

ВСЕРОССИЙСКИЙ ИНСТИТУТ НАУЧНОЙ И ТЕХНИЧЕСКОЙ ИНФОРМАЦИИ
(ВИНИТИ)

НАУЧНО • ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ

Серия 1. ОРГАНИЗАЦИЯ И МЕТОДИКА
ИНФОРМАЦИОННОЙ РАБОТЫ

ЕЖЕМЕСЯЧНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ СБОРНИК

Издается с 1961 г.

№ 10

Москва 2000

ОРГАНИЗАЦИЯ ИНФОРМАЦИОННОЙ РАБОТЫ

УДК 002.56:004.7

В. А. Глухов, О. Л. Голицына, Н. В. Максимов

Электронные библиотеки.

Организация, технология и средства доступа

Рассматривается программа "Российские электронные библиотеки", в рамках которой ведутся работы по общесистемным вопросам создания и функционирования электронных библиотек, развития инфраструктуры, разработки инструментальных средств, а также создания конкретных электронных библиотек по областям науки, культуры и образования.

С 1998 г. Российский фонд фундаментальных исследований (РФФИ) совместно с Российским фондом технологического развития при Министерстве науки и технологий приступил к реализации межведомственной программы "Российские электронные библиотеки" [1], в рамках которой ведутся работы по общесистемным вопросам создания и функционирования электронных библиотек (ЭБ), развитию инфраструктуры, разработке инструментальных средств, а также по созданию электронных коллекций по отдельным отраслям науки, культуры и образования. В настоящее время успешно функционируют электронные библиотеки РГБ, ИНИОН РАН, ГПНТБ, ВНИИЦентра, Объединенного института ядерных исследований,

Российского государственного архива кинофотодокументов и другие.

Хотя понятие "электронная библиотека" уже прочно вошло в научный обиход, определение этого термина еще не вполне устоялось. С представленной в настоящей статье точки зрения, наиболее подходящим можно считать следующее: электронная библиотека — это информационно-поисковая среда, обеспечивающая пользователю эффективный доступ к комплексу взаимосвязанных разнородных массивов информации. Помимо справочно-поискового сервиса ЭБ обычно включает службу электронной доставки [2], обеспечивающую предоставление копий документов, по каким-либо причинам отсутствующих в ее полнотекстовом "фон-

де". Важнейшая роль библиотеки как систематизированного информационного пространства, обеспечивающего получение и воспроизводство знаний, затрагивается только косвенно при рассмотрении технологий информационного поиска. Таким образом, настоящая статья отражает скорее информационно-функциональную точку зрения, включающую:

анализ способов представления информации и взаимосвязей компонентов ЭБ;

обобщенную структуру информационно-технологического комплекса, обеспечивающего накопление и управляемый доступ к информационным компонентам ЭБ;

анализ основ и особенностей технологии документального поиска как неотъемлемой части когнитивного процесса, реализация информационного обеспечения которого и является основной задачей ЭБ.

Функциональные особенности, обсуждаемые далее, приводятся в контексте архитектуры и возможностей электронной библиотеки, создаваемой в ИНИОН РАН на основе программного обеспечения АИС ИРБИС, разработанной авторским коллективом в составе Е. Н. Васиной, О. Л. Голицыной, Н. В. Максимова, П. И. Резниченко.

ЭЛЕКТРОННАЯ БИБЛИОТЕКА КАК ИНФОРМАЦИОННЫЙ РЕСУРС

Основные задачи электронной библиотеки — интеграция информационных ресурсов и обеспечение эффективной навигации в них [3]. Причем, под интеграцией понимается объединение электронных документов и справочников различного вида с целью их широкого использования для предоставления пользователю возможностей манипулирования данными предпочтительно в функционально однородной среде. Само же объединение ресурсов не обязательно должно быть физическим — оно может быть виртуальным, но восприниматься пользователем как единое информационное пространство.

