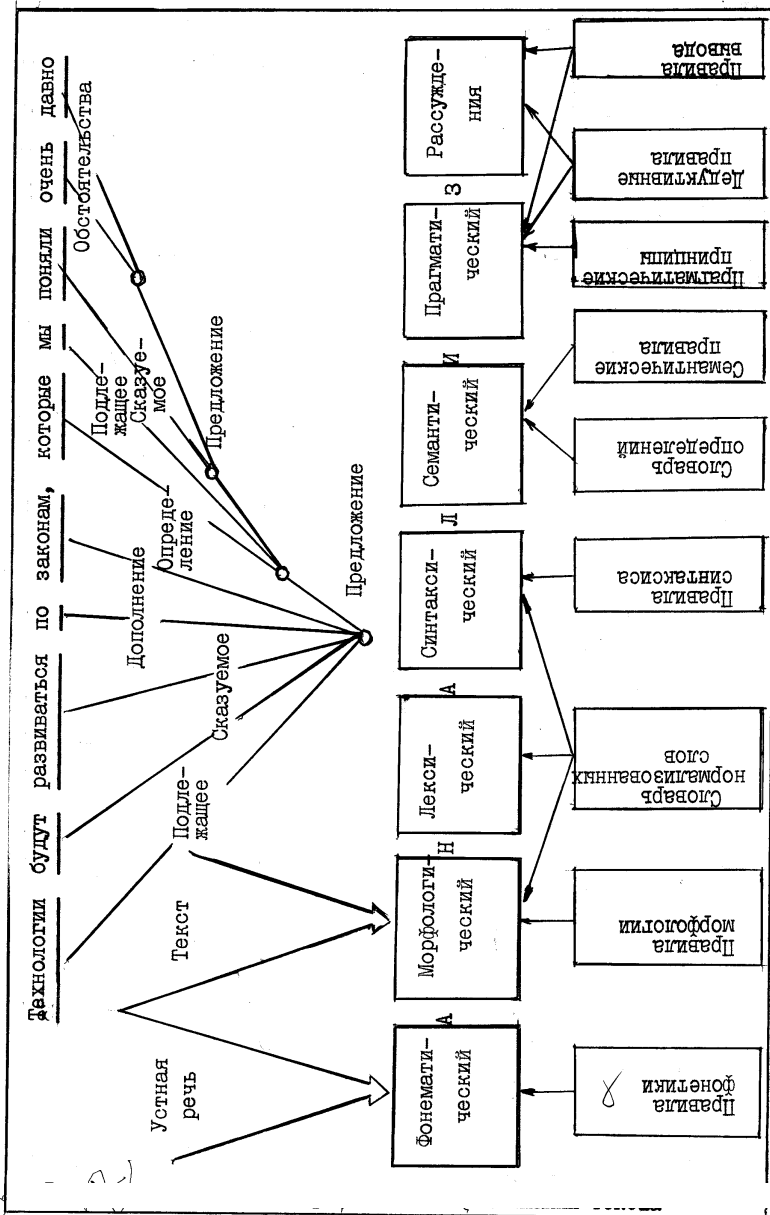
t_0 есть *момент произнесения*,

t_1 наступит после *момента произнесения t_0* ,

x развивается по y в *момент t_1* ,

t_2 был задолго до t_0 .

В ходе прагматического анализа определяется, в частности, что именно известно о переменных. Например, x – связанная квантором переменная. Она утверждает существование чего-то, но не указывает на определенный объект. Другими словами, *технологии* в данном случае это технологии вообще, а не какие-либо конкретные технологии. Точно так же переменная y есть неопределенный объект, задаваемый контекстом. Переменная z тоже остается не полностью определенной, поскольку местоимение *мы* может означать авторов высказывания, авторов и читателей, профессионалов данной области, вообще людей данного поколения. Целью машинного понимания языка является возможность диалога с машиной, в ходе которого компьютер мог бы давать логически осмысленные ответы на вопросы пользователя или же преобразовывать команды в определенные действия, учитывающие реальность. Эту задачу решает последний этап анализа, обозначенный на рис. 14 как «Рассуждения».



Каким образом, например, машина, воспринявшая нашу фразу, будет отвечать на вопрос: «Понимаем ли мы *законы*, по которым будет развиваться *реферирование*?» Для того, чтобы ответить на этот вопрос, компьютер должен знать, что *реферирование* есть процесс, относящийся к информационной *технологии*. Такое знание можно изобразить формулой исчисления предикатов: «Все, что есть *реферирование*, есть *технология*».

Есть и другие достаточно эффективные способы введения знаний в машину, такие как семантические сети, фреймы. Можно, например, ввести в машину семантическую сеть, в которой все виды технологий, включая и информационную, и все их разновидности будут связаны определенными отношениями (род–вид, часть–целое и т. п.). По такой сети можно автоматически определить, что *реферирование* есть часть, или вернее, один из процессов информационной *технологии*. Таким образом, в данной ситуации компьютер сможет дать правильный ответ на заданный ему вопрос.

Однако трудность реального представления знаний в машине заключается в многообразии конкретных ситуаций, от которых зависит понимание человеком текстов на естественном языке. В нашем примере из контекста нельзя понять, что означает выражение «*очень давно*», хотя информационные работники знают, что понимание некоторых законов информационной технологии пришло благодаря интенсивным исследованиям научных коммуникаций в середине 60-х гг. XX в.

Подобным же образом и выражение «*будут развиваться*» означает не столько будущее время, сколько продолженное действие. Можно привести много других примеров, когда фраза на естественном языке, вполне понятная человеку в конкретной ситуации, требует специальных приемов интерпретации для ее машинного понимания.

Во многих научных коллективах разрабатываются методы перевода с естественного языка на язык математической логики. Они необходимы для глубокого семантического анализа во многих автоматизированных информационных системах.

В проведении исследований важное место занимает анализатор, осуществляющий перевод синтаксического «дерева» в формулы информационно-логического языка. На каждом шаге его работы исходная синтаксическая структура приближается к логической формуле при помощи трансформаций-разверток до тех пор, пока формула не будет выражать смысл фразы. При этом, если исходная фраза неоднозначна,

система в режиме диалога предлагает пользователю уточнить, какой из найденных машиной вариантов понимания он имел в виду.

Если подытожить сказанное, то суть проблемы заключается в том, что никакая, даже самая совершенная машина не может «понимать» текст на естественном языке так, как его понимает человек. Но она может однозначно воспринимать формулы математической логики. Поэтому задача формализации текста состоит в том, чтобы научиться устранять неопределенность и многозначность текстов на естественном языке при их переводе на формальный язык логики.

Разумеется, это один из многих путей, которым исследователи пытаются обучить компьютер пониманию естественного языка.

Электронная книга

Новая концепция книги

Новые условия для одного из основных явлений человеческой культуры – книги – заключаются в появлении ее необычной физической формы – электронной. Во многих научных дисциплинах, связанных с созданием, распространением и использованием книги, разработаны собственные концепции книги. Это книго-, библиотеко-, библиографо-, архиво-, документоведение, журналистика, полиграфия, информатика и другие. Концепция электронной книги в каждой из этих дисциплин пока еще разрабатывается. Не исключено, что она будет более или менее общей для них.

В самом широком (общем) смысле книгой часто называется физическая форма законченного и единого произведения печати или письменности. Под произведением при этом понимают результат целенаправленной познавательной деятельности, имеющий определенную логическую взаимосвязь частей, завершенность в целом, и изложенный в письменном виде. Самая узкая концепция книги бытует в статистике печати, где книгой считается непериодическое многостраничное произведение печати объемом не менее 49 страниц, не считая обложки и титульного листа. Понятно, что при разработке концепции электронной книги руководствуются самым широким и общим ее пониманием.

С точки зрения физической формы любой компьютерный файл является программой, записанной в определенном формате. В двоичных кодах в памяти компьютера записаны команды управления самой машиной, обработкой данных, а также монографии, статьи, изображения, звуки, видео- и кинофильмы. Они могут быть записаны на магнитной ленте, без затруднений переписываться с одного носителя на другой. Их передают по различного рода сетям и каналам связи, в том числе и телефонным. Любой пользователь может создать любой файл

и поместить его на свою страницу в Интернете для всеобщего обозрения.

Важное для книги понятие тиража в данном случае теряет смысл. По всей вероятности, искать опору в выявлении этого производного для книги понятия следует в наличии *регистрации* данной программы, пакета программ, баз данных или, другими словами, произведений в электронной форме. Помимо этого, во всех странах издающие и предоставляющие услуги по копированию учреждения лицензируются, и указание на *лицензию* также может служить признаком издания, отличающим его от произведения, выпущенного частным образом и прежде считавшегося рукописью.

Другой не праздный вопрос заключается в том, целесообразно ли называть *электронной книгой* некоторые виды изданий или произведений в электронной форме. На самом деле они давно уже так называются (ведь почти вся терминология метафорична), но многие книговеды возражают против этого. По их мнению, книга должна быть портативной и не требовать специальных устройств для использования, а эти условия в электронной книге не выполняются. Однако вся история книги свидетельствует о том, что ее развитие сопровождалась сменой носителей информации и способов производства книги. Клинописные глиняные таблетки (плитки) сменялись папирусными свитками, за ними последовали пергаментные и бумажные кодексы.

Последняя смена происходила в связи с изобретением и распространением книгопечатания. Промежутки времени между этими сменами сокращались от тысячелетий до столетий. Ускорение темпов развития информационной технологии привело к тому, что уже на памяти одного поколения мы обсуждали вопрос о том, являются ли книгой или вернее *микронкигой* микрофильмы и микрокарты (микрофиши). Теперь настала очередь электронной книги. Смены материальных форм книги вызывались общественной потребностью в упрощении доступа к информации, появлением новых технологий ее производства и всегда приводили к ее удешевлению, новым возможностям ее использования и выполнению ею новых функций. Все это происходит и при внедрении электронной книги. Достаточно упомянуть среди обстоятельств ее появления экологическую опасность дальнейшего увеличения числа бумажных книг – вырубку лесов и изменение состава атмосферы.

По нашему мнению, электронную книгу целесообразно рассматривать именно в книговедческом аспекте, поскольку книгоиздание

во всех развитых странах уже перешло на электронный набор. При этом, к сожалению, многовековой опыт книжной культуры, сложившийся под влиянием психофизиологических особенностей восприятия текста человеком, далеко не всегда используется. Чтобы не потерять эту культуру, нужно внедрять лучшее из накопленного опыта в новую информационную технологию. А кто же сделает это лучше, чем редакторы и издатели? С другой стороны, возможности компьютера видоизменяют методы общения человека с книгой, ведут к созданию нового типа книги, которая, по-видимому, станет кумулятивной с функциональной точки зрения и заменит некоторые виды печатной книги, например, словари, справочники, учебники, библиографические указатели, реферативные журналы и т. п. Очевидно, что электронное издание позволяет обновлять и дополнять их данные без повторного набора всего текста.

Сущность, особенности и разновидности электронной книги

Электронная книга определяется как совокупность данных (текст, звук, статичное и движущееся изображение) в памяти компьютера, предназначенная для восприятия человеком с помощью соответствующих программных и аппаратных средств. В широком смысле эта разновидность книги обладает многими возможностями мультимедиа сочетать текст с аудио- и видеоматериалами, обладающими стереозвуковыми и стереоскопическими эффектами, что придает ей черты динамичной и интерактивной книги¹.

Некоторые ее виды можно не только читать, но и изменять, добавляя собственные наблюдения, размышления и результаты экспериментов (если все файлы предварительно не скомпилированы в один .exe-файл). Электронная книга получила широкое распространение после внедрения в быт в качестве внешней памяти компьютера компактных оптических дисков, на которых вначале умещалось до 150 тыс. страниц печатного текста, а теперь (по технологии DVD) – в 7 раз больше, т. е. около 1 млн.

В настоящее время сотни журналов и тысячи книг ежегодно выпускаются в электронной форме. Возможность манипулировать текстами этих изданий вызвала к жизни гипертекст, т. е. такую форму их организации, при которой смысловые единицы (фразы, абзацы, разде-

¹ Гиляревский Р. С. Электронная книга // Книга: Энциклопедия. – М.: БРЭ, 1999. – С. 729–730.

лы) представлены не в линейной последовательности, а как система явно указанных возможных переходов или связей между ними. Это создает возможности поиска и объединения по смыслу фрагментов текстов из многих источников (т. е. навигации по гипертексту), что предполагает новый тип восприятия документов (в первую очередь, специальной литературы). При этом типе восприятия чтение перестает быть линейным (от первой до последней страницы), становится выборочным, "поисковым" и может осуществляться одновременно во многих направлениях и произведениях¹.

Такая модель иногда называется "гиперкнигой". Сохраняя многие черты печатных книг с учетом функций интерфейса и других имеющихся средств (ориентация, навигация, персонализация), гиперкнига обеспечивает дополнительные преимущества в отношении поиска по логическим связям. Это динамическая система, включающая структурные и функциональные компоненты:

- Структурные, которые отражают деление книги на страницы, страниц – на текст, иллюстрации и т. п. элементы;
- функциональные – динамический и интерактивный аспекты системы. В частности, некоторые программы позволяют читателю менять "состояние системы", т. е. из многих текстов компоновать новый текст, изменять положение текста и иллюстраций на странице, менять шрифты. При этом в некоторых случаях существует (хотя официально и не признана) техническая возможность сохранять содержание библиотечной книги в собственном компьютере.

Электронный журнал: проблемы распространения и хранения

Самым распространенным видом электронных изданий является электронный журнал, который набирает силу в результате необычайно быстрого роста числа пользователей сети Интернет и особенно WWW. Теперь и крупные издательства, которые долгое время использовали Интернет для распространения экспериментальных электронных изданий, в полной мере осознали, что эта сеть стала важным средством распространения их публикаций. И хотя технология доступа к элек-

¹ Гиляревский Р. С., Субботин М. М. О возможностях оценки перспективности новых информационных технологий (на примере гипертекстовой технологии) // Научно-техн. информация. Сер.2.- 1988.- № 4; Черный А. И. Современные проблемы интеграции издательского дела, полиграфии и информационного обслуживания // Итоги науки и техники. Сер.: Издательское дело и полиграфия, т.1. – М., 1989.

тронным сериальным изданиям используется широко, осталось много нерешенных проблем. В числе таких проблем следует назвать разработку ясных и понятных механизмов установления платы за использование электронных журналов. Не решены также вопросы авторского права, доступа к материалам за предшествующие годы, гарантированной доступности журналов. А без этого невозможно определить многие положения библиотечной стратегии и тактики.

Наиболее интересными проектами подготовки электронных журналов являются проект Muse Издательства Университета Джонса Хопкинса (Балтимор, шт. Мэриленд, США), проект JSTOR (Journal Storage Project = Проект хранения журналов) ряда крупных издательств Великобритании, проект интерактивного доступа к электронным журналам Компьютерного библиотечного центра с интерактивным доступом (OCLC = Online Computer Library Center), программы издательств Academic Press и Elsevier Science Publishing.

Интернет революционным образом изменил возможности неформальной научной коммуникации. Электронная почта, доски объявлений, интернетовские «конференции» – все это действительно уничтожает расстояния и делает «невидимые коллегии» интенсивно взаимодействующими коллективами.

В связи с электронными книгой и журналом возникают и новые «жанры» научной коммуникации, пока еще не очень хорошо осознанные и освоены. Примером может служить жанр откликов на статью. Во многих электронных версиях журналов, доступных через Интернет, каждый пользователь может составить свой комментарий к любой статье и познакомиться с комментариями других читателей. В традиционной системе научной коммуникации институт опубликованных рецензий распространялся только на книги, чаще всего организовывался редакциями журналов, имел большой временной лаг и был крайне ограничен по охвату рецензируемых книг. Что касается статей, то лишь считанные их единицы могли получить отклик в "Письмах в редакцию", да и то чаще всего по принципиальному несогласию с автором опубликованной статьи. Новая демократическая возможность в Интернете позволяет каждому выразить свое отношение к любой статье и в любой момент выяснить степень интереса к ней.

Другой пример – возможность по ссылкам на статьи (включая и гипертекстовые) сразу обращаться к полным их текстам, если они доступны через Интернет. Это создает совершенно новое пространство как бы единой научной книги или журнала, служит воплощением дерз-

ких идей Поля Отле о всемирной энциклопедии и Герберта Уэлса о мировом мозге.

Электронный журнал вышел из пеленок и начал взрослую жизнь в результате необычайно быстрого роста числа пользователей сети Интернет и особенно www-технологии. Теперь и крупные издательства, которые долгое время использовали Интернет для распространения экспериментальных электронных изданий, в полной мере осознали, что эта сеть стала важным средством распространения их публикаций среди мирового научного сообщества. И хотя почти стандартной технологией доступа к электронным сериальным изданиям стала Web-технология, в младенческом состоянии осталось много важных проблем, связанных с такими изданиями. В числе таких проблем следует назвать разработку ясных и понятных механизмов установления платы за использование электронных журналов, электронного макетирования страниц. Не решены также вопросы авторского права, доступа к материалам за предшествующие годы, гарантированной доступности журналов, отражения данных и другие.

