

## МИРОВЫЕ ИНФОРМАЦИОННЫЕ РЕСУРСЫ И ПРОБЛЕМЫ ПОИСКА ИНФОРМАЦИИ

С. Н. Селетков

Рассматривается состояние мировых информационных ресурсов и ресурсов знаний, основные проблемы, которые возникают у потребителя при поиске сведений из мирового информационного пространства, место и роль информатики в совершенствовании и развитии мировых информационных ресурсов и ресурсов знаний.

### 1. Основные понятия в области проблемы изучения информационных ресурсов

Понятие «информационные ресурсы» появилось более тридцати лет назад и прочно вошло в мировую и российскую практику информационной работы, в том числе содержание этого понятия отражено в российском законодательстве. Чтобы чётко определить содержание этого понятия рассмотрим следующие определения.

*Информация* — это сведения о лицах, предметах, фактах, событиях, явлениях и процессах независимо от формы их представления [1]. Теория познания утверждает, что человек познаёт реальный мир через объекты, свойства и отношения. В каждом виде деятельности человека свой набор объектов и интересующих человека свойств. Этот набор объектов составляет предметную область данного вида деятельности. Учитывая этот фактор, а так же дух закона [1] следовало бы после перечисления «...сведения о лицах, предметах, фактах, событиях, явлениях и процессах» добавить «и других объектах предметных областей».

*Документированная информация* (документ) — зафиксированная на материальном носителе информация с реквизитами, позволяющими её идентифицировать.

*Информационная система* — организационно-упорядоченная совокупность документов (массивов документов) и информационных технологий, в том числе с использованием средств

вычислительной техники и связи, реализующих информационные процессы.

*Информационные процессы* — процессы сбора, обработки, накопления, хранения, поиска и распространения информации. В настоящее время широко используется термин «*информационные и телекоммуникационные термины*». В соответствии со смыслом используемых терминов, это технологии реализации информационных процессов и процессов обмена информацией с помощью телекоммуникаций.

Информация, зафиксированная на материальном носителе и хранящаяся в информационных системах (библиотеках, архивах, фондах, банках данных, других информационных системах) образует *информационные ресурсы* [1].

Из определения понятия «информационные ресурсы» следует, во-первых, что существуют средства их обновления и доступа в интересах пользователя, т. е. средства поиска, обработки и выдачи; во-вторых, информация должна быть зафиксирована на каком-либо материальном носителе. Необходимо обратить также внимание на то, что информационные ресурсы — это документированная информация не только в базах данных, но также и в библиотеках, архивах и других информационных системах организационного типа.

В последнее время появилось ряд публикаций, в которых встречаются термины «*ресурсы знаний*», «*управление ресурсами знаний*», «*управление знаниями*», «*экономика знаний*» и ряд схожих терминов. Эти термины в законодательстве в области информатики и информационной работы в настоящее время не нормированы. Раскроем их содержание и покажем, в каком понимании эти термины будут использованы в представленной работе.

Рассмотрим понятие «ресурс знаний» в сравнении с понятием «информационный ресурс». Под понятием «знание» обычно понимается проверенный практикой результат познания действительности, верное её отражение в сознании человека. Специалисты в области теории познания отмечают многоуровневый характер отношения знания к действительности. Знания различных уровней обладают различной степенью достоверности. Мышление движется от поверхностного (первичного) знания — достоверных фактов к более глубоким научным знаниям, являющимся обобщением достоверных фактов. Факты или, как их ещё называют, *первичные знания*, сводятся к констатации фактов и их

описанию. Научные знания поднимаются до уровня объяснения фактов.

Будем понимать под знаниями всю совокупность знаний от описаний фактов (первичных знаний) до научных знаний.

Гигантский объём знаний, накопленных человечеством, доходит до нас потому, что они документированы, т. е. хранятся на каких либо носителях. Документированные знания в информационных системах являются, по аналогии с содержанием термина «информационный ресурс», ресурсом знаний, и в этом смысле понятия «ресурсы знаний» и «информационный ресурс» — синонимы. Человечеству для своего развития должна быть представлена информация от первичных знаний до научных знаний, и специалист с его интеллектом сам определит, какие из описаний фактов уже являются знаниями. Иногда это очевидно — открытие на небосклоне новой звезды, географическое открытие и т. п.