Если рассматривать ныне создаваемые электронные библиотеки с точки зрения состава и характера информационных ресурсов, то можно заметить, что в качестве компонентов здесь выступают электронные каталоги, библиографические и реферативные базы данных, полнотекстовые массивы (электронные журналы, фактографические базы данных, хранилища электронных копий статей и других публикаций), справочно-нормативные файлы (рубрикаторы, тезаурусы, авторские, предметные, географические и другие метаинформационные продукты). Записи этих массивов могут быть взаимно связаны между собой ссылками, указателями хранения или условиями отыскания. И даже если такие ссылки в явном виде не существуют, то уже по своей сути эти компоненты всегда были и будут связаны, по крайней мере, на концептуальном уровне. Например, записи электронных каталогов содержат указания местоположения книг, а справочно-нормативные файлы традиционно используются в качестве "точек входа" в библиографические и реферативные базы данных. С появлением полнотекстовых баз данных справочно-поисковый аппарат и массив собственно информации технологически становятся единым целым, и на первый план выходит задача организации таких взаимосвязей, чтобы переход по ссылке внутри

электронной библиотеки воспринимался пользователем как простейший одномоментный процесс, подобно перевороту страницы книги.

Таким образом, в общем случае информационные компоненты электронной библиотеки образуют трехуровневую структуру, включающую:

- компоненты метаинформационного уровня, обеспечивающие "узнаваемость" коллекции в целом, идентификацию как ее наполнения, так и формы представления. Сюда относится, например, нормативная информация о коллекциях объектов (рубрикаторы, тезаурусы, авторские, предметные и другие указатели), а также каталоги и описания структуры и наполнения доступных ассоциированных ресурсов;
- компоненты справочно-поискового описания объектов, обеспечивающих семантическую идентификацию — "узнаваемость" отдельного документа за счет унификации и нормализации представления как элементов данных, так и содержания. Обычно это библиографическая запись или какая-либо другая вторичная информация, являющаяся поисковым образом документа;
- компоненты уровня объектов первичной информации: текст, графическое изображение, аудио, видео и т. д., т. е. собственно информативные документы.

На рис. 1 приведена обобщенная типология взаимосвязей компонент различного уровня, где в скобках даны примеры, характерные для автоматизированных или традиционных информационных систем.



Рис. 1. Типология взаимосвязей (ссылок) между элементами данных информационных компонент различного уровня

Рассматривая эту схему как технологию поиска, можно видеть, что ссылки вполне “узнаваемы” и представляют собой традиционные правила и приемы отыскания информации в условиях “бумажной” библиотеки, где поиск начинается с классификационной схемы, указателя или каталога, и далее через библиографические описания приводит к первоисточнику, по которому, используя теперь уже пристатейные ссылки и указатели, поиск может быть снова продолжен с метainформационного уровня.

С точки зрения характера размещения информационных ресурсов (т. е. их локализации в вычислительной среде, что в итоге и определяет технологию и условия использования) можно выделить два класса электронных библиотек:

локализованные, когда метainформационная, поисковая и информационная компоненты объединены в одной программной и общей вычислительной среде;

распределенные, когда компоненты находятся, возможно, в различном сочетании, на взаимно удаленных вычислительных системах, взаимодействующих на уровне сетевого протокола и функционирующих под управлением разных программных систем.

Здесь, в свою очередь, можно выделить два типичных варианта. Первый — ЭБ, использующие *ассоциированные* ресурсы, когда локализованная метainформационная, а также, возможно, поисковая компоненты, а информационная компонента является распределенной (удаленной и независимо изменяющейся). Второй — ЭБ с *произвольно распределенными* компонентами, когда о ресурсе известны только его Интернет-адрес и, возможно, параметры для генерации поискового запроса.

Как же сейчас реализуется связь компонентов электронных библиотек? Существующая практика, основанная на принципах гипертекста, предполагает, что сами компоненты и связи между ними, как правило, статичны или, по крайней мере, мало подвержены изменениям и поэтому могут быть жестко и надолго определены. На деле это не совсем так. Вернее, это совсем не так. Ничего более динамичного, чем компоненты постоянно развивающейся электронной коллекции, не существует. Поэтому, даже когда речь идет о взаимосвязи на уровне локальных информационных ресурсов, нельзя раз и навсегда установить такие связи.