Ценообразование. Установление обоснованных и приемлемых для потребителя цен на электронные журналы является трудной проблемой. Издатели ищут пути, как защитить свои доходы от сокращения подписки на печатные журналы и потому придумывают хитроумные схемы оплаты. Для библиотек и других потребителей оказывается неприемлемым то, что цены на электронные журналы представляются неоправданно завышенными, или то, что они неразрывно связываются с печатными версиями, которые потребитель не всегда желает получать.

Безопасность. Одна из ключевых проблем, с которой столкнулись издатели электронных журналов, заключается в следующем. Как организовать полный доступ к издаваемым ими электронным журналам и сохранить надлежащую правовую защиту для подписчиков и заблокировать доступ к ним для всех остальных. Многие издатели разрешают любому пользователю просматривать определенные фрагменты издаваемых ими электронных журналов (например, их образцы, некоторые статьи или оглавления журналов). Однако для доступа к полной публикации обычно требуется подтверждение права на него.

Наиболее часто используются следующие методы:

- а) предоставление подписчику пароля;
- б) использование специальной зоны доступа, только из которой пользователь данной организации может просматривать материалы

электронного журнала по сети Интернет. Недостатком первого метода является то, что возникают затруднения, когда подписчиком является не индивидуальный пользователь, а организация. А недостатком второго метода является ограничение места доступа к ресурсам сети Интернет. Это означает, что издателям еще предстоит разработать методы обеспечения безопасности, не имеющие указанных недостатков.

Электронное представление страниц. Хорошо форматированная печатная публикация не всегда выглядит привлекательно при ее электронном представлении. Кроме того, интерактивная технология позволяет издателям делать такие вещи, которые были невозможны в печатных изданиях. Например, эта технология допускает установление интерактивных гиперсвязей между взаимосвязанными ресурсами, установление связей со всеми видами мультимедиа, т. е. звуком, видео, графикой в двух- и трехмерном представлении. Возможна связь первичных публикаций с традиционными службами реферирования и индексирования, страничных ссылок – с соответствующими публикациями.

Однако ни одно из возможных решений не подходит для любого журнала. Если в журнале приводятся сложные математические уравнения и используются специальные символы и изображения, то такие журналы иногда сканируются. При этом используются такие форматы, как Portable Data Format (PDF) фирмы Adobe Acrobat, *.gif, *.jpeg, и результаты сканирования выдаются на экране в таком же виде, какой имеют печатные страницы. В отличие от издательств, которые производят электронные журналы путем сканирования их печатных версий, некоторые издательства (например, Elsevier и Springer-Verlag) используют для подготовки электронных и печатных журналов интегрированную технологию. Такие издательства применяют формат SGLM.

Главная трудность, с которой в настоящее время сталкиваются издатели при использовании сканированных изображений журнальных страниц, заключается в том, что такие изображения не могут использоваться как символьное представление текста. В них нельзя включить гиперссылки и использовать средства поиска и обработки текстов. Многие электронные журналы представлены непосредственно в формате Hyper Text Marked Language (HTML), который позволяет передавать символьный текст, гиперссылки, мультимедиа и другие современные формы представления информации, используемые в Интернете. В некоторых проектах последовательно используются оба метода, чтобы

обеспечить поиск в HTML-тексте и представлять изображение в его исходной форме.

Доступность материалов за предшествующие годы. Библиотеки традиционно играли роль архивов, приобретая и сохраняя журналы для их вероятного использования читателями в течение длительного времени. А кто будет выполнять эту функцию в отношении журналов в электронной форме? Разумеется, издатели могли бы выполнять эту функцию в отношении издаваемых ими журналов. Однако это потребовало бы от них затраты значительных средств, объем которых со временем мог бы превысить стоимость самих издательских процессов. Если же издательство прекратит свое существование или сольется с другими издательствами, что станет с выпускавшимися им электронными изданиями и останутся ли они доступными для использования?

Большинство издательств стремится лишь распространять пользующиеся спросом материалы. Едва ли они будут склонны тратить немалые деньги на то, чтобы поддерживать большие, но редко используемые массивы данных за прошлые годы в форме, допускающей интерактивный доступ к ним. Возможны разные решения этой проблемы, но среди них не просматриваются достаточно эффективные модели. Например, издательства могли бы размещать копии своих электронных изданий по договорам в нескольких местах, чтобы обеспечить их постоянную доступность при возникновении необходимости.

Центр OCLC Electronic Collections Online (США) взял на себя обязательство постоянно архивировать получаемые им электронные журналы. И если в будущем по каким-либо причинам он не сможет выполнить эти обязательства, то по согласованию с сотрудничающими издательствами и по соглашениям с библиотеками будет обеспечена возможность доступа к архивным копиям соответствующих электронных журналов, хранящихся в этих библиотеках.

Библиотеки и их объединения также могли бы начать согласованное избирательное архивирование электронных журналов – с согласия издательств. Роль архивов могли бы взять на себя агентства по подписке на журналы, такие как EBSCO и Swets. Такую же роль могли бы выполнять и национальные библиотеки. Возможно также, что электронным архивированием журналов займутся коммерческие фирмы, которые традиционно занимаются микрофильмированием материалов и их записью на компакт-диски, например University Microfilms International (UMI).

Надежность и доступность данных. Из-за недостаточной пропускной способности каналов связи и большого объема трафика в сети Интернет нередко возникают перебои в доступе к различным входящим в нее сетям и серверам. Поэтому для часто используемых электронных журналов может оказаться рациональным размещение их копий в различных серверах, находящихся в разных точках мира. Благодаря этому мог бы быть обеспечен повсеместный и легкий доступ к ним. Например, система KRII Dialog фирмы Knight-Reader Information Inc. имеет архивное хранилище емкостью более 8 терабайт в своем центре, находящемся в г. Маунтин Вью (шт. Калифорния, США), которое может быть связано арендованными линиями с ключевыми точками мира, в которых происходят перегрузки в трафике Интернет. Это должно существенно облегчить доступ к архивным фондам электронных журналов.

Организационные и юридические проблемы

Производство книги. Как уже было сказано, оригинал-макет, да и печатная форма большинства традиционно издаваемых книг готовятся к печати электронным способом, т. е. на компьютере. После печати тиража остается издательский файл оригинал-макета, который пока что не используется как электронная книга, но может считаться таковой. Однако подлинная электронная книга не может быть воспроизведена полиграфически, потому что, имея линейный текст, она не имеет линейной организации. Она с самого начала создается как гипертекстовая (по идеологии гипертекста) и с использованием языковых средств гипертекста. Хорошо форматированная печатная публикация не всегда выглядит привлекательно при ее электронном представлении. Кроме того, интерактивная технология позволяет издателям производить такие операции, которые были невозможны в печатных изданиях. Например, эта технология допускает установление интерактивных гиперсвязей между взаимосвязанными ресурсами, установление связей со всеми видами мультимедиа, т. е. звуком, видео, графикой в двух- и трехмерном представлении. Возможна связь первичных публикаций с традиционными службами реферирования и индексирования, постраничных ссылок – с соответствующими публикациями.

Если говорить о распространении электронных изданий, то здесь возникает много проблем, которые также нельзя решить на основе традиционного опыта.

Авторское право. Особенно большие затруднения возникают, когда на электронные издания распространяются старые законы об авторском праве. Из-за того, что электронные изображения страниц журналов находятся в Интернете, становится очень трудно контролировать их использование. Пока еще нет ясного представления о законах, защищающих интересы издателей в этом отношении.

Впервые создатели и читатели электронных книг и журналов столкнулись с проблемами авторского права совсем недавно, с возникновением World Wide Web – всемирной паутины, позволяющей блуждать по сети людям, не имеющим специальных знаний и определенных навыков, которые требовались для работы в Интернете с момента возникновения сети в 1969 г. Разобраться с документами, представленными в “овеществленной” электронной форме на дискетах и компакт-дисках, гораздо проще, так как их приобретают традиционными, аналогичными покупке печатной продукции способами.

Электронные документы в библиотеках в подавляющем большинстве своем появляются в результате их воспроизведения, т. е. записи в память ЭВМ путем сканирования, оцифровки имеющихся документов, а также получения по высокоскоростным каналам связи через сеть, как с отечественных, так и с иностранных сайтов.

В России функционирование документов, в том числе и электронных, регулируется следующими нормативными актами в области авторского права:

- Закон РФ об авторском праве и смежных правах от 9 июля 1993 г., № 5351-1;
- О государственной политике в области охраны авторского права и смежных прав (Указ Президента РФ от 07.10.93, № 1607);
- Вопросы присоединения РФ к ряду международных конвенций в области охраны авторских прав (Распоряжение Президента РФ от 25.03.94, № 152-рп);
- О присоединении Российской Федерации к Бернской конвенции об охране литературных и художественных произведений от 9 сентября 1886 г., пересмотренной в Париже 24 июля 1971 г. и измененной 2 октября 1979 г., Всемирной конвенции об авторском праве, пересмотренной в Париже 24 июля 1971 г. (Постановление правительства РФ от 3 ноября 1994 г., № 1224);
- Об информации, информатизации и защите информации (№ 24-ФЗ, 20.02.95, Федеральный закон принят Государственной Думой 25 января 1995 г.);

- Об участии в международном информационном обмене (N 85-ФЗ, 04.07.96, Федеральный закон, принятый Государственной Думой 5 июня 1996 г.).

Кроме национального законодательства, на территории Российской Федерации введен в действие ряд международных соглашений. К международным актам, участником которых является Россия и действие которых распространяется на защиту прав авторов, относятся:

- Бернская конвенция об охране литературных и художественных произведений в редакции 1971 г. (Бернский Союз), к которой Россия присоединилась с 1994 г.;

- Всемирная конвенция по охране авторского права, разработанная по инициативе ЮНЕСКО, подписанная в Женеве 6 сентября 1952 г. (СССР с 1971 г.) и пересмотренная в Париже 24 июля 1971 г., которая начала действовать в России с 1995 г.;

- Стокгольмская конвенция от 14 июля 1967 г. об учреждении Всемирной организации интеллектуальной собственности.

Общий принцип Женевской и Бернской конвенций состоит в том, что произведения созданные в странах-участницах конвенции, в каждой из стран пользовались бы той же охраной, что и национальные произведения.

Электронная библиотека реальная и виртуальная

Во второй раз на протяжении жизни нашего поколения библиотеке как социальному институту брошен вызов. В первый раз это было в 50-е – 60-е гг., когда во всех развитых странах государство стало заботиться о создании систем научно-технической информации. На этот первый вызов библиотеки ответили агрессивно. Один из ведущих американских теоретиков библиотечного дела писал: «Библиотекари, на баррикады!» Время нас рассудило – все становится на свои места. Мы пользуемся, как всегда, одними и теми же техническими средствами и в рамках общей цели и информационной технологии решаем разные задачи. Библиотекари осуществляют доступ к опубликованным документам и формируют информационные потребности общества, а информационные работники удовлетворяют их при помощи содержательной переработки первичной информации во вторичную.

На этот раз вызов оказывается посерьезнее. В течение, по меньшей мере, пяти тысячелетий библиотека работала с текстовыми, по преимуществу, документами. Иллюстрации – таблицы, графики,

формулы, рисунки – были вспомогательными средствами хранения и передачи знаний. Теперь появились и быстро получают повсеместное распространение в этом качестве звук и движущиеся изображения, совмещенные с текстом в цифровом машиночитаемом виде. Это чаще всего обобщенно называется электронной книгой, а собрания таких книг – электронной библиотекой, хотя это встречает возражения со стороны ряда книгovedов и библиотечарей. Дело не в названиях, но и в них, потому что за терминами стоят концепции.

Проблема, стоящая перед библиотечным делом и ее наукой – библиотечковедением, – заключается в том, сохранять ли многотысячелетнюю традицию работы с преимущественно текстовыми книгами (в печатной и электронной форме) или взять на себя ответственность за доступ к знаниям во всех мультимедийных видах. Первое решение этой альтернативы проще реализовать и в человеческом, и в экономическом, и в организационном плане. Но только для библиотечного дела. Общество все равно создаст другой социальный институт для обеспечения доступа к знаниям в мультимедийной форме. Для него уже есть название – медиатека. Библиотека перестанет быть единственным учреждением, решающим в обществе эту задачу. Как знать, не потеряет ли она свой уникальный престиж и не утратит ли позиции и в своей традиционной сфере?

Убежден, что мультимедийные средства передачи знаний быстро овладеют сферами науки, образования, просвещения, т. е. теми областями общественной жизни, где доступ к знаниям является решающим фактором. Возможно, вопросу «Читали ли Вы?» будут предшествовать вопросы «Слушали ли? Видели ли?» В сфере развлечений и досуга это уже произошло. Еще в 1970 г. на Всемирном конгрессе историков Д. А. Поспелов убедительно доказывал, что компьютерная графика существенно повлияет на изменение характера интеллектуальных способностей подрастающего поколения. Если до сих пор усвоение знаний предполагало большее развитие левого полушария мозга, то теперь оно будет способствовать более гармоничному развитию.

Разумеется, второе решение этой альтернативы потребует коренной перестройки, прежде всего, профессионального мировоззрения библиотечарей, а затем и всей сферы библиотечного дела. Мультимедийные средства – музыкальные и видеокассеты, компакт-диски, соответствующие сайты в Интернете – потребуют не только приобретения и освоения соответствующей технологии. Это самое простое и оно необходимо при любом выборе, и, собственно говоря, уже происходит.

Гораздо труднее перестроить профессиональное сознание и формирующие ее научные дисциплины: библиотеко-, библиографо- и книговедение. Может быть, не зря их объединили в одну научную специальность.

Убежден, что этот вызов библиотека должна принять позитивно и не откладывая до того времени, когда альтернатива решится сама собой. Вынужденные решения всегда хуже сознательных. Правильное решение предполагает активное участие в создании условий для развития электронной книги и библиотеки, в теоретическом и практическом решении кардинальных проблем этого развития.

Электронное издание – это теперь уже не прогноз, не предмет спора или сомнения. Это реальность: число таких изданий ежегодно измеряется десятками тысяч, и в библиотеках они занимают все более весомое место. Перед всеми участниками коммуникационной сферы, и перед библиотекарями в особенности, возникают в связи с этим вопросы и проблемы, требующие ответов и решений. Вот некоторые из них, представляющиеся наиболее важными.

Что считать электронным изданием? Ведь с точки зрения физической формы любой компьютерный файл является либо программой, либо данными, записанными в определенном формате. В двоичных кодах в памяти компьютера записаны команды управления самой машиной, обработкой данных, монографии, статьи, изображения, звуки, видео- и кинофильмы. Они могут быть записаны на магнитной ленте, жестком, мягком или компакт-диске, без затруднений переписываться с одного носителя на другой, передаваться по различного рода сетям, в том числе и телефонным. Любой пользователь может создать любой файл и поместить его на свою страницу в Интернете для всеобщего обозрения. Привычное понятие тиража в данном случае теряет смысл. По всей вероятности, искать опору в выявлении этого важного для библиотеки понятия следует в наличии *регистрации* данной программы, пакета программ, баз данных или, другими словами, произведений в электронной форме. Помимо этого, во всех странах издающие учреждения лицензируются, и указание на *лицензию* также может служить признаком издания, отличающим его от произведения, выпущенного частным образом, и прежде считавшегося рукописью.

Другой не праздный для *библиотеки* (так как ее название включает слово *книга*) вопрос заключается в том, целесообразно ли называть *электронной книгой* некоторые виды электронных изданий или произведений. На самом деле они давно уже так называются (ведь поч-

ти вся терминология метафорична), но многие книго- и библиотековеды возражают против этого. По их мнению, книга должна быть портативной и не требовать специальных устройств для использования, а эти условия в электронной книге не выполняются. Однако вся история книги свидетельствует о том, что ее развитие сопровождалась сменой носителей информации и способов производства книги. Клинописные глиняные таблетки сменялись папирусными свитками, за ними последовали пергаментные и бумажные кодексы.

Последняя смена происходила в связи с изобретением и распространением книгопечатания. Промежутки времени между этими сменами сокращались от тысячелетий до столетий. Ускорение темпов развития информационной технологии привело к тому, что уже на жизни одного поколения мы обсуждали вопрос о том, являются ли книгой или вернее *микрокнигой* микрофильмы и микрокарты. Теперь настала очередь электронной книги. Смены материальных форм книги вызывались общественной потребностью в упрощении доступа к информации, появлением новых технологий ее производства и всегда приводили к ее удешевлению, новым возможностям ее использования и выполнению ею новых функций. Все это происходит и при внедрении электронной книги. Достаточно упомянуть среди обстоятельств ее появления экологическую опасность дальнейшего увеличения числа бумажных книг – вырубку лесов и изменение состава атмосферы.