Важным в практической деятельности являются «живые знания», т. е. те знания, которые в настоящее время существуют в сознании людей. Организация или фирма имеет специалистов, соответствующих профилю этой организации. От знаний этих специалистов и возможностей использования этих знаний в практической деятельности зависит успех фирмы или организации. Естественно, что знания специалистов не могут быть отнесены к ресурсам знаний по существу определения термина «ресурс», часто их называют корпоративными знаниями.

Человек, владеющий знаниями, может отразить их в документах (статьях, монографиях, заявках на изобретения), т. е. документировать. Если документированные знания попадают в информационные системы, то они становятся доступными потребителю и относятся к информационным ресурсам. Проблематика совершенствования и использования знаний специалистов (не документированная информация) в данной работе, несмотря на её важность, не рассматривается.

## 2. Мировые информационные ресурсы и актуальные проблемы их использования

В последние годы часто используется термин «*мировое информационное пространство*» или «*мировые информационные ресурсы*». Под этим термином понимается совокупность информационных ресурсов, доступных мировому сообществу, то есть

потребителям различных стран мира. Доступность мировых информационных ресурсов потребителям из различных стран стала возможной благодаря развитию информационных технологий. Мировые информационные ресурсы предоставляются бесплатно или на коммерческой основе.

Мировые информационные ресурсы в имеющейся литературе обычно разделяются на три сектора:

- сектор деловой информации;
- сектор научно-технической и специальной информации;
- сектор массовой, потребительской информации.

Сектор деловой информации, в свою очередь, подразделяется на следующие группы: биржевая и финансовая информация, статистическая информация, коммерческая информация и деловые новости.

Сектор научно-технической и специальной информации включает документальную библиографическую, реферативную и полнотекстовую информацию о фундаментальных и прикладных исследованиях и профессиональную информацию для специалистов различных предметных областей: юристов, врачей, работников образования, инженеров и т. д.

Сектор массовой, потребительской информации включает новости и справочную информацию, потребительскую и развлекательную информацию (погода, расписание транспорта, аренда машин, цены на оптовых рынках, справочники служб сбыта и т. д.)

Как видно из приведённой выше классификации, мировые информационные ресурсы обеспечивают необходимыми сведениями все виды деятельности общества, в том числе экономику.

С позиции используемых средств, мировые информационные ресурсы разделяются на информацию, хранимую в традиционных информационных системах (библиотеках, архивах), и информацию, содержащуюся в электронных профессиональных базах, деловых ресурсах Интернет и электронных библиотеках [5]. Традиционные средства, содержащие документированную информацию, библиотеки и архивы также имеют компьютеры, но эти средства автоматизируют лишь часть функций поиска — сами же документы хранятся в библиотеках и архивах в печатном виде.

Основой мировых информационных ресурсов являются профессиональные базы, которых в настоящее время насчитывается свыше 20 тыс. Формируют информацию и продают её по-

ставщикам на коммерческой основе около 8 тыс. информационных агентств. К более чем 60% профессиональных баз имеется непосредственный доступ пользователей через телесети. В связи с тем, что источники информации в профессиональных базах тщательно отбираются, информация характеризуется высокой степенью достоверности и постоянно обновляется. При хранении в профессиональных базах заранее определено, какая информация будет накапливаться в каком хранилище, что облегчает потребителю её поиск.

Основой деловых ресурсов Интернет являются *web-страницы*. Этот ресурс представляет собой страницы гипертекста, в которых может также содержаться графическая, звуковая и видеоинформация. В основном, этот ресурс содержит рекламную информацию и он может быть использован как дополнение к профессиональным базам.

Последнее время получили своё развитие *электронные библиотеки*, однако область их применения существенно ограничена законодательством в области авторского права.

Одна и та же информация может быть представлена в библиотеках, в профессиональных базах, в деловых ресурсах Интернет и в электронных библиотеках. Объём информации в профессиональных базах — свыше 10 млрд записей. В России в 150 тыс. библиотек содержится свыше миллиарда экземпляров изданий, а количество электронных библиотек лишь около 400.

В настоящее время в мире издаётся свыше 200 тыс. периодических изданий, из них две трети — без ограничений по направлению. Чуть более 10% периодических изданий представлено в Интернет. Из 16 тыс. журналов в области экономики и бизнеса 12% журналов в полном объёме представлено в Интернет. Число web-страниц в Интернете в 2002 году превысило 4 млрд.

Приведённые характеристики объёмов информации, накопленных в мировых информационных ресурсах различных типов, свидетельствуют о гигантском объёме знаний, созданных мировым сообществом и доступных пользователям. По мнению специалистов, мировой объём накопленной информации удваивается каждые три года.