С широким внедрением телекоммуникационных сетей и некоторой стандартизации представления данных в Интернет задача взаимосвязи становится еще более сложной. Ее решение путем создания статичных связей практически невозможно, даже если бы все компоненты имели свои уникальные идентификаторы и незыблемое место в информационном пространстве (чего зачастую невозможно добиться даже для локальных массивов). Таким образом, любое изменение местоположения информационного объекта влечет за собой возникновение “ложных” связей в распределенных электронных библиотеках. Поэтому на смену статичным связям приходят так называемые “активные”, динамические, связи. Основной их особенностью является то, что они генерируются программно, по предопределенным алгоритмам во время обращения к объекту, позволяющим связать информационные объек-

ты при условии достаточной определенности (специфицированности) элементов.

Связи внутри ЭБ могут быть построены на таких идентификаторах, как давно применяемые ISBN и ISSN или недавно возникшие DOI. В тех случаях, когда такие идентификаторы отсутствуют (а таких случаев большинство), одним из решений может быть генерация динамических связей. В качестве основы для построения идентификаторов здесь могут выступать либо уникальные элементы записи, либо их свертки.

Такого рода связывание можно показать на примере электронной библиотеки Института научной информации по общественным наукам РАН, создаваемой при финансовой поддержке Российского фонда фундаментальных исследований. Электронная библиотека ИНИОН включает сейчас массив библиографии (свыше 1.8 млн записей) и хранилище оцифрованных копий источников (около 3 тысяч документов). Массив библиографии разделен по тематическому признаку на несколько проблемно-ориентированных баз данных. Электронные копии хранятся в специальных разделах сервера. При этом библиографические описания отдельных книг и статей из журналов могут быть включены сразу в несколько баз. Связь библиографического описания и электронной копии здесь определена на уровне структуры идентификатора копии, который генерируется из элементов библиографического описания источника — шифра хранения, года издания и некоторых других элементов библиографического описания [4]. Когда пользователь обращается к той или иной библиографической записи, система поиска генерирует идентификатор предполагаемой копии и, в случае ее наличия в хранилище, делает ее доступной пользователю, визуализируя сгенерированную гиперссылку. Если копия по каким-либо причинам удаляется из электронного хранилища, то, соответственно, ссылка во всех библиографических базах данных генерироваться не будет.

Кроме того, в качестве идентификаторов, используемых для установки активных связей, могут служить части компонентов ЭБ, организованные, например, в виде поисковых индексов. Таким образом можно связывать, скажем, элементы записей справочно-нормативных файлов и библиографических массивов. А для установления ссылок от пристатейной библиографии в полнотекстовых базах данных к записям библиографических баз данных, и далее — к оцифрованным копиям, равно как и для связывания элементов самих полнотекстовых баз данных, могут использоваться свертки библиографических описаний источников. Конечно, это возможно лишь при достаточно строгой структуризации данных и применении алгоритмов свертки, допускающих минимальный процент дублирования. Установленные подобным образом связи допускают возможность ошибочного определения объекта, однако, это решение кажется приемлемым, так как позволяет не отслеживать частые изменения компонентов электронных библиотек.

Еще к одному типу активных связей относятся так называемые внешние связи, направленные к компонентам удаленных электронных коллекций, поддерживаемых разными производителями.

Рассмотрим несколько возможных реализаций связей электронных библиотек с внешними информационными ресурсами. Если такой ресурс пред-

ставляет собой полнотекстовую базу данных с поисковым аппаратом, то связеобразующий механизм может функционировать на уровне обмена запросами. При этом запрос от электронной библиотеки к внешнему (ассоциированному) ресурсу должен формироваться программно, базируясь на данных исходной записи. Для примера рассмотрим реализованный в электронной библиотеке ИНИОН механизм связывания записей библиографических баз данных с электронными журналами издательства Elsevier из Научной электронной библиотеки (НЭБ) РФФИ [5–6].

При формировании результатов поиска в базах ИНИОН для записей, которые соответствуют статьям из журналов издательства Elsevier (это определяется по справочнику соответствующего *ассоциированного* ресурса), система строит URL-запрос к Научной электронной библиотеке, размещаемый в соответствующей библиографической записи. При инициировании ссылки полный текст статьи, в случае его обнаружения на сервере НЭБ и после проверки прав доступа, выдается пользователю. Таким образом, создаются активные гипертекстовые ссылки, которые распространяются в данном случае только на определенный набор ресурсов.