Библиотека или медиатека? С какими видами электронных изданий должна работать библиотека, со всеми или с некоторыми? Вопрос непростой, и ответы на него со временем быстро меняются. Уже давно к традиционным книгам, журналам, газетам добавились сначала грампластинки, а затем аудио- и видеокассеты. С появлением в библиотеке электронных изданий этот вопрос переходит из сферы материального (по видам носителей) в сферу идеального (по содержанию), поскольку, как мы выяснили, носители перестали играть существенную роль. Практика опережает теорию – библиотекари уже предлагают читателям фильмы и другие мультимедийные произведения на компакт-дисках и других носителях. А как быть с программами? Не знаю, но думаю, что без некоторых пакетов прикладных программ (например, текстовых редакторов, программ для просмотра мультимедийных файлов) читателям библиотек не обойтись. Так что придется работать и с этими видами электронных изданий. Может быть правы французские библиотекари, которые первыми начали называть свои учреждения медиатеками?

Но и традиционные виды изданий в электронном виде настолько меняют свой характер, что перед библиотекарями возникают непростые проблемы.

Лучше ли электронная книга традиционной печатной?

Доводы за электронную книгу

Наличие поисковых систем. Главным достоинством любой электронной книги считается наличие в ней *поисковой системы*, а значит возможности анализа содержания книг, поиска нужных сведений при помощи компьютера. В этом смысле, электронная книга должна удовлетворять следующим требованиям:

- осуществлять быстрый поиск заданной информации (по выделенному ключевому слову или так называемой "горячей области");
- по требованию пользователя предоставлять ему справочную информацию по определенной теме;
- обеспечивать интерактивный режим работы пользователя;
- реализовывать различные режимы работы с системой ("чтение", "просмотр", "выборочное чтение", "поиск заданного смыслового фрагмента" и т.п.);

Очевидно, что в этой области электронная книга далеко обошла печатную. Найти какую-либо цитату, фразу, провести поиск текста по синонимам, с учетом морфологии языка – все эти возможности недоступны, если вы имеете дело с печатной продукцией. Именно данными особенностями обусловлен тот факт, что первыми в электронном виде были изданы разного рода энциклопедии и словари.

Доступность и оперативность. Еще одним преимуществом электронного издания является то, что получить его на свой рабочий стол можно в считанные минуты, заказав его в Интернет-магазине или "скачав" из электронной библиотеки. Важно, что к такой информации одновременно могут обращаться тысячи людей. Электронные технологии уже заметно ускорили и упростили процесс издания книги. Теперь временное расстояние между написанием рукописи и массовым прочтением книги может сократиться еще значительно и станет занимать недели, дни.

Возможность персональных изданий. Упрощенный процесс обработки рукописи и возможность электронной доставки позволяют авторам самим издавать свои произведения и распространять и даже продавать их. Еще в начале 80-х годов автор идеи гипертекста Теодор

Нельсон предсказал возникновение нового круга проблем в области взаимоотношений между автором, редактором и читателем книги. По его мнению, доступность электронных версий текстов в Интернете и на дешевых цифровых носителях информации сделала возможной "компиляцию на лету"; в результате компиляция, монтаж текстов, становится вполне достойной разновидностью авторского труда.

Согласитесь, что сегодня в обиходе слово "компилятор" применительно к определению человека пишущего звучит пренебрежительно. Но для учителя структурирование и компиляция материалов, полученных из самых различных источников, всегда оставались рабочим инструментом.

Как известно, издатель не станет тратить деньги на печать, хранение и доставку текстов на бумаге, не убедившись в существовании определенного спроса на книгу. В электронном издательстве подобных расходов нет. Поэтому журналы, специализирующиеся на электронных технологиях, все чаще пишут о "благодатных временах, наступающих для тех, кто мечтает стать писателем". "Теперь уж никто не сможет пожаловаться, что он пишет "в стол", скорее, в компьютер. Новые технологии издательского дела сравнивают с большим взрывом, из которого родится новая книжная вселенная. На этот раз электронная.

Легкость аннотирования и рецензирования. Перед изданием версии книг могут стать доступными через Интернет или электронную почту. Это позволяет многим специалистам иметь доступ к содержанию. Рецензирование может осуществляться прямо в сети на сайте автора. Аннотацию же возможно составить автоматически в редакторской программе Word 2000, Word XP или специальной программе, например, «Либретто» фирмы Media Lingua.

Экономия в производстве. Известно, что огромная часть издательских расходов относится к необходимым для этого материалам. При издании газет, например, до 40 % расходов приходится на бумагу и краску для печати. При создании электронных книг, журналов, газет эти траты сводятся к нулю. Кроме того, некоторые специалисты отмечают и низкую стоимость редактирования и подготовки. С помощью имеющихся издательских пакетов и программного обеспечения для распространения электронных документов авторы могут выполнять большую часть операций по редактированию и макетированию. Неслучайно поэтому некоторые российские издательства заинтересовались идеей выпуска книг в цифровой форме.

Экологические преимущества. Один из сильнейших доводов в пользу электронной книги – ее производство не требует бумаги. Это позволит спасти от вырубки множество деревьев. Для создания электронной книги не используются печатные станки, не расходуется дорогая, а иногда и вредная краска. Все еще тяжелый физический труд людей заменяется интеллектуальным.

Сохранность хрупких и ценных изданий. Проблема сохранности документов в библиотеках и архивах сегодня напрямую связана с компьютерными технологиями. Многие хранители информации на бумажных носителях создают в своем штате центры по оцифровке своих материалов или отделы создания электронных библиотек. Реализуются многолетние программы по оцифровке таких изданий, например, «Память Америки» в Библиотеке конгресса США, «Память мира. Память России» в Российской государственной библиотеке.

Компактность. Сегодня книги, занимающие в бумажном виде несколько стеллажей, можно уместить на один или несколько компакт-дисков. Специальные устройства для чтения электронных изданий, весящие в среднем один килограмм, вмещают в себя целый книжный шкаф.

Гипертекст. Гипертекст как метод нелинейного чтения давно уже стал фактом современной мировой литературы. Книги написаны так, что их можно читать не только "линейно", последовательно, но и по диагонали или даже по определенной схеме. Комбинация разных литературных "гиперприемов" приводит к различному прочтению одной и той же книги. Часто такие манипуляции с текстом не только при чтении, но и при написании удобно делать в электронном режиме. Нелучайно некоторые современные авторы выпускают книгу сразу в цифровом формате.

Мультимедийные возможности. Явное преимущество электронных книг – сопровождение текста звуком и видеоматериалами. Таким образом, возможности автора и читателя уникально расширяются, а содержание становится более понятным и интересным. И это не просто для развлечения. Некоторые специальные издания становятся при этом более понятными. Описание операции для студентов-медиков, например, или описание шумов сердца для начинающих кардиологов с видео и аудио эффектами приобретает значительно большую ценность.

Электронные библиотеки – бесплатная книга. В последние годы в России хорошая книга превратилась из редкого в дорогой про-

дукт. Электронные библиотеки позволяют найти и прочитать бесплатно нужное издание в Интернете. Поскольку книги часто находятся в этих библиотеках в текстовом формате, их можно легко “скачать” на свой компьютер и прочитать (а если понадобится, то и распечатать. Правда тут возникнет проблема стоимости бумаги и картриджа для принтера).

Возможность найти вышедшую из продажи или редкую книгу. По мнению некоторых книготорговцев на Западе, понятие вышедшей из продажи книги попросту исчезнет. Ведь у электронной книги не может быть тиража в традиционном смысле этого слова. Если представить, что созданная книга никогда не удаляется из архива издателя, ее всегда можно будет приобрести. В настоящее время гарантом сохранности книги служат электронные библиотеки. Если издание вышло из продажи, а повторный тираж не планируется, Интернет – единственный шанс для тех, кто желает прочесть книгу, но опоздал ее купить.

Укрепление контактов между авторами, издателями, переводчиками, читателями. Когда электронная книга доступна в Интернете, читатели часто имеют “прямую” возможность высказать свое мнение автору или издателю. В электронных библиотеках это стало нормальной практикой. Таким образом, расстояние между создателем книги и ее читателем может резко сократиться. Известно, что многие писатели нашли через электронные библиотеки своего издателя или переводчика.

Легкость редактирования. В электронное издание легче вносить исправления, чем в бумажное. Цифровые книги могут постоянно модернизироваться автором. Мы уже говорили о поисковых системах, о возможности составления собственных изданий. Создатели устройств для чтения электронных книг также заботятся о том, чтобы работа с книгой стала как можно удобнее или, по крайней мере, не уступала работе с бумажным изданием.

Доводы против электронной книги

Неудобство чтения с экрана. Известно, что обычный человек способен читать страницу, набранную мелким шрифтом, со скоростью до 400 слов в минуту. (Люди с хорошими неврологическими связями могут читать со скоростью до 600 слов в минуту). Устная речь, переда-

ваемая по телевидению или по радио, может восприниматься со скоростью 150 слов в минуту. Скорость чтения с компьютерного дисплея находится где-то посередине. Зачем же тогда соглашаться на тридцатипроцентное, если не больше, сокращение эффективности чтения? Символы с разрешением 72 точки на дюйм трудно читаются, любой свет, попадающий на экран, порождает слепящие блики. Есть и эстетические соображения: тексты на экранах чаще всего отформатированы в соответствии с параметрами *по умолчанию*, с не выдерживающими никакой критики интервалами и выравниванием, а для их отображения используются стандартные, набившие оскомину шрифты Times или Arial. К тому же вы оказываетесь заложником своего рабочего стола. Даже если вы обладаете портативным компьютером, ваше положение перед экраном далеко от той расслабляющей позы, которую человек принимает в удобном кресле за чтением книги.

Создатели электронных устройств для чтения усиленно работают над исправлением этих недостатков. Корпорация "Microsoft" проводит испытания, выясняющие, не вредит ли чтение экранного текста зрению и не вызывает ли оно головные боли. По словам инженера "Microsoft" Билла Хилла, технология ClearType позволяет программным способом увеличить эффективное разрешение экрана стандартно-ноутбука со 110 до примерно 250 точек на дюйм.

Зависимость от связи с Интернет и электричества. Традиционную книгу можно читать дома, на балконе, на даче и для этого не нужны ни электричество, ни средства доступа к Internet. Говоря об экономических преимуществах электронных книг, нужно решить вопрос, что на окончательном этапе дешевле для потребителя. Цены на услуги Интернет не высоки и будут становиться все доступнее. Текущие ограничения на пропускную способность связи являются временным явлением. Коммерческое использование Интернет требует очень высоких скоростей передачи для таких приложений, как видео по требованию или покупки, не выходя из дома. Такие скорости являются более чем достаточными для распространения всех данных.

Проблема защиты авторского права. Благодатные времена настали сейчас для пиратов авторского права. В бумажном мире пиратские копии тоже существуют, однако потребители все же пытаются достать оригинальные издания, тогда как покупать пиратские CD в России стало нормой не только для нас, но и для иностранных гостей. Пиратская копия электронного продукта подчас ничем не отличается от настоящего ни по качеству, ни по документации, а в большинстве

случаев даже и по поддержке (по крайней мере, в России). Если в недалеком прошлом автор мог технологически контролировать распространение произведений, сейчас он просто не в состоянии отследить движение битов по проводам. И обществу предстоит заменить устоявшиеся законы, на те, которые бы соответствовали текущему положению.

Как известно, довольно многие периодические издания не брезгают перепечатывать материалы из Интернета без ссылки на автора, без выплаты гонорара, без предварительной договоренности.

Отсутствие стандартов на форматы электронной книги. Поскольку стандартов на форматы электронных книг или распространяемые копии текстов пока не существует, для опубликования электронного издания нужно выбрать определенную модель устройства и подготовить текст с помощью собственных инструментов производителя для издания и распространения. В каждой компании есть собственные технологии для верстки публикаций и защиты от копирования, в итоге каждый текст готовится исключительно для определенной платформы электронных книг.

Такая ситуация сохранится только на начальном этапе развития индустрии электронных книг; это повторение сценария, совсем недавно разыгранного в Web. Первопроходцы рынка охраняли "свою территорию" с помощью собственных уникальных технологий, но большинство из них быстро осознали, что для процветания рынка нужны открытые стандарты.

Ненадежность электронных библиотек. Одной из крупных проблем Интернета является непостоянство его «жителей». Интернет приучает человека получать нужную информацию в любой момент, не отходя от стола. Для многих активных пользователей задержка в доступе к необходимым данным является раздражающей, а то и катастрофичной. Однако известно, что сервер, еще вчера предоставлявший необходимую информацию, сегодня может быть отключен, испорчен, взломан.

Неприспособленность психологии к нелинейному чтению. Многие специалисты считают гипертекстовый формат документов, характерный для электронных книг, одним из выдающихся достижений. Наверное, так оно и есть. Однако, по мнению многих читателей, перевод всех произведений в такой формат представляется более чем сомнительным, прежде всего, с познавательной точки зрения. Когда читаешь большой документ (техническую книгу, руководство пользователя,

пояснительную записку), то гипертекст далеко не всегда вызывает восторг. Человеческая психология более приспособлена к последовательному восприятию информации, без "перескоков" с пятое на десятое. Гипертекстовые документы можно сравнить с книгой, где в конце каждого предложения имеется ссылка на другую страницу или печатное издание. Единственное различие такой книги от документации в формате HTML заключается в том, что для перехода на другую страницу или в другой документ достаточно нажать кнопку мыши.

Дорогая инфраструктура. Несмотря на свою экономическую привлекательность (часто бесплатный доступ через Интернет) электронная книга находится неизмеримо дальше традиционной от широкого круга читателей. Для того чтобы начать пользоваться этой новой формой издания, необходим компьютер и подчас доступ к Интернету, а для большинства людей (не только в России) это дорого.

Хотя стоимость инфраструктуры достаточно велика, создатели электронных книг считают, что, учитывая стоимость всех печатных публикаций, которые больше не нужно будет покупать, устройства для отображения электронных изданий окупятся в довольно короткие сроки. Производители подобной техники и потребители надеются на быстрое снижение цен.

Быстрое старение компьютерных технологий. В течение многих лет компьютерные специалисты говорили, что цифровые данные в виде единичек и ноликов будут храниться вечно. Но они ошибались. Испытания доказали, что магнитные пленки живут всего 10 лет в зависимости от условий хранения. Судьба флоппи-дисков, видеопленки, винчестеров тоже туманна. Даже компакт-диски, некогда названные вечными, оказались уязвимы к влиянию магнитных полей, окислению, влажности и разрушению материала, из которого они сделаны. Да и большая часть "железа" и "софта", предназначенного для извлечения информации с хранящихся дисков и пленок, исчезает во имя технического прогресса.

Как было сказано, общий объем информации об электронной книге достигает во всемирной паутине 30 тыс. страниц. Однако мы пока не нашли ни слова о решении проблемы старения компьютерных технологий. В самом же компьютерном мире так и не решен критически важный вопрос старения компьютерных технологий. Несмотря на привлекательность электронных копий, множество библиотек и архивов продолжает снимать свои хрупкие книги на черно-белый микро-

фильм. Потому что эти хранители "памяти мира" думают не о ближайшем десятилетии, а дальше, и боятся идти на риск.

Заменит ли электронная книга бумажную?

Споры по этому вопросу занимают большую часть статей, посвященных электронным книгам. Представители корпорации "Microsoft" сделали ряд смелых прогнозов, заявив, что к 2003 г. электронные книги станут весить менее 500 г, смогут работать от аккумуляторов в течение восьми часов, а стоить будут впятеро дешевле, чем сейчас. Еще несколько лет спустя электронные книги станут продавать в газетных киосках, а через 20-30 лет цифровые издания начнут обгонять по объему продаж своих печатных "собратьев".

Возможно, споры будут идти до тех пор, пока история не докажет, кто был прав. Оптимальным выводом кажется то, что электронная книга в ближайшие годы не заменит бумажную, но будет ее дополнять. Мы рассмотрели многие преимущества электронной книги. Понятно, что в связи с развитием компьютерных технологий число "плюсов" будет увеличиваться: работать с текстами станет легче, издательские программы станут проще, а читающие устройства – удобнее. Многие юристы уже сейчас активно занимаются проблемой авторского права в электронном мире. Возникает впечатление, что сейчас это одна из самых популярных областей работы в юриспруденции. Со временем должны будут выработаться не только законы, но и стандарты.