Изменения, которые происходят в обществе и экономике под влиянием увеличения объёмов доступной информации и внедрения современных информационно-телекоммуникационных технологий (ИКТ), наиболее чётко сформулированы в Окинавской Хартии глобального информационного общества, подписанной

в 2000 г. Великобританией, Германией, Италией, Канадой, Россией, США, Францией и Японией.

1. Информационно-телекоммуникационные технологии оказывают определяющее влияние на формирование общества XXI в. Это влияние оказывается на экономику, образ жизни людей, взаимодействие государственных структур и населения.

2. Существо изменений в экономической и социальной сфере заключается в максимальном использовании знаний и передовых идей. Должен быть преодолен разрыв стран в области доступа к информации и знаниям.

3. Хартия подтверждает приверженность принципу — все люди должны иметь равные права пользоваться преимуществами глобального информационного общества.

4. Хартия обращается с призывом ко всем ликвидировать разрыв между странами в области информации и знаний.

Рассмотрим теперь, каковы потребности пользователя мировых информационных ресурсов и удовлетворяют ли мировые информационные ресурсы эти потребности. Очевидно, что потребности пользователей, работающих в различных областях, будут отличаться. Учитывая, что Окинавская Хартия и другие документы, определяющие задачи мирового сообщества в XXI в., ставят задачу максимального использования информации и знаний в экономической и социальной сфере, будем рассматривать требования потребителей информации применительно к экономике. Для предприятий и фирм в условиях рыночных отношений потребности в информации делятся на две группы:

— потребности в информации о собственных ресурсах (внутренней информации) — 10–15% от общих потребностей фирмы в информации;

— потребности в информации о внешней среде — 85–90% от общих потребностей в информации.

Информацию о внешней среде, в которой ведёт деятельность предприятие или фирма, предоставляет мировой рынок информации, мировые информационные ресурсы.

Информация о внешней среде содержит следующие группы данных:

1) сведения о состоянии мирового и региональных финансовых рынков;

2) сведения, характеризующие рынки сбыта;

3) сведения, характеризующие возможных партнёров;

- 4) сведения о степени коммерческих и финансовых рисков в различных странах и регионах мира;
- 5) сведения о финансовом состоянии компаний различных стран;
- 6) сведения о продукции и услугах компаний различных стран;
- 7) сведения о рынках недвижимости;
- 8) сведения о патентах и новейших технологиях;
- 9) сведения о законодательстве различных стран мира;
- 10) отчёты аудиторских фирм;
- 11) данные о внешней торговле по различным странам;
- 12) сведения о состоянии страхового рынка;
- 13) официальная политическая и экономическая информация по странам;
- 14) деловые новости ведущих мировых Агентств.

Перечисленные группы сведений в полном объёме представлены в мировых информационных ресурсах и используются в развитых странах при принятии решений в бизнесе и государственном регулировании экономики.

Могут быть выделены следующие основные требования пользователей к мировым информационным ресурсам.

1. Пользователь хотел бы получить прямой ответ на свой вопрос. Статистика мировых информационных ресурсов показывает, что в настоящее время в общем объёме мировых информационных ресурсов содержится лишь 16 % формализованной информации и 84 % текстовой. Из этих соотношений следует, что лишь шестая часть накопленной мировым сообществом информации представлена в мировых информационных ресурсах в виде, удовлетворяющем потребности пользователя.

*Формализованная информация* — это значение свойств каких-либо объектов, представленных в виде высказываний. Значения свойств могут быть представлены в виде чисел, слов или словосочетаний (например, фамилия физического лица).

Формализованная информация понятна человеку и компьютеру в смысле возможностей автоматизированной обработки такой информации. Выданная формализованная информация в целом удовлетворяет потребностям пользователя и по его указанию может быть подвергнута автоматизированной обработке.

Текстовая информация может содержать необходимую пользователю информацию или смысл текста с позиций потребности пользователя. Однако, выделить смысл из произвольного текста,

в связи с избыточностью и недостаточностью естественного языка, может только человек. Компьютер не обладает способностями человека по анализу текста с целью выделения его содержания, однако, он обладает колоссальными возможностями обработки информации, если она формализована. Этот факт очевиден, компьютер может сравниться с человеком по интеллекту, если он пройдёт весь путь развития вместе с человеком, начиная с его рождения. С момента появления вычислительной техники принимались неоднократные попытки опровергнуть этот очевидный факт, то есть создать автоматизированные алгоритмы выявления смысла из произвольных текстов. Эти попытки окончились неудачей.