Однако когда речь идет о распределенных ресурсах, каковыми являются Интернет-страницы, содержащие полные тексты отдельных источников, ситуация становится более сложной. Во-первых, выявление описываемых в библиографических базах данных и опубликованных в Интернет документов — электронных аналогов печатной продукции, весьма трудоемкий и до сих пор мало автоматизированный процесс. А во-вторых, связи между компонентами электронных библиотек и Интернет-страницами в этом случае также должны быть динамическими, поскольку такой тип ресурсов еще в большей степени подвержен изменениям, чем организованные полнотекстовые хранилища. Поэтому необходимо создание программ, которые будут собирать и периодически обновлять информацию о таких ресурсах, создавая для этого справочные элементы локальной электронной библиотеки. В задачи этих программ также должно входить выявление областей на Интернет-страницах, относящихся непосредственно к электронным публикациям, и их индексирование. Если говорить, например, о нахождении и индексировании электронных журналов в Интернет, то в качестве поисковых предписаний могут служить библиографические описания журналов и их издателей. Программы поиска и просмотра в этом случае будут обращаться к созданным индексам и формировать активные связи с внешними ресурсами.

Спецификации связей рассматриваемой нами ЭБ, когда ресурсы — источники связей — реализованы в виде баз данных, обеспечиваются средствами определения схемы БД. Возможность генерировать динамические связи основана на механизме спецификации типов и значений параметров каждой связи на логическом уровне как функции значений элементов данных текущей записи. Элемент записи (например, отдельный термин самого документа, пиктограмма или комментарий), визуализирующий связь, определяется в схеме как генерируемое поле, с которым связывается гипертекстовая ссылка-запрос. При активации такой ссыл-

ки происходит поиск или прямое обращение к указанному в запросе ресурсу, а найденные документы будут визуализированы как результат перехода по ссылке. Поскольку для базы данных можно определить неограниченное количество схем, такой способ обеспечивает достаточную гибкость как с точки зрения типологии, так и с точки зрения степени насыщения ссылками объектов ресурса, причем выбор схемы предоставлен пользователю при работе с ресурсом.

ОБОБЩЕННАЯ СХЕМА ЭЛЕКТРОННОЙ БИБЛИОТЕКИ И ОРГАНИЗАЦИЯ ДОСТУПА

В составе электронной библиотеки, общая схема которой представлена на рис. 2, можно выделить три блока:

сервер связи (Web-сервер или Z39.50-сервер), обеспечивающий доступ к программно-информационному комплексу ЭБ;

собственно информационный сервер — информационно-поисковая система, обеспечивающая доступ к информационным ресурсам ЭБ;

сервер управления доступом пользователей библиотеки к информационным ресурсам.

Сервер связи обеспечивает сетевое взаимодействие клиентской программы с программным обеспечением информационного сервера.

Как следствие интенсивного развития WWW-технологий сегодня “де факто” стандартом для электронных библиотек является HTTP-протокол, хотя Z39.50 создавался и развивался как специализированный протокол поиска и обмена данными именно между ЭБ. И хотя возможности организации поисковых процессов Z39.50 изначально существенно выше, чем HTTP-протокола, реализующего не более, чем точечное взаимодействие “без памяти”, простота последнего позволила, тем не менее, использовать его почти повсеместно, разрабатывая при этом собственные средства поддержки сессионного уровня, или игнорируя необходимость такового, ограничивая пользователя схемой “вопрос—ответ”.

Информационный сервер — это хранилище данных, представляющих коллекции разнородных логически или физически взаимосвязанных документов (образующих компоненты трехуровневой модели ЭБ), доступ к которым обеспечивается средствами информационно-поисковой системы. При этом поисковые (реферативные или библиографические) и метаинформационные компоненты традиционно реализованы в виде структурированных баз данных, в то время как компоненты уровня первичной информации могут быть представлены как полнотекстовыми БД, так и коллекциями файлов, каждый из которых хранит чаще всего образ в том или ином графическом или печатно-ориентированном формате.