С другой стороны, остается нерешенной проблема организации нового массива электронной информации, а главное, отбора самой высококачественной, подлинной и достоверной ее части. Постоянно увеличивающееся количество электронных документов усложняет современную информационную систему. В среднем за последние 120 лет вся техническая продукция удваивается каждые 20 лет. После второй мировой войны этот ритм достиг удвоения каждые 10 лет. Скоро потребителям информации придется справляться не просто с ее огромным потоком, а со шквалом. Читателям нужно будет превратиться в настоящие поисковые машины. Выбрать необходимое из бесконечно пополняющегося списка названий произведений и изданий – дело непростое.

Сведения об Интернет

Интернет как глобальная компьютерная сеть

По-английски Internet является аббревиатурой слов INTER-connected NETworks, что означает ВЗАИМОсвязанные СЕти. Интернет представляет собой свободное сообщество независимых сетей, объединенных только общими правилами передачи данных и общими интересами пользователей. Именно возможность совместной работы в Интернет компьютеров разного типа, архитектуры и стандартов, а также ее изначальная ориентация на пользователей неспециалистов в этой области сделали эту сеть универсальной и общедоступной. Универсальность сети привлекала как пользователей, так и поставщиков информации, что определило важнейшую роль сети как быстро развивающейся среды мировых информационных ресурсов. Общедоступность сети – это удобные технические, экономические и правовые условия для всех пользоваться ее ресурсами и службами обмена данными и разнообразными услугами.

Интернет началась в 1969 г. с сети Арпанет (ARPANET – Advanced Research Projects Agency Network), которую финансировал Пентагон с целью предотвратить или по крайней мере минимизировать потерю национальных информационных ресурсов в случае ядерного конфликта, рассредоточив их на тысячах связанных компьютеров, каждый из которых мог работать независимо, сохраняя информацию в гипотетическом случае поражения США. В 1973 г. были созданы первые интернациональные связи Интернет с Англией, а затем – с Норвегией. В 1977 г. начались процессы слияния ARPANET с другими сетями. В 70-е годы в США Интернет базировалась на сети National Science Foundation Network (NSFNet), которая была предназначена в первую очередь для исследователей и запрещала коммерческий трафик. Глобальный характер сети, обширные информационные ресурсы, широкая доступность, массовость и дешевизна услуг, а также многоаспектность

возможностей использования сети и ее постоянное развитие, сделали Интернет привлекательной для пользователей многих стран мира.

В последние несколько лет, когда задача создания национальной информационной инфраструктуры была осознана обществом и поставлена государствами развитых стран как приоритетная, она привлекла к себе многие фирмы – гигантов телекоммуникации и массовой информации. Хотя они не жалели средств на эксперименты и демонстрации скоростных электронных сетей, которые связывали частный дом или рабочее место с мировыми информационными сокровищами, предлагая новые возможности для образования и развлечений, дело так и заканчивалось экспериментами, а то и вовсе разговорами. Этот новый информационный мир был создан в Интернет, причем, в основном усилиями энтузиастов – программистов, компьютерных хакеров, инженеров и студентов. Он быстро развивается по собственным законам и правилам, существуя как бы параллельно с миром, о котором трубят большие корпорации, и без всяких инвестиций уже дал людям то, что корпорации даже не смеют пока обещать.

Каждый человек видит секрет успеха Интернет по-своему. Одни считают, что Интернет олицетворяет успех нерегулируемого, настоящего свободного рынка, триумф рыночной экономики над плановой. Другим кажется, что в Интернет проявились преимущества малого бизнеса перед крупными фирмами и корпорациями. Третьи, глядя на Интернет, делают вывод о неизбежности победы демократии над диктатурой, четвертые – о силе неконтролируемого индивидуального созидания, и объясняют быстрый рост Интернет тем, что сеть развивается сама по себе и постоянно находит неудовлетворенный спрос. Некоторым кажется, что Интернет – это и есть информационная супермагистраль будущего, которая просто выглядит по-другому, чем существующие диалоговые службы информационных компаний. Однако в Интернет до настоящего времени так и не нашли отражения важнейшие черты информационной супермагистрали, такие как скорость передачи информации, интеграция всех каналов коммуникации от телефона до мобильной связи и спутникового телевидения.

Скептики, безусловно, правы, когда отмечают, что сегодня Интернет выглядит хаотичной и разочаровывающей при работе с представленной в сети информацией. В Интернет еще мало средств, ориентирующих пользователя в ресурсах сети и облегчающих утомительный поиск и анализ информации. Использование ее мультимедийных возможностей по силам только тем, кто имеет доступ к высокоскорост-

ным каналам, так как даже при небольших скоростях очень легко утонуть в море информации, случайных фактов и пустой риторики.

Действительно, после появления важнейшего компонента современной Интернет – World Wide Web (WWW) всемирной паутины серверов, связанных взаимными ссылками и предлагающих информацию в мультимедийной форме от текста до видеоизображения и звука, – стало ясно, что объем малополезной информации в сети растет более высокими темпами, чем сама сеть (хотя рост сети превышает все ранее достигнутые показатели роста, например, связанные с распространением факс-машин или персональных компьютеров). Это явление получило название феномена аквариума. Когда в рамках Интернет стало возможным создавать так называемые домашние страницы пользователей, многие из них содержали чепуху, неинтересную никому, кроме хозяина, например, изображение его аквариума. Однако этот информационный шум не в состоянии обесценить те ресурсы, которые привнесли в сеть миллионы ее участников, ранее выступавших лишь как пассивные потребители.

Первоначально Интернет создавалась как бы на общественных началах, получая мощную поддержку со стороны правительств ряда высокоразвитых стран (прежде всего США), а также крупнейших корпораций. Допустив коммерческое использование, она быстро вышла за рамки объективных пределов роста, характерных для частных и специализированных сетей и обусловленных недостатком потребителей. Сегодня Интернет может рассматриваться как основа информационной инфраструктуры мира и, несмотря на уже достигнутые успехи, имеет большие резервы для дальнейшего развития. Наиболее продуктивными ее пользователями пока все же являются компании и университеты, работающие с сетью через арендованные каналы связи. Для индивидуальных пользователей, работающих дома по обычным телефонным линиям, сеть WWW пока недостаточно эффективна, так как эти линии связи являются слишком медленными и не позволяют использовать мультимедийные возможности WWW в полной мере.

Со второй половины 80-х годов в качестве коммерческого продукта особую популярность получили базы данных на компактных оптических дисках (CD-ROM). Они позволяют организовать диалоговый поиск информации непосредственно на компьютере потребителя, без обращения к услугам операторов и сетей передачи данных. Этим они привлекают новых конечных пользователей. Базы на компакт-дисках оказались наиболее удобной формой вывода на рынок мультимедий-

ной информации и технологий типа гипертекст. Именно появление такой информации в значительных объемах и соответствующих технологий и стандартов типа HTML (Hyper Text Marked Language) подтолкнуло технологические инновации, позволившие создать WWW в рамках Интернет. Технология гипертекста, являющаяся изюминкой WWW, дает возможность связывать документы через систему ссылок, строить сами документы иерархически и переходить от одного фрагмента текста к другому простым нажатием на кнопку мыши.

Присоединение к Интернет критической массы пользователей (5–6 млн) привело к резкому росту объемов информации, доступной в сетевом режиме. Не рассчитывая на прибыль, тысячи людей создали миллионы страниц информации, доступных в диалоговом режиме, начиная от баз данных НТИ и кончая личными дневниками или мини-каталогами ресурсов самой Интернет. Одни подготавливают эту информацию по личным мотивам – из тщеславия, желания высказаться, обмениваться идеями, потому что вхождение в Интернет кажется простейшим способом прославиться или удовлетворить личные притязания на реализацию своей роли в обществе. Другие через Интернет пытаются обойти недоступные для них каналы массовой информации, маркетинговые каналы или каналы для связи с общественностью. Третьи видят в Интернет зарождение нового мира и хотят принять участие в этом процессе. Четвертые работают просто потому, что Интернет существует и их ничто не ограничивает. Действительно, если витрина представляет интерес, то десятки, а то и тысячи людей могут посещать ее ежедневно.

До появления Интернет основными поставщиками услуг для конечных потребителей были центры обработки баз данных (ЦОБД или, как их называют в США, vendors, а в Европе – hosts). Они осуществляют обработку БД, полученных от генераторов, и предоставляют потребителю услуги по работе с ними. Прежде всего эти центры организуют диалоговый доступ к большим базам данных, а иногда также и пакетный поиск для неограниченного круга пользователей. Как правило, базы данных приобретаются на условиях аренды или лицензирования. Некоторые центры специализируются также на подготовке и распространении баз данных на компакт-дисках. В России специализированных ЦОБД фактически не было и нет, и данные задачи традиционно решались крупными генераторами в области НТИ: Всероссийским институтом научной и технической информации (ВИНИТИ), Институтом научной информации по общественным наукам (ИНИОН), Все-

российским центром научно-технической информации (ВНТИЦ), Всероссийским институтом межотраслевой информации (ВИМИ).

Несмотря на доступность современных видов информационных услуг для конечного потребителя значительная их часть все еще попадает к непосредственным пользователям через традиционные каналы информационного обслуживания – библиотеки и информационные подразделения предприятий и организаций. В большинстве случаев как библиотеки, так и отделы информации взаимодействуют со своими потребителями на некоммерческой основе, однако общее удорожание информационных услуг в сочетании с тенденцией ограничения бюджетных ассигнований заставляет их внедрять элементы коммерческих отношений. Появление возможностей работы с Интернет через библиотеки, по-видимому, не только не подорвет их традиционные услуги, но, напротив, вдохнет в них новую жизнь, так как позволит привлечь новых читателей.

В последние годы содержание понятия постиндустриального общества было переосмыслено, и его экономика теперь воспринимается не как экономика услуг, и даже не как информационная экономика, а как экономика, основывающаяся на знаниях. В развитых странах не менее половины валового национального продукта вырабатывается в непроизводственном секторе, а более 25 % промышленной продукции производится в высокотехнологических отраслях, т. е. тоже основывается на знаниях. Работники в области знаний стали основой благосостояния современного общества. Во многих западных компаниях появилась новая должность – менеджер знаний, то есть формируется группа своего рода собственных информационных брокеров и даже вводится должность директора по знаниям.

В Интернет WWW предлагает пользователю услуги качественно нового уровня, которые не могут быть получены при представлении информации в печатной форме или даже на компакт-диске. Основным элементом новизны при этом выступает возможность для потребителя, работающего с текстом, без клавиатурных и интеллектуальных усилий практически мгновенно переходить от одного фрагмента текста к другому, связанным с ним. При этом связи отдельных объектов (слов, понятий, графических символов) используются не только в ограниченных рамках данного текста (что возможно в случае компакт-диска), но и с текстами из множества удаленных баз данных, которые могут содержать значительно более полную информацию в любой форме, включая изображения, звукозаписи и т. п.

Другой новой уникальной услугой, доступной, правда, не только в Интернет, но и через коммерческие диалоговые службы, выступают персонализированные газеты и бюллетени. Эти газеты представляют собой результат работы специальных систем по отбору и фильтрации в реальном масштабе времени информации по определенной теме или проблеме из множества источников данных в электронной форме для индивидуального пользователя, помещаемые непосредственно в его электронный почтовый ящик. Отличие данной услуги от традиционно избирательного распространения информации состоит в том, что поиск ведется не по одной или нескольким базам данных, как в случае диалоговых служб, а одновременно по сотням и тысячам серверов Интернет. Информация поступает пользователю не периодически (каким бы коротким не был период), а в виде непрерывного потока новостей, причем, при несравнимо более низких затратах, чем в коммерческих центрах обработки баз данных.

Данные услуги сначала появились на серверах внутренних клиентских сетей крупных организаций, но уже с осени 1995 г. на рынке стали предлагать специальные версии для WWW, которые используются не только в корпоративных сетях, но и для организации публичного доступа в рамках Интернет. Примером выступает служба доставки оперативной спутниковой информации DirectPC, разработанная Hughes Network Systems и используемая корпорацией Network News, которая интегрирует персонально-ориентированную информацию, поступающую из десятков спутниковых информационных каналов, таких как Knight-Ridder, Newsbyte или BusinessWire.

Бил Гейтс – владелец фирмы Майкрософт – считает, что даваемая Интернет миллионам людей возможность высказаться и опубликовать свое мнение или информацию при незначительных затратах, является одним из важнейших социально-экономических результатов, привнесенных этой системой, и будет способствовать как развитию свободы слова, так и индивидуального творчества. Другим изменением, которое вызвала Интернет в обществе, стало формирование собственной культуры, появление своеобразного стиля общения и даже сетевого этикета (явление, которое ранее в информационной деятельности не было замечено или не рассматривалось как существенное). Через этот этикет массовый потребитель стал оказывать существенное влияние на поведение информационных служб и на развитие и формы предпринимательства в сети. Этикет сети даже получил собственное название Netiquette (сетевой этикет).

Среди наиболее активных пользователей специалисты выделяют так называемых netizens – граждан сети, число которых пока составляет лишь один процент от общего числа пользователей. Netizens не жалуют денег и рассматривают свои затраты на работу в сети в качестве

расходов на удовлетворение важнейших жизненных потребностей, таких как потребность в пище, жилье или развлечениях. Рост числа Netizens будет всячески стимулироваться информационными службами в качестве одного из главных источников будущих доходов.

Организация доступа к первоисточникам

Научно-техническая информация

Интернет представляет собой один из основных источников доступа к публичным информационным ресурсам в области НТИ. С одной стороны, эта сеть с самого начала предназначалась именно для распространения такого рода информации между специалистами. С другой стороны, в США и в некоторых других странах с начала 90-х годов Интернет официально стал рассматриваться в качестве основного канала доступа к научно-технической информации самых широких кругов общественности и, частично, именно в связи с этим и был открыт для публичного доступа. Примерами ведущих Web серверов, предлагающих доступ к большим базам НТИ для пользователей Интернет, могут служить:

EasyNet 2.0 (\$) (<http://www.telebase.com/>) – коммерческая служба фильтрации и доставки информации через Интернет, организованная шлюзовой службой доступа к более чем 250 базам данных. Доступ к службе бесплатный, но пользователь платит за каждый поиск от \$2 до \$7 за полный текст статьи. Особенностью выступает мощная служба поддержки пользователя по электронной почте, которая позволяет потребителю точно сформулировать запрос и провести поиск с минимальными затратами;

UnCover Periodicals (<http://www.carl.org/uncover/unhome.html>) – служба диалогового поиска в базах данных более чем тысячи библиотек, охватывающих более 17 тыс. наименований периодических изданий;

Internet Patent Search (<http://sunsite.unc.edu/patents/intropat.html>) – сервер, предлагающий услуги поиска патентной информации и доступа к полному тексту патентов с иллюстрациями.

Следует отметить, что приведенные выше службы Интернет обеспечивают доступ к НТИ не только в области фундаментальных, прикладных и технических наук, но и в области общественных наук и прежде всего экономики, управления и т. п.

Некоторые информационные службы, представленные в Интернет, взяли на себя заботу об обеспечении пользователей копиями статей из научно-технических и деловых периодических изданий. Для России это является особенно актуальным, так как за годы реформ средства, выделяемые бюджетом для подписки на зарубежные журналы и книги, сократились почти до нуля и даже ведущие научно-

технические библиотеки страны, такие например, как ГПНТБ или РГБ и информационные центры, например, ВИНТИ или ИНИОН остались фактически без источников комплектования. Правда, некоторые зарубежные издания удается получать в рамках книгообмена, но этот источник является случайным и нерегулярным. В последнее время все больше научных журналов доступны через сеть в электронной форме, так как это позволяет снизить издательские затраты на 70 %. В числе служб – диалоговых платных источников копий статей, представленных в Интернет:

UMI InfoStore (<http://www.umi.com/ach/index.htm>) – служба коммерческой компании, предоставляющая пользователю доступ к статьям из 15 тыс. сериальных изданий, на основе которых подготавливаются ее базы данных;

HeadsUpNewsService

(gopher://gopher.eneews.com:2100/11/news_services/headsup/)– служба фильтрации и доставки, предоставляющая пользователю ежедневный доступ к 12 тыс. статей из сериальных изданий по цене \$3 за статью. При этом в электронный почтовый ящик сначала приходят библиографические описания, на основе которых делается окончательный заказ текстов статей.