Поиск документов, в которых встречаются заданные слова, имеет очень отдалённое отношение к задаче выявления смысла в произвольном тексте. Отметим, что выделить смысл из текста применительно ко всему мировому информационному пространству и создать фактографическую систему, содержащую эту информацию, невозможно. Эту задачу можно решить лишь применительно к потребностям какой — либо конкретной предметной области, где определены классы объектов, их взаимосвязи, свойства объектов и области возможных значений этих свойств, то есть информационные потребности, вытекающие из задач данной предметной области.

Человек, обладая аналитическими способностями по выявлению содержания смысла текста, сталкивается с непреодолимыми трудностями, если ему нужно сопоставить смысл нескольких текстов. Предел его способностей — это анализ 4–5 текстов, если они не очень большого объёма. Следовательно, в большинстве случаев пользователю будут выданы тексты документов, в которых может содержаться нужная ему информация, а не ответ на вопрос пользователя. Если документов много, пользователь зачастую просто отказывается от их анализа. Таким образом, успех в работе с мировыми информационными ресурсами может быть обеспечен только в случае возложения задач выявления смысла из текстов на человека и использования компьютера для выполнения массовых операций по сопоставлению и обработке формализованной информации.

Анализ процентного соотношения формализованной и текстовой информации в мировых информационных ресурсах в динамике показывает, что двадцать лет назад в мировых информационных ресурсах формализованной информации было 52 %, а тек-



стойкой 48%, что говорит об устойчивой тенденции ухудшения удовлетворения потребностей пользователя.

В настоящее время в мире насчитывается свыше 660 млн компьютеров, и этот гигантский потенциал не может быть эффективно использован для общения и обработки основной части мировых информационных ресурсов. Компьютер используется сейчас при работе с текстовой информацией используется сейчас лишь как телеграфный аппарат или средство электронной почты. Компьютер отправляет запрос и распечатывает ответ. Компьютеры информационных агентств в части обработки текстовой информации осуществляют упрощённый поиск текстов по словам с качеством хуже, чем в информационных системах поиска в текстах, которые использовались 40 лет тому назад.

**2.** Потребителю хотелось бы, чтобы мировые информационные ресурсы адекватно отображали реальный мир в соответствии с требованиями каждой предметной области.

На самом же деле 84% объёма мировых информационных ресурсов — это текстовая информация, которая отражает реальный мир лишь фрагментарно. В хранимых документах источник описывает реальный мир в виде фрагмента, его части, то есть то, что он видит или то, что он получил в результате анализа, обобщения. Отсутствие возможности автоматически обработать смысловое содержание каждого документа, представленного на естественном языке, и, следовательно, обобщать полученную информацию, ставит перед потребителем сложную, а иногда и невыполнимую задачу — в условиях крайне низкой полноты и большого шума проводить анализ и обобщение фрагментов отображения реального мира для решения задачи, стоящей перед потребителем.

Опустим вопрос о достоверности самого источника информации. Компьютеры не в состоянии выявить смысл текста и по этой причине не могут интегрировать информацию и из отдельных фрагментов сформировать более общее описание реальной действительности, чем это содержится в отдельном тексте (фрагменте). Потребителю в этих условиях бессмысленно задавать запрос, требующий смысловой обработки ряда текстов.

**3.** Потребитель нуждается в том, чтобы в ответе на запрос отсутствовало дублирование информации. Рассмотрим проблему *избыточности информации* в мировых информационных ресурсах. Новые знания должны появляться в публикациях: статьях в междотраслевых и отраслевых журналах, монографиях, диссер-

тациях, отчётах по научно-исследовательским работам и в ряде других источников. Когда автор приходит со своей публикацией в редакцию, то, в идеале, должна быть система, которая должна определить, имеются ли новые знания в представленной публикации или же основной смысл представленного материала уже опубликован. К сожалению, мировое сообщество до сих пор не разработало эффективных средств исключения дублирования в мировых ресурсах знаний.

В этом направлении больших успехов добились патентные службы. К автору изобретения предъявляются жёсткие требования в части формулировки существа изобретения. Автор обязан сформулировать, что он предлагает, какой эффект будет достигнут от применения его идеи, за счёт чего достигается этот эффект. Работник патентной службы сопоставляет предложение автора с имеющимися в патентных базах заявками других авторов, представленных ранее, и на основании результатов анализа принимается решение, что представленная заявка обладает новизной. Другим исходом является отказ и уведомление автора, что данное предложение было зарегистрировано ранее по заявке другого автора.