Сервер управления доступом к информационным ресурсам, реализующий ограничения доступа к компонентам электронной библиотеки, является специализированной системой, обеспечивающей поддержку сессионного уровня поискового процесса, а также возможности административного учета и управления доступом на уровне отдельного пользователя к отдельному информационному и функ-

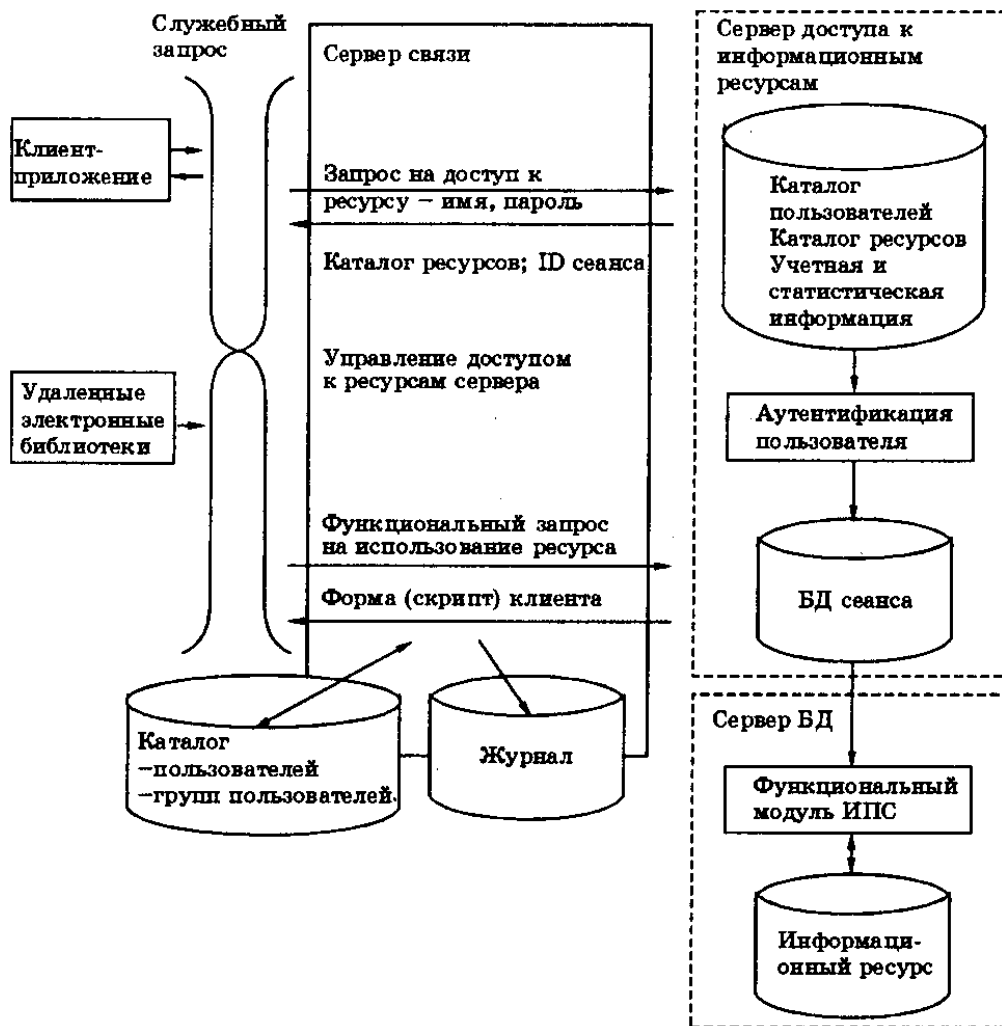


Рис. 2. Обобщенная схема электронной библиотеки

циональному ресурсу. Реализуется в виде специализированных баз данных, в том числе:

каталога информационных ресурсов, хранящего сведения о локализованных и ассоциированных ресурсах;

каталога пользователей, хранящего учетные данные о каждом пользователе и список доступных ему информационных ресурсов;

технологической базы данных сеансов, содержащей информацию о текущем состоянии сеанса каждой поисковой сессии;

базы данных учетно-статистической информации об использовании ресурсов.