Для российского пользователя, не привыкшего задумываться о соблюдении авторских прав, любопытным может оказаться, что упомянутые выше службы уделяют этим вопросам самое серьезное внимание и, пользуясь их услугами, можно быть абсолютно уверенным, что права ни автора, ни издателя не будут нарушены. В мире уже выработаны определенные правила в этой области, предусматривающие тщательный учет случаев соответствующей платы авторам и издателям. Интересно, что несмотря на публичную ориентацию НТИ, многие крупные издатели и поставщики баз данных, специализирующиеся на этом виде информации и имеющие свои WWW серверы и Web страницы, доступ к информации этого типа предоставляют за плату даже в том случае, если они являются профессиональными обществами или получают бюджетную поддержку. Примерами могут служить:

- * CAS (<http://www.cas.com/>),
- * NTIS (www.ntis.gov),
- * NASA Star (<http://www.nasa.gov/>),
- * STN International (<http://www.karlsruhe-fiz.de/>).

Причин здесь несколько и основной, вероятно, является то, что именно государственные и полугосударственные ЦОБД были среди первых информационных служб, которые государство в целях экономии бюджетных средств заставляло внедрять элементы хозрасчета и самофинансирования. Видимо привычка настолько укоренилась, что смогла преодолеть традиции бескорыстия, характерные для некоммерческих серверов и Web страниц Интернет. Подобная ситуация харак-

терна и для России. Например, прекрасный источник доступа к НТИ – сервер ВИНТИ (<http://viniti.msk.su>), дающий возможности поиска в текущих и ретроспективных базах данных около 30 наименований с годовым приростом до 1 млн записей, – также предлагает свои услуги только за плату. Цены являются даже более высокими, чем в случае западных аналогов (36 тыс. руб. в час плюс 300 руб. за просмотр одной записи в ценах 1997 г.). С другой стороны, Web страницы ВИНТИ организованы в соответствии с лучшими стандартами и ничем не уступают страницам других западных источников НТИ, например, такому как STN International. При этом, пользователь может загрузить в свой компьютер методические материалы, например, описание формата, тезаурус, методiku выполнения поиска и т. п., но также за плату.

Надо заметить, что с НТИ связано и большинство работающих в России серверов Интернет, которые первоначально в основном создавались на базе институтов Российской академии наук и ведущих университетов и ВУЗов про поддержке со стороны государства. Большую роль сыграли и международные программы, а также деятельность ведущих компьютерных фирм мира, например, SUN Microsystems. Поэтому и небезызвестный Д. Сорос также способствовал развитию сетевой активности в области НТИ, так как старается предоставлять свои гранты на развитие телекоммуникации только при условии, что финансирование не меньшего объема будет выделяться из российского бюджета. Среди других российских Web серверов, предлагающих доступ к НТИ:

МЦНТИ (<http://www.icsti.su/>) – сервер, предлагающий базы данных НТИ по различным областям знаний;

РИНКЦЭ (<http://www.extech.msk.su/>) – сервер, содержащий список государственных научных центров по отраслям науки, а также тексты документов, формирующих нормативно-правовую базу науки и техники в России;

RD MNTS-Service Company (<http://www.mnts.msk.su/>) – сервер, охватывающий информацию о российских научно-технических достижениях;

РосНИИРОС (<http://www.ripn.net/>) – сервер, обеспечивающий в рамках службы ИНФОМАГ доступ к оглавлениям научно-технических журналов и зарубежных электронных бюллетеней, кратким тезисам статей [96].

Instauratio Magna (<http://www.free.net/IM/main.ru.html>) – информационная система ВУЗов России, создаваемая по программе Госкомвуза РФ – Национальная система баз данных и баз знаний высшей школы России;

Курьер РАН и высшей школы (<http://www.free.net/courier/ENTRY.ru.html>) – электронный ежемесячный журнал, включающий материалы о жизни научного сообщества России и международную информацию (сведения о грантах, вакансиях, зарубежных научных организациях);

Международная поддержка российской науки и высшей школы – по адресу (<http://www.free.net/ENTRY.ru.html>) – страница, содержащая справочник по негосударственным источникам финансирования российской науки и образования;
Казанский гос. университет (<http://www.ksu.ras.ru/home.ru.html>);
Корпорация *Университетские сети знаний* (<http://www.rc.ac.ru/ENTRY.ru.html>);
Нижегородский гос. университет (<http://www.unn.ac.ru/>);
Новгородский гос. университет (<http://www.novsu.ac.ru/entry.ru.html>) – сервер, который наряду с НТИ содержит информацию о Новгороде (туризм, путешествия, жизнь города, архитектура, археология, иконопись);
Пензенский гос. технический университет (<http://www.stup.ac.ru/main.ru.html>);
Пермский гос. технический университет (<http://www.pstu.ac.ru/>);
Ростовский гос. университет (<http://www.unird.ac.ru/>);
Ярославский гос. университет
(<http://www.yars.free.net/Russian/Univ/main.ru.html>);

Перспективы развития данного сектора Интернет в России представляются обнадеживающими, и можно надеяться, что российские академические и учебные службы WWW станут своего рода инкубаторами для запуска на их основе коммерческих служб.

Научная информация в области экономики может быть получена из следующих Web страниц:

Economic Literature Index (EconLit) (<ftp://kermit.columbia.edu/kermit>) – одно из ведущих в мире собраний библиографической научной информации в области экономики;

Review of Economic Studies (<http://www.ecn.bris.ac.uk/Restud/revstud.htm>) – сервер, предлагающий информацию о результатах исследований в области экономики с 1991 г.

Политическая информация

Информация этого типа получила широкое распространение в Интернет, прежде всего, в связи с попытками использования сети в политических целях (например, в предвыборной борьбе) – сначала в качестве средства массовой информации, а затем и массовой коммуникации. Различие между массовой информацией и коммуникацией состоит в том, что в первом случае пользователь пассивен, а во втором с ним возможен прямой диалог, который, с одной стороны, требует от политической инфраструктуры более тщательной и внимательной работы с аудиторией (так как дозировать информацию больше не удается), а, с другой стороны, позволяет оказывать на мысли и поведение человека более сильное воздействие.

Специалисты выделяют четыре основных причины популярности использования Интернет в политической борьбе: (1) низкая стоимость создания Web страниц для любого пользователя; (2) неогра-

нические возможности размещения и распространения информации; (3) выход на избирателя нового типа, ранее недоступного; (4) слабые рычаги регламентирования, цензуры и фильтрации.

Через Интернет избиратели могут получить пресс-релизы, тексты речей, биографии, видео клипы, аудиозаписи. Следует отметить, что среди российских политиков уровень ориентации на новые технологии и их оперативное восприятие достаточно высок. В России имеется Web сервер Федеральной службы новостей (<http://www.ru/fns/>), которая приводит полные тексты пресс-конференций и интервью с ведущими российскими политиками.

В области российской политической информации работают как отечественные, так и зарубежные серверы. Например, *Политика* (<http://www.ru/>), предлагающая базу о политических партиях, движениях, ведущих политиках, созданная совместно с агентством *Панорама*. Блок информации *Национальной службы новостей* (НСН) "Федеральная власть России" содержит информацию о Президенте, Правительстве, Федеральном собрании, судебной власти и Центризбиркоме. Другие блоки политической информации НСН включают описания политических партий и общественных движений России, телефонный справочник федеральных органов РФ, хронику законодательной деятельности (подробные отчеты о работе Федерального собрания), календарь ожидаемых политических событий, описание групп влияния (средств массовой информации, аналитических центров, экономических структур), провинциальную хронику и справочные материалы по регионам, карты. НСН предлагает клиентам доступ к фактографической базе, в которой содержатся досье государственных и политических деятелей федерального и регионального уровней, бизнесменов, промышленников и иных влиятельных лиц.

Новости и литература в электронной форме

Новости и литература в Интернет являются одним из основных информационных ресурсов. Интернет обеспечивает более оперативный доступ к новостям, чем сами агентства новостей, так как работая с сетью, пользователь получает доступ к источнику новостей в электронной форме практически одновременно с их генератором – Reuter, Time Warner, The New York Time, CNN и др. – то есть находится с ним в примерно равных условиях.

Использование сервера новостей имеет ряд преимуществ по сравнению с телевизионной программой: интерактивный доступ к большим объемам информации и простота выбора из них, возможность сохранения нужной информации в мультимедийном виде (с фото, картами, интервью) и т. п. Большинство ведущих газет и журналов мира сегодня уже имеют серверы Интернет, и через них необходимую информацию можно получить намного дешевле и быстрее, обычно еще до выхода печатной версии.

Важнейшие серверы новостей, доступных через Интернет:

* CNN (<http://www.cnn.com>) – предоставляет доступ к новостям еще до отражения их на экранах телевизоров, а информация обновляется каждые полчаса и группируется по нескольким основным разделам;

* The New York Times (<http://www.nytimes.com>) – информация появляется за несколько часов до выхода газеты.

Среди служб, уже сегодня предлагающих пользователям персональные газеты, подготовленные с использованием систем фильтрации информации и доступные через Интернет:

* The Times (<http://www.the-times.co.uk/>) – бесплатная персональная газета Personal Times, подготавливаемая на основе материалов газет The Times и Sunday Times;

* IBM (<http://www.infosage.ibm.com/>) – персональная газета InfoSage, подготавливаемая на основе материалов служб новостей и предлагаемая по подписке;

* Los Angeles Times (<http://www.latimes.com/>) – бесплатная персональная газета Hunter, подготавливаемая на основе материалов LA Times и агентств новостей;

* Adaptix (<http://www.daptix.com/>) – бесплатная персональная газета Rex, подготавливаемая на основе новостей, представленных в Интернет;

* Pointcast (<http://www.pointcast.com/>) – бесплатная персональная газета Pointcast Network (PN), подготавливаемая на основе материалов агентств новостей и некоторых ведущих американских газет.

Почти все ведущие российские службы новостей имеют собственные серверы WWW. В качестве примеров можно назвать:

* РИА Новости (<http://ss10.elvis.msk.su/alexey-bin/RIAN-Index.rus>);

* TACC (<http://www.trib.com/WIRE/TASS/INTERNATIONAL/>);

* TADI (<http://www.tadi.tomsk.su/>) – сервер информационного агентства

TADI.

Сектор новостей привлек внимание и новых коммерческих структур. Например, в середине мая 1996 г. компания Bit Software устроила презентацию работы систем распознавания текста в Национальной службе новостей (НСН) (<http://www.nns.ru>), в рамках которой в компьютерную форму преобразуется около 240 российских периодических изданий (до 800 газетных страниц ежедневно), включая 160 ежедневных газет. Этот проект может принести России своеобразный аналог службы Lexis-Nexis. НСН не только предоставляет доступ к

тексту изданий, но и включает в себя несколько тематических и обзорных блоков, например, таких как:

Информационно-справочная система Выборы; Федеральная власть России; Политические партии и общественные движения; Группы влияния (СМИ, аналитические центры, экономические структуры); Дайджесты и обзоры СМИ; Пресс-рейтинги; Регионы России; Телефонный справочник федеральных органов РФ; Карты стран бывшего СССР; Хроника законодательной деятельности; Календарь ожидаемых событий.

Многие российские электронные газеты и журналы имеют электронные версии. Помимо уже названных это:

- * Аргументы и факты (<http://205.198.180.2/AiF/>);
- * Независимая газета (<http://www.glas.apc.org/ng/index.html>);
- * Учительская газета (<http://www.redline.ru/ug/>);
- * Иностранец (<http://www.elvis.msk.su/koi8/foreigner/index.html>);
- * Экстра-Пресс (<http://angar.dtk.kiae.su/Express/Welcome.html>);
- * Эхо-Инфо (<http://www.uniphys.spb.su/IHQ/HTML/HelpInfo.html>);
- * С.-Петербургское

эхо

(<http://www.dux.ru/win/enpp/newspapers/echo/echohome.html>);

* Неделя (<http://www.astro.ru/centro/>) – электронная газета, включающая два раздела: *справочная информация* (справочные базы данных и в т. ч. коммерческая информация); *развлекательная информация* (кино, театры, концерты, программа телевидения, музеи, выставки и т. п.);

- * Сегодня (<http://www.eastview.com/segodnia/index.html>);
- * St. Petersburg Press (<http://www.spb.su/sppres/>);
- * Экспресс-хроника

(<http://www.online.ru/nlist/expchronicle/choronicle~weekly/>).

В числе российских электронных изданий журналов, доступных через Интернет, можно назвать:

- * Огонек (<http://www.ropnet.ru/ogonyok/win/tech-sup.html>);
- * Скифы (<http://www.simtel.ru/cgi-bin/snews.pl/simbirsk.skyfy>) [96];
- * Лица (<http://www.izhmark.udmurtia.su/~vd.faces/>);
- * Реклама-Шанс (<http://www.chance.sp.su/>);
- * Ярмарка (<http://www.mplink.ru/public/jarmarka/Head.html>);
- * Независимое военное обозрение (<http://glasnet.ru/ng/>);
- * Знакомства (<http://www.astro.ru/cgi-bin/emz/emz/>).

Говоря о литературе в Интернет, специалисты отмечают, что эта сеть как бы продолжила традицию самиздатовских ксероксных элитарных литературных, политических и музыкальных журналов, распространяемых всего в нескольких десятках экземпляров, и тем ни менее, получающих значительный общественный резонанс. Эти бюллетени зародились еще в рамках систем электронной почты и получили наименование e-zine (вместо magazine – печатный журнал), а теперь именуются web-zines. Главным в этих изданиях, распространяемых через Web, выступает не только доступность и широта распространения, но и гипертекст и мультимедийность.

В Интернет помимо деловых и специальных уже есть настоящие литературные и литературно-критические журналы, например:

- * Salon (<http://www.salon1999.com>);
- * Hotwired (<http://www.hotwired.com>);
- * Word (<http://www.word.com>);
- * Feed (<http://www.feedmag.com>);
- * Suck (<http://www.suck.com>) [36].

Российская электронная литература представлена Публичной электронной библиотекой содержащей произведения Есенина, Пушкина, Тургенева, Чехова (<http://win.www.online.ru/sp/eel/russian/>),.

Среди российских литературно-художественных электронных журналов, доступных через Интернет:

- * Не спать – журнал для ночных жителей Москвы (<http://www.enjoy.ru/>);
- * Пятое колесо (<http://www.chance.sp.su/>);
- * Комиксы (<http://win.www.online.ru/rmain/rentert/rcomics.xhtml>);
- * Magazine – Online (<http://win.www.online.ru/sp/magazine/>);
- * Юношеская газета (<http://www.glas.apc.org/~asebran/edut/publicat.html>);
- * Петербургский журнал искусств (<http://warcom.spb.su/art/frontdoor.html>);
- * БИТ (<http://www.spb.su/beat/1/koi/index.html>)

Некоторые Web страницы по русской литературе представлены на зарубежных серверах. Например, на странице *Русская литература в Интернет* собраны ссылки на все существующие в WWW российские газеты и журналы, а также на коллекции прозы, поэзии и песен (<http://iaiwww.uni-muenster.de/cgi-bin/simplex/lat/lit.html>).

В числе ведущих источников справочной информации общего назначения в Интернет:

- * Argus Clearinghouse (<http://www.lib.unmich.edu/chouse/>) – справочник по предметно-ориентированным справочникам ресурсов Интернет;
- * CIA World Factbook (<http://www.odci.gov/cia/publications/96fact/index.html>) – сервер ЦРУ, через который предоставляется информация по всем странам мира, включая экономическую статистику, карты, характеристику политической обстановки и т. п.
- * Britanica (<http://www.eb.com/>) – электронная Web версия Британской энциклопедии, включающая более широкий состав статей и иллюстраций, чем печатная и 200 тыс. гипертекстовых связей;
- * Webster's Dictionary (<http://c.gp.cs.cmu.edu:5130/prog/webster>) – Web версия словаря Webster с разнообразными гипертекстовыми связями.

В Web представлены все крупные издательства, которые предлагают пользователю информацию о своей работе и планах, а также каталоги публикаций и интервью с авторами. На многих страницах можно ознакомиться с проспектами еще не вышедших книг или статьями в работе, а также деловыми новостями из области издательской деятельности и полиграфии. Среди Web страниц издательств:

- * Bantan Doubleday Dell (<http://www.bdd.com/>);

- * HarperCollins (<http://www.harpercollins.com/>);
- * Random House (<http://www.randomhouse.com/>);
- * Penguin (<http://penguin.com/>);
- * Little, Brown (<http://pathfinder.com/twep/Library/Library.html>) [42].
- * McGraw-Hill (<http://www.mcgraw-hill.com/>);
- * Macmillan (<http://www.mcp.com/>);
- * Oxford University Press (<http://www.oup.co.uk/ouphome.html>);
- * Ventana (<http://www.vmedia.com/>);
- * West Publishing (<http://www.westpub.com/>);
- * John Wiley & Sons (<http://wiley.com/>);
- * Ziff-Davis (<http://ziff.com/>) [42].

Развлекательная и бытовая информация

Потребительская развлекательная и бытовая информация представлена ресурсами в области: образования; музыки; музеев, галерей и художественных выставок; фильмов и фото; спорта; развлечений; бытовой информации.