Такие средства позволяют создавать каталоги доступных ресурсов для каждого отдельного пользователя, защищая определенные базы данных и/или массивы библиотеки от несанкционированного обращения. Кроме того, можно запретить доступ в случае, если физический адрес отправителя запроса отличается от адреса, зафиксированного в каталоге.

Таким образом сервер поддерживает уровень *поисковой сессии* как последовательность сеансов — операций, осуществляемых пользователем в среде информационно-поисковой системы электронной библиотеки. Такими операциями могут быть: выбор БД для поиска; выбор формата представления (схемы); установление параметров поиска, предпочтительных для данного пользователя; составление запроса; осуществление поиска; просмотр результатов; навигация в протоколе; завершение работы и т. д. (необязательно именно в та-

кой последовательности — например, просмотрев результаты поиска, можно провести новый поиск).

При этом будем считать, что *поисковый сеанс* — это одна операция, выполняемая в рамках одной транзакции — физического сетевого соединения, т. е. поисковая сессия — это последовательность функционально связанных поисковых сеансов, направленная на получение логически целостного результата.

Сеансовый уровень в рамках сеансового HTTP-протокола реализуется с помощью специализированной БД сеанса, в которую заносятся параметры всех операций поиска: какие установлены базы, схемы просмотра и параметры поиска (тип поиска, объем выдачи, способ сортировки результата, способ отображения найденных документов, четкий/нечеткий поиск, с маскированием или нет), каков механизм поиска (обычный поиск, эвристический), сколько документов найдено. Каждому сеансу присваивается уникальный идентификатор, с которым все функциональные запросы и их параметры заносятся в БД сеанса. Идентификатор сеанса также заносится в регистрационную базу, и при следующем посещении пользователем электронной библиотеки, после осуществления аутентификации, по этому идентификатору можно восстановить и продолжить поисковую сессию. К тому же, пользователь имеет возможность сохранить свой запрос с тем, чтобы воспользоваться им в дальнейшем, т. е. поисковая сессия реализуется как последовательность поисковых сеансов, где ка-

ждый последующий сеанс использует параметры предыдущего.

ОСОБЕННОСТИ ПРЕДСТАВЛЕНИЯ И ПОИСКА ИНФОРМАЦИИ

Особенности представления информации определяются их назначением — обеспечением эффективного (быстрого и исчерпывающего прежде всего по полноте) поиска нужных данных или, если таковые не обнаружены, — сведений о документах, предположительно их содержащих.

Именно поэтому для достижения общности представления, по возможности не зависящей от точек зрения и особенностей изложения, создаются каталоги и базы данных вторичной информации, когда содержание документа редуцируется до формы перечисления основных понятий, в той или иной степени однозначно характеризующих его в контексте той предметной области, для которой создается база данных. В свою очередь, для обозначения таких характеристических понятий используется ограниченная (нормализованная) лексика, снижающая влияние свойств синонимии и полисемии. Представление информационного содержания конкретных документов в виде поисковых образцов обеспечивает, с одной стороны, очень эффективную вычислительную процедуру (когда отбор производится по условию простого сопоставления термина запроса с термином документа), а с другой стороны — дает пользователю возможность получить достаточно полные и точные ответы на запросы, выражаемые упрощенным, “телеграфным” стилем. Кроме этого, в ряде случаев, когда предметом проблемного поиска является что-то неизвестное (по крайней мере пользователю), удовлетворительным результатом поиска скорее всего будет не отдельный документ, содержащий ответы на все вопросы, а некоторое множество документов, образующее для пользователя информационное пространство (фактов, идей и т. д.), достаточное для построения неизвестного до того решения, и, таким образом, — генерации новой информации. Т. е. в случае такого рода проблемных поисков, запрос — это не столько вопрос, предопределяющий ответ, сколько определение (через уже известное) предметной, документальной и лексической области для поиска неизвестного.

С другой стороны (внутримашинного представления информации), семантическая тривиальность структур данных, реализующих документальные системы в архитектуре фон-Неймановских машин, предопределяет, что развитие запроса и смысловая оценка результата поиска — исключительно прерогатива пользователя, а система — ассистент. И даже принимая во внимание определенные успехи в области разработок искусственного интеллекта, по крайней мере два фактора не позволяют надеяться на скорое равноправие сторон:

1) выражение запроса (даже хорошо формализованное и обязательно представленное на естественном языке) слишком лаконично для того, чтобы можно было бы выделить глубинную сущность и аспектные взаимосвязи проблемы, для решения которой должна быть найдена информация;

2) в ряде случаев пользователь не может однозначно специфицировать информационную потреб-

ность, особенно если она связана с этапом постановки задачи.

Таким образом, роль системы — не более чем информационно-технологическая поддержка на различных этапах многостадийного процесса информационного поиска, который в общем случае включает [7]:

1. Определение (локализация и формализация) темы запроса и идентификация (тематический выбор) ресурса, где система обеспечивает метаинформирование о тематике, наполнении, структуре и методах доступа к выбранному ресурсу.

2. Формирование (форма, способ ввода), а также адаптивное (структурное и лексическое) выражения запроса, где система предоставляет вспомогательные информационные объекты (словари, тезаурусы, шаблоны и т. д.).

3. Отбор документов по критерию, адекватному степени неопределенности информационной потребности, где система предоставляет выбор механизма поиска или, например, автоматически расширяет лексику запроса.

4. Формирование и управление выдачей найденных документов, где система обеспечивает масштабирование (форматирование) пространства представления выданных документов, а также сортировку и, возможно, ранжирование по некоторому формальному критерию соответствия.

5. Оценку результатов поиска с точки зрения полного удовлетворения информационной потребности (т. е. завершение поискового процесса) или их соответствия цели поиска и степени освоения информационного ресурса в части, соответствующей теме запроса (для обоснования решения о завершении процесса из-за отсутствия информации), где система позволяет количественно оценивать динамику выданных и обеспечивает возможность выборочного обращения к результатам отдельных этапов процесса поиска.

6. Развитие запроса по технологии “реформулирования по обратной связи по релевантности” или использование других ресурсов, где система обеспечивает регистрацию оценки пользователя и генерацию нового выражения запроса (включающего, например, статистически значимые термины из документов, которые пользователь отметил как действительно релевантные), а также связывание ресурсов, например ассоциированных баз данных вторичной или справочной информации.

Следует также отметить и некоторые особенности взаимоотношений человека и автоматизированной поисковой системы, обусловленные “заторможенным развитием” последних, сохранивших пакетную, вопросно-ответную идеологию, свойственную уже ушедшим системам информационного обслуживания, когда гарантом качества поиска был информационный работник. Задачей такого информационного посредника были понимание и интерпретация потребностей пользователя и, далее, использование знаний особенностей конкретной базы и поисковых средств, проведение эффективного (по крайней мере — профессионального) поиска. В современных условиях “информационного самообслуживания” пользователь, привыкший к интуитивному освоению программных сред (в основном стандартных средств операционной системы, большинство из которых имеет существенно более простой и дружелюбный интерфейс), часто неадекватно оценивает состояние и результаты поиска.

Крайними примерами ошибок может быть:

принятие безапелляционного решения о “плохой” базе данных или поисковой системе после получения неудовлетворительного или нулевого результата по первому же запросу, иногда даже не являющемуся правильным с точки зрения поискового языка;

прекращение пользователем развития запроса, если он получает известные или собственные публикации, т. е. когда происходит подмена критерия остановки процесса поиска по условию нахождения нужной новой информации или остановки по условию отсутствия новой информации в каждой следующей выдаче условием подтверждаемости “результативности” выражения запроса.

Кроме того, пользователь рискует обрести некоторую уверенность в том, что ИПС (особенно, когда речь идет о поисковых средствах Интернет) всемогущи и вездесущи, а их способности извлекать информацию (знания) из текстов и массивов документов бесконечно выше человеческих. Такой подход в самом безобидном случае приводит к некритичному отношению к результату поиска, т. е. пользователь удовлетворяется уже фактом получения выдачи (а современные поисковые системы часто устроены так, чтобы практически всегда пользователю выдавались какие-нибудь документы, пусть даже и в минимальной степени соответствующие запросу).