Образование

В сети широкое распространение получили различные обучающие программы для самообразования, например:

- * ICS Learning Systems (<http://www.icslearn.com/>) – сервер одной из крупнейших в мире служб заочного образования, организованной в 1890 г. и с тех пор подготовившей более 350 тыс. специалистов [42];

- * CIBER Web (<http://www2.mgmt.purdue.edu/Centers/CIBER/ciber.htm>) – сервер Центра международного образования и исследований в области бизнеса (Center for International Business Education and Research – CIBER) – правительственного учреждения, предназначенного для содействия в развитии бизнес-образования в США. Центр обеспечивает связи с серверами 25 крупнейших университетов США;

- * Training & Seminar Locator (<http://tasl.com/tasl/home.html>) – сервер, предлагающий доступ к базе данных по семинарам и учебным курсам с возможностями тематического поиска;

- * AccessNet (<http://accessnet.ru/>) – страница, содержащая список школ по всему миру с указанием специализации.

Ресурсы Интернет в области образования пока не получили значительного развития в России. Тем не менее, на WWW сервере сети FREEnet (<http://www.free.net>) установлена информационная система о высшем образовании в России Instauratia Magna, содержащая данные о 40 российских университетах и 25 тыс. страниц разнообразной информации в области образования.

Музыка

Доступ к страницам, предлагающим потребителям развлекательную информацию в области музыки и музыкальные произведения, постоянно расширяется. В России

такого рода страницы также имеются в достаточно большом количестве и предлагают пользователю многообразие жанров:

- * Interactive Music Encyclopedia (Announcements)
(<http://www.fe.msk.ru/win/ime.html>);
- * Музыка (<http://win.www.online.ru/rmain/rentert/rmusic.xhtml>);
- * Фестиваль экстремальной музыки
(<http://www.fe.msk.ru/win/feelee/welcome.html>) – страница, на которой представлена информация о рок-группе Наутилус Помпилиус, группе ЛЮБЭ, Наталии Штурм и Александре Новикове;
- * Русская классическая музыка
(<http://www.fe.msk.ru/win/records/welcome.html>);
- * Авторская песня (<http://win.www.online.ru/mlist/ksp/news/>);
- * Alien Network (<http://lobin.alien.ru/>) – музыкальная страница, предлагающая тексты песен.

Музеи, галереи и художественные выставки

- * Smithsonian Institute (<http://www.si.edu>) – сервер, предлагающий коллекцию из 500 произведений искусства Национального музея американского искусства и портреты их авторов;
- * WebMuseum (<http://sunsite.unc.edu/wm>) – постоянно расширяющаяся коллекция наиболее известных произведений живописи, формирование которой начал парижский студент в начале 1995 г.;
- * Real Louvre (<http://www.louvre.fr>) или (<http://www.paris.org>) – сервер, предлагающий доступ к коллекции музея Лувра, а также других парижских музеев, включая D'Orsay ;
- * Christus Rex (<http://www.cristusrex.org>) – коллекция частной некоммерческой организации, распространяющей информацию о произведениях искусства, находящихся в церковных учреждениях в Риме, Ватикане и по всему миру;
- * Andy Warhol Museum (<http://www.warhol.org/warhol/>);
- * George Eastman House International Museum of Film and Photography
(<http://www.itrit.edu/gehouse>) – сервер музея кинематографии и фотографии;
- * A Passion for Art – коллекция полотен постимпрессионизма, принадлежащая Фонду доктора Альберта Барнса (Филадельфия). Бил Гейтс, хозяин Microsoft, владеет также компанией Corbis, специализирующейся на создании компьютерной базы данных изображений различного характера. Эта компания в настоящее время владеет, по-видимому, одной из самых больших диверсифицированных баз данных визуальной информации, накопленной человечеством.

Справочная информация по музеям и выставкам также представлена в Интернет:

- * The World Wide Web Virtual Library Museums Page
(<http://www.comlab.ox.ac.uk>);
- * Cultural Treasures of the Internet
(<http://shirley.cs.widener.edu/clark/clark/html>);
- * World Wide Arts Resources (<http://www.concourse.com/wwar/default.html>).

В числе российских музейных страниц в области живописи и изобразительного искусства:

- * ЦНИИРТК (<http://www.rtc.neva.ru/>) – страница Русского музея;
- * Галерея Гельмана – (<http://www.fe.msk.ru/win/gallery/guelman/welcome.html>) или (<http://win.www.online.ru/rmain/rart/guelman.rcgi>) – сервер самой известной в Москве галереи современного искусства;
- * Галерея фирмы RASER (<http://win.www.online.ru/Art/Exh/indexRussian.html>);
- * Выставка икон 11–16 веков (<http://www.UKG.Wiv.Ua/Gallery/MainHall.html>);
- * Contemporary Russian Fine Arts (<http://www.kiae.su/www/wtr/artinfo/begin.html>);
- * Петербургский вернисаж (<http://palantiri.spb.su/English/Companies/Addvertisement/Petersburg.Art.Show/index.html>);
- * Митьки (<http://www~personal.umich.edu/~yegor/mitki/>);
- * Связь-Сервис (<http://www.bryansk.ru/>) – сервер, имеющий страницу, посвященную Сальвадору Дали (биография, альбом).

Кино и фотография

- * Фильмотека (<http://win.www.online.ru/rmain/rentert/rfilm.xhtml>);
- * Russian Top Cartoonist – Victor Bogorad (<http://www.spb.su/victor/>);
- * ElectronicPhotoGalleryHotPictures (<http://www.kiae.su/www/wtr/hotpictures/gallery.html>).

Спорт

Россия Онлайн Спорт (<http://win.www.online.ru/rmain/rentert/rsport.xhtml>).
The Russian Football Homepage (<http://www.quark.lu.se/~oxana/football.html>) – данные о футбольной статистике, составе команд, результатах игр Российской сборной и клубных команд [96].

Развлечения

Игры (<http://win.www.online.ru/rmain/rentert/rgames.xhtml>);
Танцы (<http://win.www.online.ru/rmain/rentert/rdance.xhtml>);
Юмор (<http://win.www.online.ru/rmain/rentert/rhumor.xhtml>);
Русский юмор (<http://www.pa.uky.edu/~verner/lat.cgi/an.html>);
Н.Ж.М.Д. (<http://pobox.com/~tema/rrr/choose.html>).

Бытовая информация

Примерами бытовой информации, представленной в Интернет, выступают многочисленные серверы погоды, серверы полезной информации, такой как программы телевидения, а также серверы, создаваемые коммунальными службами для информирования жителей о текущих районных новостях:

- * Local Time Around the World (<http://www.hilink.con.au/times/>) – сервер, через который можно получить сведения о местном времени в большинстве городов мира;

* World Telephone Area Codes
(gopher://gopher.cs.ttu.edu:4320/worldareacode) – гофер по телефонным кодам большей части городов мира;

* Currency Conversions (<http://bin.gnn.com/cgi-bin/gnn/currency>) – система расчета курсов всех валют мира через основные конвертируемые валюты.

Бытовая информация, доступная через Интернет в России:

* Институт космических исследований
(<http://smis.iki.rssi.ru/dataserv.htm#wmap>) – спутниковая карта погоды:

* TV Program (Classical Music) (http://www.fe.msk.ru/win/tv_prog.html);

* Программа телепередач (<http://win.www.online.ru/rmain/rentert/rtv.xhtml>).

* Сервер Московского Правительства (<http://193.232.232.88/>) – муниципальная и коммунальная информация.

Информация по Интернет

* Meckler iWORLD (<http://www.mecklerweb.com/>) – страница для участников Интернет, подготовленная в связи с возникновением WWW и характеризующая развитие Интернет и WWW;

* NetInSights (<http://www.netinsights.com>) – конференции по изучению путей развития Интернет;

* Open Market's Internet Index
(<http://www.openmarket.com/diversions/internet-index>) – собрания фактов по Интернет;

* The Internet Reference Guide (<http://204.120.185.2/netindex.html>) – системы обучения работе с Интернет.

В России также появились серверы, основным объектом которых выступает Интернет:

* CrazyWEB (<http://www.ru/crazyweb/>) – диалоговый журнал, посвященный российскому сектору Web;

* План-карта информационных ресурсов в России (<http://www.ras.ru/>).

Заключительная лекция

Идеи и методы информатики

За время становления информатики в ней выработан ряд специфических идей и методов, которые не были известны библиотеко- и библиографоведению или какой-либо другой смежной науке или научной дисциплине. Важнейшими из этих идей и методов являются следующие:

- координатное индексирование документов и других объектов соответствующими характеристиками – ключевыми словами и дескрипторами;

- использование в качестве дескрипторов классов условно эквивалентных ключевых слов;

- инвертированная организация поисковых массивов;

- использование ключевых слов, выбираемых из заглавий, рефератов или текстов документов, для их координатного индексирования;

- использование *тезауруса* как средства представления знаний, полезного для информационного поиска, автоматического перевода текстов и других информационных процессов;

- избирательное распространение информации о релевантных документах;

- интерактивный (диалоговый) поиск документов и информации;

- ранжирование документов, выдаваемых при поиске, по степени их релевантности;

- кластерный и другие методы автоматической классификации документов;

- использование библиографических ссылок для информационного поиска научных документов и решения наукометрических задач путем создания специальных указателей и баз данных;

– методы изучения явлений роста, рассеяния (С. Бредфорд) и старения публикаций, а также определения их информационной ценности.

На первый взгляд некоторые из перечисленных идей и методов могут показаться мелкими или не поднимающимися над уровнем здравого смысла. На самом деле, за внешней простотой этих идей и методов скрыт глубокий смысл, не всегда понимаемый в полной мере даже многими из тех специалистов, которые используют эти идеи и методы в своей практической деятельности. Даже для многих библиотекведов и информатиков до сих пор остается непонятной логическая сущность координатного индексирования. Понятие дескриптора тоже в большинстве случаев трактуется ими весьма упрощенно, что, несомненно, препятствует разработке эффективных информационно-поисковых языков дескрипторного типа¹.

Далеко не все специалисты по информатике понимают, что информационно-поисковый тезаурус – это не только двуязычный словарь, но и специализированный справочник, в котором приводится логическая информация, необходимая для адекватного перевода с естественного языка на информационно-поисковый. Это непонимание проявляется в том, что информационно-поисковыми тезаурусами называют любые упорядоченные множества терминов – таблицы десятичной классификации, рубрикаторы информационных изданий, словники предметных указателей. До сих пор не разработана научная методика построения и использования таких тезаурусов, хотя от их качества в значительной мере зависят результаты информационного поиска.

Построение теоретических основ любой науки и научной дисциплины невозможно без разработки системы ее основных понятий и специальной терминологии. В информатике это было сделано уже на начальных этапах ее формирования. Информатики сознавали необходимость изучения структуры, содержания и динамики потоков научной и другой семантической информации. Для этого были разработаны и апробированы соответствующие методы, которые могут быть использованы не только в сфере научной коммуникации, но и в других сферах общественной жизни. Было продолжено и углублено начатое еще в библиографоведении изучение рассеяния публикаций определенной

¹ Черный А. И. Заметки об информатике и дескрипторах // Азгальдов Э. Г., Высочанская О. А., Хаскина М. И. Дескрипторный словарь по информатике. – М.: ВИНТИ, 1991. – С. 3–13.

отрасли знаний, предмета или проблемы по периодическим изданиям, которое известно как «закон рассеяния Бредфорда».

Изучение специфики информационных потребностей ученых и специалистов, изыскание наиболее эффективных способов их удовлетворения привели к выработке понятий «релевантности» и «пиртинентности». Эти понятия легли в основу теории информационного поиска, которая несомненно является крупным вкладом информатики в современную науку. В рамках этой теории были сформулированы понятия «информационно-поискового языка», «поискового образа документа», «поискового предписания» и «критерия выдачи». Были также разработаны способы оценки эффективности информационного поиска, без чего было бы невозможно сравнение разных информационно-поисковых систем с целью их совершенствования. Особенно значительным достижением информатики является разработка метода координатного индексирования документов и других объектов поиска, а также дескрипторных информационно-поисковых языков и тезаурусов.

Идеи и методы информатики используются во всех сферах интеллектуальной деятельности – везде, где сообщения облакаются в форму документа и становятся объектом аналитико-синтетической переработки, долговременного хранения и достаточно частого поиска и использования. В качестве таких сообщений могут выступать не только научные документы, но и газетные статьи, фотоснимки, опубликованные тексты радиопередач, архивные материалы, банковские счета, бухгалтерские документы и т. п.

При создании автоматизированных информационных систем, которые охватывают документы, не относящиеся к сфере науки и техники, как правило, не требуется применения каких-либо иных методов и средств, чем те, которые разработаны информатикой для научных документов. Более того, для этого обычно оказываются достаточными наиболее простые методы. Сказанное в полной мере относится и к сфере материально-технического снабжения, когда объектами классификации (индексирования), хранения, поиска и выдачи являются не документы, а материальные объекты – промышленные изделия, запасные части и т. п.

Информатика все больше ориентируется на использование компьютеров, микрофотографии и средств массовой коммуникации. В последние десятилетия для поиска документальной и фактографической информации, а также для дистанционной передачи копий научных документов все шире используются каналы и средства телеграфной, те-

лефонной и телевизионной связи. В последние годы особенно интенсивно развиваются компьютерные сети. Ясно, что использование компьютеров и других новейших технических средств, а также их быстрое совершенствование оказывают сильное влияние на развитие определенных методов информатики.

Следует, однако, понимать, что компьютеры и другая современная информационная техника служат лишь инструментами для решения задач информатики. Их использование для решения задач аналитико-синтетической переработки информации базируется в основном, если не исключительно, на идеях и методах, разработанных в рамках информатики. Между тем иногда складывается неверное впечатление, что решающим фактором в «индустрии информации» являются применяемые в ней технические средства, а идеи и методы, на основе которых производится переработка информации, играют второстепенную роль. Хорошо известно, что компьютеры работают по определенным программам, разработанным на основе идей и представлений, лежащих за пределами программирования. Рассмотрение конкретных примеров решения информационных задач при помощи средств автоматизации (например, автоматизированных информационных систем в медицине) показывает, что все они основаны на идеях и методах информатики.

Особенно широкое распространение получил метод координатного индексирования ключевыми словами, который ныне используется в самых разнообразных областях. Изолированные ключевые, стержневые слова (*нем.* *Stichwort*), выбираемые из заглавий книг, использовались для составления каталогов немецких библиотек еще в XIX в. Сущность метода информатики – в установлении координатной связи между ключевыми словами, выбранными для индексирования документа, в их соподчинении, логическом умножении. С практической точки зрения этот метод позволил автоматизировать индексирование документов, а также их поиск по произвольным сочетаниям признаков (с использованием булевой и «пороговой» логик) – особенно в режиме диалога. Не в меньшей степени этот метод ценен и теоретически, поскольку сделал объектом соответствующих исследований семантическую и знаковую структуру текстов на естественных языках, без чего невозможна автоматическая переработка таких текстов в системах искусственного интеллекта. Нельзя забывать и о том, что этот метод был разработан и применялся еще до появления первых компьютеров.

Примером того, насколько важную роль в современных автоматизированных информационных системах играют идеи и методы ин-

форматики, может служить японский проект создания ЭВМ пятого поколения. Главным принципом проекта являлось создание сверхбыстродействующих компьютеров, ориентированных на использование «баз знаний», т. е. накопленных и особым образом организованных декларативных и процедурных знаний о соответствующих фрагментах внешнего мира. Но такая задача в информатике начала обсуждаться несколько десятилетий назад. Концептуальным аналогом «базы знаний» можно считать тезаурус как семантическую сетевую структуру. Тезаурус, используемый при поиске документов, является проблемно-ориентированным вариантом «базы знаний».

Поиски фундаментального закона

На рубеже 70-х и 80-х годов XX в. значительно возрос интерес к теоретическим аспектам информатики и, в особенности, к ее методологическим проблемам. Необходимость в решении этих проблем и построении теории информатики, адекватной ее практическому развитию и расширившейся сфере ее применения, была четко выражена на официальном заседании информатиков в Национальном научном фонде США в 1979 г., получившем название "Круглого стола". Этот же период ознаменован появлением крупных работ по теории информатики в Великобритании, принадлежавших известным ученым Б. Бруксу и Дж. Фаррадейну.

В четырех последовательных номерах английского "Journal of Information Science" (1980.— Vol. 2, N 3–6; 1981.— Vol. 3, N 1) была опубликована работа Б. Брукса "Основания информационной науки". Автор начинает с того, что, когда он посещал американские университеты, ему представляли коллег как преподавателей лингвистики, вычислительной техники, статистики для информатиков. Когда же он спрашивал, кто учит студентов самой информатике, ему отвечали, что информатика – смесь из этих дисциплин вместе с библиотечной технологией и исследовательской методикой. Но автор утверждает, что "информатика это дисциплина с собственной и притом уникальной областью исследований, проблемами и точкой зрения на человеческую деятельность, и теперь она должна создать собственные принципы и методы. Как смесь не связанных между собой элементов из произвольного набора отдельных дисциплин информатика не имела бы будущего".

Б. Брукс справедливо полагал, что информатика должна опираться на фундаментальные законы, согласующиеся с определенной

философской системой. Как и большинство ученых на Западе, он развивает свою теорию в рамках объективного идеализма и берет за основу критический рационализм современного английского философа К. Поппера, в частности, его учение о существовании трех миров: физического мира, мира субъективного знания и мира объективного знания ментального мира. Исследование Б. Брукса направлено на поиски количественной меры знания, которое он понимает как структуру понятий и отношений между ними. Информацию он рассматривает как небольшую часть этой структуры, а взаимосвязь между информацией и знанием выражает уравнением:

$$K [S] + \Delta I = K [S + \Delta S],$$

описывающим в общем виде преобразование структуры знания: $K [S]$, которая превращается под воздействием информации ΔI в $K [S + \Delta S]$.

Из этого уравнения в псевдоматематической форме автор делает следующие выводы:

- если информация и знания измеримы, то они должны измеряться в одних и тех же единицах,
- информация отлична от чувственно воспринимаемых данных, хотя и зависит от них,
- появление новой информации может приводить к радикальному изменению структуры знания не только путем добавления к ней новых элементов.

Б. Брукс переосмысливает полученные им ранее результаты в общую теорию и считает, что классические статистические методы игнорируют то обстоятельство, что при образовании больших классов потеря информации соответствует потере индивидуальности объединяемых в эти классы сущностей. Поэтому в общественных дисциплинах, к которым принадлежит информатика и в которых индивидуальность этих сущностей значительна, необходим частотно-ранговый анализ.

В дальнейших рассуждениях он приходит к формулированию логарифмического закона информации, который интерпретируется как закон равномерного распределения информации в логарифмической шкале. В качестве аналога этого закона он использует закон перспективы (видимые размеры объекта обратно пропорциональны расстоянию до него), преобразованный им следующим образом: видимое расстояние до объекта пропорционально логарифму физического расстояния до него.

Другую аналогию, иллюстрирующую логарифмический закон информации, Б. Брукс находит в наблюдении звездного неба. В те-

чение шестнадцати веков существовала Птолемеява модель космоса, в точности соответствовавшая абстрактному информационному миру, в котором информация распределена равномерно. (Человеку, находящемуся на Земле и наблюдающему небо, нелегко делать поправку на искажение расстояний за счет закона перспективы). Эта модель служит примером субъективной точки зрения на трехмерный мир, наблюдаемый в перспективе. Чтобы перейти к объективной картине космоса, понадобилось открыть постоянство скорости света, сформулировать понятие светового года и сделать его объективной мерой астрономических расстояний. Нечто подобное должно произойти и в информатике после нахождения меры семантической информации, что, по мнению автора, может быть сделано на основе теории информации К. Шеннона.

Иная попытка создания теории информатики была предпринята Дж. Фаррадейном в статьях "К истинной информационной науке", "Природа информации", "Знание, информация и информатика", опубликованных в 1976–1980 гг. в том же журнале (и его предшественнике "Information Scientist". –1976. – Vol. 10, N 3; 1979. – Vol. 1, N 1; 1980. – Vol. 2, N 2). Эти статьи не составляют законченного цикла, но обладают внутренним единством и посвящены тому же кругу вопросов, разрешаемых, впрочем, в ином аспекте. Основная задача данного исследования заключалась в том, чтобы операционным путем нащупать связь между информацией и ее содержанием, между мыслью и ее физической формой, что, по мнению автора, возможно лишь путем изучения реакции получателя (потребителя) информации. В соответствии с этой задачей он определяет информацию как физическую форму представления знания или отдельной мысли, как их суррогат, предназначенный для коммуникации. В этом определении отсутствует связь информации как с ее создателем (отправителем), так и с ее воздействием на получателя (потребителя).

Дж. Фаррадейн считал, что большинство определений информации как фрагмента знания содержит характеристики, свидетельствующие о непонимании природы мысли ("новизна", "приращение знания", "полезность", "интерпретация внешних стимулов", "основные закономерности материи и сознания", "повышение уровня знаний получателя", "устранение неопределенности", "значение для принятия решения", "структура или организация", "изменение структуры тезауруса получателя"). Альтернативная этим определениям концепция информации как системы, включающей людей, их деятельность и потребно-

сти, а также влияние передачи информации на принятие решений, социальное поведение и т. п., представляется автору тоже неприемлемой для эффективного анализа проблемы, поскольку не позволяет изучать каждый элемент системы изолированно.

Автор рассматривает разные варианты реакции потребителя на получение незапрашиваемой информации:

– если она ему уже известна, он может отвергнуть ее или же принять как подтверждение прежнего опыта, но не закладывать в долговременную память как поверхностную, тривиальную или не соответствующую его интересам,

– если же информация не была получена ранее и не отвергается по названным причинам, то она может быть добавлена к имеющимся знаниям, не изменив их, заполнить пробел в структуре знаний получателя, изменить существующую у него структуру знаний.

При получении информации, отвечающей на запрос, реакция потребителя обычно ограничивается двумя последними вариантами, т. е. заполнением пробела в знании или изменением его структуры.

Рассматривая информационную науку как экспериментальную, Дж. Фаррадейн не ждет больших результатов от философских размышлений и абстрактных математических моделей, если они не дают объяснений экспериментальным данным. Большинство же экспериментов в информатике, по его мнению, недостаточно корректно, так как опирается на произвольную интерпретацию данных. Поэтому он призывает обратить большее внимание на реакцию потребителя при получении первичной и вторичной информации. Что же касается эффективности информационно-поисковых систем и всей информационной деятельности, то основным инструментом и этапом достижения этой эффективности он считает индексирование. При этом центральная роль отводится индексатору даже и в автоматизированных системах, где эта роль лишь завуалирована, но может проявляться в тезаурусах, алгоритмах, стратегиях поиска, грамматических и лексических правилах. Хочется отметить, что и в этом исследовании подчеркивается настоятельная необходимость создания теоретической базы информатики, которая позволила бы этой новой дисциплине стать исключительно интересной в будущем.

Важное значение, на наш взгляд, имела также дискуссия, прошедшая в ГДР в 1978–1979 гг. на страницах журнала "Informatik" под девизом "К выяснению теоретических основ научной информации как предпосылки совершенствования информационной работы в области

науки и техники". В дискуссии приняли участие преподаватели и научные работники Берлинского университета им. В. Гумбольдта, Высшей технической школы Ильменау, Академии наук ГДР. Не будет преувеличением сказать, что все имевшиеся к тому времени в литературе мнения об информатике (информационной науке, науке об информации и документации), ее предмете, объектах и проблемах исследования были подвергнуты критическому осмыслению. Эти дискуссии в разных странах проводятся до настоящего времени.

Это хорошо, поскольку привлекает внимание к важной проблеме. Ведь в отличие от политических, научные дискуссии не завершаются принятием согласованного решения, а служат средством выяснения точек зрения и побудительным мотивом для дальнейших исследований.

Трудно возражать против того, что все возрастающее число исследований во многих точных, естественных и общественных науках, посвященных информационным аспектам этих наук, представляют для нас интерес и служат прояснению общих концепций информатики. Можно понять и стремление декларировать создание всеобщей науки об информации – проникновение в сущность этой философской категории является одной из актуальных общенаучных задач. В конечном счете, все научные проблемы взаимосвязаны и каждая имеет свой информационный аспект.

Однако в силу многих причин и особенностей организации современной науки как социального явления мы вынуждены дробить ее на специальные научные дисциплины. Вопрос о предмете и объектах исследования, о границах каждой и особенности такой формирующейся дисциплины, как информатика, это, прежде всего, вопрос о выделении того круга понятий и явлений, который необходим и достаточен для открытия фундаментальных объективных законов и закономерностей. Одной из устойчивых тенденций в развитии информатики является расширение понимания ее предмета. Эта тенденция нуждается в разумном контроле и ограничении, если мы не хотим, чтобы информатика из специальной научной дисциплины превратилась в комплексный раздел науки, объединяющий разнородные темы, связанные с изучением различных видов и аспектов информации. Это вынуждает нас обращать особое внимание на определение и предметную область информатики.

Определение и предметная область информатики

Использование термина "информатика" в значении информационной технологии совместилось с уже принятым у нас более полутора десятилетий обозначением им научной дисциплины, изучающей структуру и общие свойства научной информации и общие свойства и закономерности научной коммуникации. Это породило некоторые терминологические трудности. Существо и путям преодоления этих трудностей в нашей стране посвящена достаточно большая литература. Среди высказанных точек зрения особого внимания заслуживают мнения А. П. Ершова, А. А. Дородницына и В. С. Михалевича. Серьезную методологическую помощь в этом вопросе оказывают и философы. Среди них заслуживают быть особо отмеченными А. Д. Урсул и Э. П. Семенюк.

Зарубежные специалисты иногда посмеиваются над нашим, как им кажется, чрезмерным пристрастием к обсуждению чисто терминологических проблем. Действительно, как просто решаются эти затруднения в английском языке, где термин "информатика" употребляется редко, а вся совокупная область, обозначаемая этим словом по-русски, именуется "информационная наука и технология". Как будто бы о словах спорить не принято. Но это неверно. Термин это не просто слово, даже не более или менее удачное название понятия или дисциплины, за термином обычно (а в данном случае особенно) стоит концепция и предметная область.

Наибольшее понимание вызывает позиция А. П. Ершова, который считал, что информатика должна быть "фундаментальной естественной наукой, изучающей процессы передачи и обработки информации". В этой позиции смущает только категорическое отнесение информатики к естественным наукам, тем более, что довод самого А. П. Ершова "о единстве законов обработки информации в искусственных, биологических и общественных системах" скорее подтверждает мнение Э. П. Семенюка об общенаучном характере информатики.

Менее приемлемой представляется позиция А. А. Дородницына, который определял состав информатики как "три неразрывно и существенно связанные части – технические средства, программные средства и алгоритмические средства". Трудно представить себе фундаментальную научную дисциплину, в основе которой лежат технические средства. Нельзя согласиться и с тем, что информатика как научная

дисциплина должна объединить две различные, хотя и близкие дисциплины, в англоязычной практике называемые "информационная наука" и "компьютерная наука". Соответствующие им в русском языке "информатика" и "вычислительная техника" уже сложились как самостоятельные дисциплины и имеют определенные аналоги в мировой науке.

Не может не настораживать и неумеренное расширение предмета информатики путем включения в него всех философских, социологических, психологических и других обществоведческих проблем, связанных с процессами передачи и обработки информации. Такая опасность возникнет, если без всяких ограничений последовать мнению В. С. Михалевича. Он характеризует информатику как новую комплексную научную дисциплину, изучающую все аспекты разработки, проектирования, создания, "встраивания" в социальную среду машинных систем обработки данных, а также их воздействие на ту или иную область общественной практики. С такой трактовкой предмета информатики также нельзя согласиться. Дело в том, что наряду со становлением информатики закономерно развиваются информационные аспекты других наук от математики и биологии до социологии и философии. В некоторых науках эти аспекты оформляются в самостоятельные разделы. Объединение всех этих разделов в информатике превратило бы ее в "науку всех наук", что и неразумно и невозможно.

Любопытно, что приведенные мнения относительно информатики как научной дисциплины высказываются так, будто она возникла внезапно и не имеет никаких исторических корней. Более того, создается впечатление, что у этой науки нет аналогов за рубежом. Говорится о том, чем должна быть или стать информатика, но совершенно игнорируется ее современное состояние. Мы знаем, что научные дисциплины не возникают на пустом месте, а наука в целом интернациональна. Говоря о развитии информатики, мы уже знакомимся с мнением ряда специалистов США о ее предметной области. Прокомментируем их позицию.

Хотелось бы обратить внимание на три момента. Во-первых, информационная наука в США на протяжении последних десятилетий проходила те же этапы, что и информатика в СССР, включая и нынешний этап ее превращения в фундаментальную науку о семантической информации. Во-вторых, хотя в США нет терминологических трудностей с "информатикой", но как и в СССР имеется склонность к неправомерному отождествлению "информационной" и "вычислительной" наук. Примечательно, что русский термин "информация" шире, чем

английский, и включает не только сообщения, переданные по каналам связи, но и управляющие сигналы в автоматических системах, не имеющие семантики (смысла). В-третьих, и там и здесь проявляется особое внимание к семиотическим (знаковым) и семантическим (смысловым) аспектам информатики и осознается ее междисциплинарное значение. Разумеется, имеются и существенные различия. Путь, который американцы проходили более или менее эволюционно, мы пытаемся преодолеть скачком, сразу переходя к информационной технологии, основанной на компьютерах новых поколений.

Поэтому для нас представляют особый интерес те теоретические дискуссии, в которых рождалась современная информационная наука на Западе. Из многих таких обсуждений серьезностью подхода и достигнутыми результатами выделяются два. Одно из них проходило в 1972 г. в Северн Спрингс (шт. Пенсильвания, США) под названием "Информационная наука: поиски своего лица". Несмотря на участие очень многих известных специалистов, обсуждение не привело к ясной формулировке предмета и границ информационной науки. Было высказано много полезных мыслей о природе социальной информации, необходимости переходить от изучения процессов обработки документов к изучению структуры знания, о связи информационной науки с психологией творчества и искусственным интеллектом.

М. Кочен, известный социолог науки, заключая доклад, вынужден был признать что, "информационная наука в прямом смысле началась с документации в химии, с одной стороны, и с информационного поиска, с другой. На всем протяжении ее существования ей недоставало научной инфраструктуры". Закрывая обсуждение, Дж. Ликлайдер, специалист в области информационных систем, призвал сконцентрировать усилия на двух направлениях: поисках фундаментальных оснований науки и включении ее в разработку проблем информационных сетей.

Второй из названных исследовательских форумов проходил в 1977 г. в Копенгагене под названием "Теория и применение информационных исследований". На нем среди многих других проблем обсуждались и определения информационной науки. Наиболее известное из них дано Международной организацией по стандартизации: "Информационная наука это область знания, изучающая структуру и функции информации и информационных систем, которые включают создание, сбор, анализ, представление, хранение, поиск, распространение и использование информации". Британский стандарт, принятый в

1976 г., указывая на цель информационной науки, состоящую в изучении теории и практики сбора, оценки, организации хранения и распространения информации, определяет ее как "науку о закономерностях и свойствах информации и ее обработки, преимущественно на ЭВМ для удовлетворения информационных потребностей людей".

Б. Веккери подчеркнул, что современная проблематика информационной науки сложилась постепенно, главным образом, из нужд обучения информационных работников: "Термин информационная наука употребляется в разных контекстах. Сначала учеными в этой области назывались специалисты, которые вели информационное обслуживание других ученых и инженеров. По мере развития этой профессии и курсов по подготовке ее специалистов совокупность знаний, необходимых для их обучения, стала называться информационной наукой. Возрастание роли системного анализа, проектирования и оценки систем информационного обслуживания, в также все более интенсивное использование таких технических средств, как микрофотография, вычислительные машины и телекоммуникация привели к тому, что многие стали считать, что все эти объекты также изучаются информационной наукой". Однако сам Б. Веккери считает, что "информационная наука изучает информационную коммуникацию в обществе", а его американский коллега А. Дебонс подчеркивает необходимость развития теоретических основ этой дисциплины.

В заключительном отчете форума было признано, что в информационной науке целесообразно различать фундаментальные аспекты (изучение информации и ее преобразований как явления) и прикладные аспекты (изучение методов преобразования данных в информацию) и использовать результаты исследования других наук: философии (общие методы подхода к проблеме), лингвистики, социо- и психолингвистики (представление документов и обслуживание потребителей), психологии и социальной психологии (создание и восприятие информации), социологии (организация информационных потоков), компьютерной науки (проектирование информационно-поисковых систем), статистики (оценка эффективности систем), а также системного анализа, исследования операций и математики как общенаучных методов.

Чтобы закончить с определениями информатики, приведем еще одно, принадлежащее лондонскому Институту информатиков, ответственному за поддержание высокого уровня профессиональных знаний своих членов. "Информационная наука занимается теорией, методикой и практикой предоставления информации преимущественно, но не ис-

ключительно для поддержания профессиональной и академической деятельности. С этой целью она изучает информацию от ее создания до ее использования, а также ее передачу в различных формах и по разным каналам". В этом определении, относящемся к 1977 г., признается, что информатика изучает не только теорию, но также методику и практику, а изучаемая информация, хотя и называется специальной, но не ограничивается определенными видами научной, технической, экономической или какой-либо другой. К сожалению, за прошедшую со времени этих дискуссий четверть века поиски оснований информатики существенно не продвинулись.