Поскольку система является всего лишь инструментом, используемым человеком при поиске, а не интеллектуальным автоматом для поиска информации, эффективность ее использования зависит от того, насколько хорошо человек знает природу объектов и свойства инструмента, посредством которого он с этими объектами работает.

В этом смысле, говоря об организации поиска информации, представленной в документальной форме (и отдавая должное истории развития АИПС), можно говорить о системах, основанных на одном из двух типов “стратегических” решений. Первые представлены “классическими” АИПС, берущими начало от библиотечных систем, информационный вход в которых реализуется через дополнительные (вторичные по отношению к текстам документов) справочные структуры. Вторые — гипертекстовые системы, в которых переход к потенциально полезному документу реализуется через контекстно определенную ссылку, статически размещенную непосредственно в тексте документа, обычно при его подготовке.

Относительная независимость (а иногда даже некоторая альтернативность) развития этих двух направлений в значительной степени обуславливалась функциональным различием информационного продукта и техническими ограничениями среды хранения. Имея в виду тождественность конечной задачи — обеспечение доступа к реально полезной и полной информации (а не только выдача документов по формальным признакам), гипертекстовые системы использовались для реализации локальных справочных систем, базирующихся на предметно-ориентированных, ограниченных коллекциях полнотекстовых документов, а эффективность поиска документов или их фрагментов достигалась через более или менее полную систему контекстно-определенных ссылок. ИПС же предназначались для обработки большого количества однородных

(регулярных), чаще всего вторичных документов, а контекст запроса заранее не может быть предопределен.

Можно сказать, что у этих двух типов систем принципиально различаются сценарии и средства выборки пользователем реально нужной информации. Координатное индексирование документов и использование в запросе терминов вне контекста предопределяет для классических АИПС необходимость последующей пользовательской оценки документов после их отбора системой по формальным признакам. Гипертекстовые же системы позволяют более целенаправленно (хотя лишь в рамках отдельного, текущего документа) управлять переходом к следующему документу за счет контекстной определенности ссылки. Соответственно (субъективно), легче выработка решения о завершении поиска по критерию удовлетворения потребности или исчерпания релевантных документов в просматриваемом потоке. Однако, по крайней мере для класса задач поиска новой информации, не укладывающейся в предопределенность типологии и графов ссылок, более гибким представляется метод координатного индексирования с его свойством комбинативности.

С появлением технических возможностей реализации полнотекстовых ИС альтернативность выбора какой-либо одной из двух “стратегических” технологий практически перешла в плоскость экономических решений о степени насыщения отдельных документов ссылками и приемлемых объемах словарей, индексирующих все или отдельные поля документов. Наиболее показательным примером является WWW-технология Интернет, где массивы документов изначально создаются по гипертекстовой технологии, а в дальнейшем строятся индексы, реализующие “классические” поисковые технологии.

Таким образом, в общем случае процесс поиска может быть представлен как *навигация* — целенаправленное и управляемое перемещение в документальном и лексическом пространстве базы данных, обеспечивающее оцениваемый уровень удовлетворения информационной потребности или объективно подтверждающее отсутствие информации.

Целенаправленность здесь предполагает наличие некоторой цели, обычно в сфере основной деятельности, которая в свою очередь может быть представлена как комплекс локальных (информационных) целей тематического (многоаспектный поиск) и/или технологического типа. Это означает, что для того, чтобы обеспечить целенаправленность избирательного обращения к поисковым объектам, интерфейсные средства системы должны позволять фиксировать и идентифицировать цель (точнее, технологические объекты, использованные в процессе ее достижения).

Управляемость же предполагает, с одной стороны, возможность выбора средств и/или параметров работы, а с другой — выборочное, в том числе повторное обращение к результатам и их обработку (например, статистическую или структурно-форматную). Кроме того, так как выделенная последовательность результатов (физически соответствующая отдельным шагам поиска, а логически — отдельной цели) образует ряд, это позволяет использовать статистически вычисляемые показатели (разностного типа), характеризующие сходи-