В этом месте предвижу недоуменный вопрос: "А где же то определение, которое нужно выучить, чтобы потом отвечать на экзамене или зачете?" Оно не приводится, потому что цель данных лекций заключается в том, чтобы пробудить интерес к обсуждаемой проблеме и стимулировать работу мысли. А верное определение имеется, и не одно. Их может быть много, в зависимости от аспекта, в котором рассматривается информатика. Да и не одним только определением задается предметная область. Это, как мы уже установили в начале данного курса лекций, может быть сделано и перечнем ее выдающихся представителей, и анализом тематики диссертаций, кластеров социотирования, и многими другими методами, дающими относительно объективные результаты.

Существует разница в организации знаний, необходимых для создания технологии или другой практической цели, и знаний, составляющих отрасль науки или научную дисциплину. Именно поэтому теоретические и прикладные аспекты одной и той же науки могут сильно различаться не только последовательностью, но и составом включаемых в нее знаний. Не случайно при обсуждении предметной области информатики выделяются ее особенности как фундаментальной науки, прикладной области знаний и учебной дисциплины.

Перспективы информатики

Немаловажное значение для перспектив развития информатики имеет ее структура, в значительной мере определяющая направление проводящихся в ней научных исследований. Эта структура в разных модификациях (в соответствии со спецификой того или иного вида деятельности) находит отражение в рубриках реферативных и других информационных изданий, монографиях, излагающих основы этой дисциплины, диссертациях, программах и планах научных исследова-

ний и разработок, учебниках и т. п. Наиболее чуткими к объективным изменениям структуры научной дисциплины являются рубрикаторы, поскольку они испытывают одновременное (хотя и не одинаковое по силе) воздействие как потоков литературы, так и интересов читателей. Несмотря на это, рубрикатор реферативного журнала ВИНТИ "Информатика" сохраняет устойчивую структуру, включающую семь разделов:

общие вопросы теории и методологии,
организация информационной деятельности,
документальные источники информации,
аналитико-синтетическая переработка,
информационный поиск,
информационное обслуживание,
технические средства.

Нетрудно заметить, что эта структура (если отбросить обязательные разделы теории, организации и технических средств) отражает последовательность этапов информационной деятельности:

отбор – обработка – хранение – поиск – распространение.

Эта структура характерна и для большинства названных выше источников. Прослеживается она и в нашем курсе лекций. Не нужно забывать и о том, что предмет и структура информатики, круг разрабатываемых ею проблем имеют не только теоретическое, но и практическое значение. Подготовка кадров, повышение их квалификации, формирование профессионального мышления, информационное обеспечение, создание коммуникационных традиций (как по формальным, так и по неформальным каналам) – вот далеко не полный перечень жизненно важных вопросов, решение которых зависит от того, какой круг вопросов будет исследовать информатика. Но, учитывая, что научная дисциплина это не только совокупность знаний, но и институция, сообщество ученых, объединенных единым кругом изучаемых проблем и подходов к их изучению, инфраструктура дисциплины оказывает влияние на ее развитие.

Как в этих условиях будет развиваться информатика? Если не выходить за очерченные нами рамки, то потребуется значительно интенсифицировать исследования свойств и структуры семантической информации, подходов к определению ее сложности, количества, ценности, полезности, кумулятивности, старения, рассеяния и т. п. Необходимы дальнейшие исследования информационных систем, разработка их математических и других формальных моделей. Особую актуальность приобретает построение интеллектуальных информационных

систем, позволяющих прогнозировать исследуемые свойства веществ, процессов, явлений на основе неполной информации. Информационный поиск смыкается здесь с автоматизацией исследований и проектирования, что требует глубокой логической и лингвистической проработки.

В области технологии информационной деятельности возникает ряд важных проблем взаимодействия человека с развитыми техническими системами: унификация представления научно-технических данных, методы кумуляции и концентрации информационных ресурсов, методы структурирования информации в условиях ее машинного накопления и обработки, рациональное представление нетекстовой информации, эффективный ввод речевой и графической информации. Малоизученной, но крайне важной областью информатики являются ее экономические, социологические, психологические и правовые аспекты. Вся сфера информационной коммуникации недостаточно изучена с точки зрения ее соответствия экономическим структурам производства и процессам принятия решений. Перспективными представляются исследования влияния новых информационных технологий на внедрение достижений в народное хозяйство, в сферу образования, культуры, массовой коммуникации.

Если же пойти еще дальше, за пределы того круга проблем, которые легко укладываются в рамки обсуждаемой нами научной дисциплины информатики, то мы выйдем в необозримое пространство проблем, обозначаемых терминами "искусственный интеллект", "информационные технологии", "теория программирования", "теория формальных языков" и т. п. Споры нет, информатика уже сейчас наталкивается на многие вопросы, разрабатываемые в этих проблемных комплексах. К ним относятся "понимание" текста на естественных языках, логико-смысловые методы его обработки, представление знаний в базах данных, программирование задач информационного обеспечения и многие другие вопросы подобного типа. По всей вероятности, они займут в информатике соответствующее их значению место.

Но это не потребует расширения тех границ, которыми очерчены предмет и объекты изучения информатики. Все эти вопросы не выходят за рамки изучения структуры и свойств семантической информации и ее воздействия на многообразные сферы человеческой деятельности.

Ясно, что ближайшей перспективой развития информатики является развертывание широкого фронта теоретических исследований.

Это не значит, что прикладные разработки, направленные на совершенствование нынешних информационных систем и методов информационного обеспечения должны быть отодвинуты на второй план. Они также будут развиваться и расширяться, но не они, а теоретические исследования определяют будущее информатики как фундаментальной науки.

Словарь терминов

Значение терминов объясняется в словаре лишь в том аспекте, в котором они употреблены в данном курсе лекций.

Адаптер – набор электронных цепей для связи процессора с периферийными устройствами.

База данных – совокупность взаимосвязанных данных, хранящихся во внешней памяти ЭВМ и используемых различными программами.

База знаний – формализованная совокупность сведений о некоторой предметной области, содержащая данные о свойствах объектов, закономерностях процессов и явлений и правила использования этих данных в задаваемых ситуациях для принятия решений.

Видеотека – система, использующая клавиатуру и телевизор для доступа по телефону к информационным службам.

Графопостроитель – см. плоттер.

Дисплей – устройство визуального отображения информации, в качестве которого обычно используется телевизионная трубка.

Гипертекст – (1) форма организации текста, при которой его смысловые единицы (фразы, абзацы, разделы) представлены не в линейной последовательности, а как система явно указанных возможных переходов, связей между ними; (2) формат представления информации, позволяющий логически связывать отдельные слова текста, графические и аудио-фрагменты.

Драйвер – управляющая программа в операционной системе мини- и микро-ЭВМ, предназначенная для организации обмена информацией между оперативной памятью ЭВМ и конкретным внешним устройством.

Индексирование – процесс выражения главного предмета или темы документа на информационно-поисковом языке.

Интеллектуальная система – человеко-машинная система, в которой ЭВМ имитирует рассуждения квалифицированного эксперта, усиливая интеллектуальные возможности пользователя.

Интерфейс – технические и программные средства, осуществляющие сопряжения устройств или их узлов.

Информатика – научная дисциплина, изучающая структуру и основные свойства семантической информации и служащую теоретической основой информационной технологии.

Информации источник (в информатике) – синоним информационного документа, т. е. логически завершенного фрагмента информации, имеющего указание на время и место создания и закрепленного на материальном носителе.

Информационный поиск – совокупность логических операций, в результате которых в ответ на запрос выдается необходимая информация, документ, в котором она может содержаться, или адрес этого документа, – потребитель – лицо, являющееся абонентом (возможно, потенциальным) какой-либо информационной системы (технической или социальной).

Информационная среда – совокупность источников информации, методов и средств доступа к информации.

Информационная сфера – совокупность общественных институтов и видов деятельности, обеспечивающих создание, передачу, хранение, поиск и использование

социальной информации (журналистика, редакционно-издательское, библиотечное и архивное дело, библиографическая и информационная деятельность).

Информационная технология – методы и средства создания, представления, передачи, хранения и получения информации.

Информационно-поисковая система – совокупность методов и средств, обеспечивающих осуществление информационного поиска.

Информация:

* **массовая** – сообщения, передаваемые по каналам массовой коммуникации;

* **научная** – сведения, получаемые в процессе познания (научных исследований, производственной и общественной деятельности), соответствующие современному уровню знаний и проверяемые практикой;

* **семантическая** – имеющая смысл (содержание) и выраженная на языке;

* **социальная** – циркулирующая в человеческом обществе (бытовая, справочная, медицинская, политическая, экономическая, техническая, научная, эстетическая, спортивная и т.п.).

Искусственный интеллект – (1) система, имитирующая решение человеком сложных задач, (2) направление научных исследований, сопровождающих и обуславливающих создание таких систем.

Кластер (*англ.* гроздь, скопление) – подмножество близких между собой объектов, образуемое в процессе (обычно автоматическом) разбиения (классификации).

Коммуникация интеллектуальная (информационная) – передача информации в процессе человеческой деятельности.

Коммуникация массовая – распространение информации через печать, радио, телевидение, кино, звуко- и видеозапись.

Коммуникация научная – процессы представления, передачи и получения научной информации.

Компактный оптический диск (CD ROM) – компакт-диск постоянного запоминающего устройства компьютера, на который информация записана во время изготовления и не может быть изменена (аналогичен стандарту акустического компакт-диска).

Логарифмическая шкала – совокупность делений и цифр, обладающая тем свойством, что расстояния между делениями пропорциональны разности логарифмов чисел, соответствующих этим делениям.

Межбиблиотечный абонемент (МБА) – система взаимного использования фондов библиотек страны.

МикроЭВМ – малогабаритный компьютер с большой производительностью, емкостью оперативной памяти, разрядностью машинного слова, внешней памятью и низкой стоимостью.

Монитор – устройство, применяемое для контроля и управления, включающее обычно видеотерминал (дисплей).

Плоттер (принтер-плоттер, графопостроитель) – печатающее устройство компьютера, предназначенное для растровых графических изображений.

Принтер – печатающее устройство компьютера.

Процессор – аппаратное или программное средство для обработки информации.

Рубрикатор – схема классификации материала в информационном издании (обычно в реферативном журнале).

Семiotика – наука об общих свойствах знаковых систем (языков естественных и искусственных – информационных, алгоритмических, программирования, международного общения, науки, а также систем сигнализации), включает три раздела: синтактику, семантику и прагматику.

Система управления базами данных (СУБД) – совокупность программ для создания, ведения и использования баз данных.

Сканнер – устройство оптического (бесклавиатурного) ввода в ЭВМ текста или изображения.

Тезаурус (греч. клад, сокровищница) – в широком смысле – знания, накопленные человеком, в более узком – справочник для поиска слов по их смыслу.

Телетекс – средство передачи текстовой информации по коммутируемым сетям общественного пользования (вскоре заменит телекс).

Телетекст – система широковещательной передачи текстов по резервным каналам на адаптированный бытовой телевизор.

Терминал – оконечное устройство, которое служит для оперативного ввода/вывода данных в/из ЭВМ.

Цитирование (англ. упоминание) – библиографическая ссылка на источник информации (не путать с обычным смыслом этого слова, означающего дословное воспроизведение фрагмента текста).

Экспертная система – искусственная система, выполняющая функции эксперта в конкретной предметной области.

Электронная почта – система, в которой ЭВМ берет на себя все функции по хранению и пересылке сообщений между пользователями вычислительных систем.

ASCII (American Standart Code for Information Interchange) – Схема кодировки, в которой знаки представляются наборами из семи или восьми битов; печатный (printable) ASCII — часть полного набора ASCII.

Boolean searching – Метод поиска информации, в котором поисковый запрос состоит из последовательности поисковых терминов, объединенных логическими операторами, такими как "and", "or", "not".

Browser – Пользовательский интерфейс общего назначения, используемый в веб и в других онлайн-информационных службах. Другое название — веб-браузер.

Cache – Временное хранилище, которое используется для копий недавно запрошенной информации или данных, которые, как ожидается, будут часто использоваться.

Database producers – (центры подготовки баз данных) – информационные службы, специализирующиеся на сборе информации, ее обработке и подготовке баз данных.

Database vendors – (центры обработки баз данных) – информационные службы, специализирующиеся на предоставлении баз данных для работы в диалоговом режиме.

DOI (Digital Object Identifier) – Идентификатор, используемый издательствами для обозначения материалов, опубликованных в электронной форме; частная форма

handle.

DTD (Document Type Definition) – Спецификация разметки классов документов, определенная в рамках SGML.

Dublin Core – Упрощенный набор элементов метаданных, используемый в электронных библиотеках, главным образом для описания цифровых объектов и управления коллекциями, а также для обмена метаданными.

Electronic journal – Онлайн-публикация, организованная по образцу традиционного печатного журнала; либо онлайн-версия печатного журнала, либо только электронная версия.

Emulation – Создание на современном компьютере программными методами виртуальной модели устаревшего компьютера, на котором "выполняется" программное обеспечение таких компьютеров.

FTP (File Transfer Protocol) – протокол сетевой передачи файлов.

Full text searching – Метод поиска в текстовых материалах, в котором поисковые термины сопоставляются с содержанием всего текста.

Gopher – разновидность FTP, предоставляющий пользователю более понятное меню для работы, которым можно управлять "мышью". Само слово gopher не имеет технического смысла и означает шуточное прозвище "суслик" для жителей штата Миннесота, в университете которого этот протокол был разработан.

Home page (домашняя страница) – Web страница, созданная для представления организации или частного лица в Интернет.

HTKP (Hypertext Transfer Protocol) Базовый протокол веб, используется для коммуникации между браузерами и веб-сайтами.

HTML (Hypertext Markup Language) Простой язык разметки и форматирования текста, с гиперссылками между объектами, используемый в WWW.

HTTP (Hypertext Transport Protocol) – протокол для сетевого доступа к информации в гипертекстовой форме.

Information broker (информационный брокер) – профессиональный информационный работник или фирма, оказывающая за плату услуги по поиску информации для заказчиков, а также оформляющая и интерпретирующая результаты информационного поиска.

Internet (Интернет) – аббревиатура слов INTERconnected NETworks, что означает ВЗАИМОсвязанные Сети.

ISDN (Integrated Services Data Network) – цифровая сеть передачи информации, которую можно использовать одновременно для телефонной и телеграфной связи, передачи данных, распространения видео программ и т.п.

Lexis – юридическая информационная служба, первой предоставившая полную текстовую информацию он-лайн.

Markup language – язык разметки — система команд, включаемых в документ и описывающих его структуру или формат.

Memex – концепция онлайн-библиотеки, предложенная Ванневаром Бушем в 1945 г.

Metadata – данные о других данных (метаданные), обычно подразделяемые на описательные метаданные (библиографическая информация), структурные метаданные (форматы и структура) и административные метаданные (для управления информацией).

Morphology – грамматические и иные варианты слов, произведенных от одного корня или основы.

On-line – (диалоговый, онлайн) – режим взаимодействия пользователя с компьютером, при котором происходит непосредственный обмен сообщениями (используется как составная часть многих понятий).

Optical character recognition – автоматическое преобразование текста из графического представления в кодированный текст.

Scanning – метод преобразования формы представления; при котором физический объект, например, печатная страница, превращается в набор точек (графическое цифровое изображение).

Speech recognition – автоматическое преобразование произносимых слов в компьютерный код (распознавание речи).

Stop word – слово, которое настолько распространено, что обычно игнорируется при поиске информации. Набор таких слов называют стоп-списком.

Tag – команда, специальная последовательность знаков в размеченном тексте, указывающая на структуру или формат.

Thesaurus – лингвистический инструмент, который связывает слова между собой по их значениям.

Transliteration – последовательная замена знаков одного алфавита (набора знаков) знаками другого алфавита (набора).

TREC (Text Retrieval Conferences) – ежегодные конференции по методам обработки текстов, представленных на естественных языках. (Тестирование систем на стандартных коллекциях текстов и задач).

URL (Uniform Resource Locator) – ссылка на ресурс в Интернете, определяющая протокол, компьютер и файл на этом компьютере, а также параметры. Полный URL указывает расположение как доменное имя или IP-адрес; относительный URL указывает положение относительно текущего файла.

Virtual (виртуальный) – электронная форма представлений объектов с использованием возможностей технологии мультимедиа.

World Wide Web (WWW, веб) – Связанный гиперссылками набор информационных источников в Интернете и технологии, которые при этом используются, включая HTML, HTTP, URL и MIME. Совокупность серверов Интернет, предлагающих информацию в гипертекстовой и мультимедийной форме. (WWW–Всемирная паутина).