Отметим, что относительно успешный пример идентификации знаний в патентном деле стал возможным при участии эксперта патентной службы, наличия у эксперта доступа к массиву ранее поступавших заявок и принятых по ним решений, а также жёстких требований к заявителю по форме представления заявки. Ясно, что такая процедура идентификации знаний будет требовать соответствующих затрат.

В требованиях к диссертационным работам указано, что сам автор обязан провести сравнение своих результатов с уже известными по основным показателям, в том числе, по стоимости. Представленная работа проходит проверку, в том числе, и на новизну, которая выполняется оппонентами, головной (ведущей) организацией и организациями, дающими отзывы на рефераты. После защиты Учёный Совет в своём заключении должен отметить новизну основных положений работы. Естественно, что сам Совет принимает решение на основе заключений оппонентов и организаций, представивших отзывы, которые при решении вопросов оценки новизны опираются лишь на собственные знания и не имеют доступа к мировым информационным ресурсам. По существу, это оценка отдельных специалистов, а не сравнение с мировыми информационными ресур-

сами, содержащими накопленные всем мировым сообществом знания.

В издательствах, особенно авторитетных, публикации на этапе рассмотрения проходят рецензирование, в том числе и проверку на новизну. Качество рецензирования, особенно оценка новизны, определяются уровнем знаний рецензента и только.

По мнению специалистов, в ресурсах знаний имеется десятикратная избыточность, то есть новое знание десятикратно повторяется. Отсутствие у мирового сообщества эффективной системы исключения избыточности в ресурсах знаний не просто увеличивает время просмотра информации. Во многих случаях потребитель просто отказывается от поиска необходимой информации, не имея времени на просмотр десятков, а иногда и сотен, документов при такой избыточности, т. е. шума.

**4.** Пользователь нуждается в максимальной полноте выдачи информации. Для оценки качества поиска нужных сведений в информационных системах используют такие показатели, как полнота и точность. Эти показатели характеризуют способность системы выдавать только документы, релевантные именуемому запросу. Под *релевантностью* понимается объективно существующее смысловое соответствие между содержанием документа и запроса [3].

Под *полнотой выдачи* понимается отношение множества релевантных и выданных системой документов к множеству находящихся в системе релевантных документов.

Под *точностью выдачи* понимается отношение множества релевантных и выданных системой документов к множеству выданных системой документов (релевантных и нерелевантных).

### **3. Основная теоретическая задача информатики в определении общих закономерностей создания, преобразования и использования информации**

Первоначально, полвека тому назад, под информатикой понималась дисциплина, изучающая структуру и общие свойства научной информации, а также закономерности её создания, преобразования, передачи и использования в различных сферах человеческой деятельности, включая теоретические и практические аспекты оценки эффективности поиска информации в произволь-

ных текстах. В те годы основные усилия были направлены на совершенствование информационных технологий в библиотечных сетях и системах научно-технической информации. В этой области были достигнуты значительные успехи. Большое внимание уделялось теоретическим и практическим вопросам создания автоматизированных информационно-поисковых систем. Велись глубокие исследования в части анализа возможностей автоматизированной обработки текстовой информации, построения специальных поисковых языков, разработки критериев полноты и точности, характеризующих эффективность поиска в документальных системах, и многие другие проблемы.

Исследовался естественный язык с позиций возможности эффективного поиска информации в автоматизированных информационно-поисковых системах. Было показано, что источник, представляющий документ в информационную систему, описывает фрагмент реального мира исходя из своих субъективных представлений о предметах и явлениях реального мира. Также субъективно описывает запрос и потребитель. Неоднозначность представлений о средствах описания реального мира у источников и потребителей резко снижает эффективность поиска в информационных системах.

Для повышения эффективности поиска специалисты в области информатики пошли по пути разработки семантических средств информационных систем. Под *семантическими средствами* понимаются информационно — поисковые языки, методы (алгоритмы) индексирования, методы поиска информации.

Основная идея использования информационно-поисковых языков в общем виде заключается в том, что информация, поступающая от источников, с помощью специалистов описывается на информационно — поисковом языке и это описание вводится в систему как поисковый образ документа. Запрос в систему также описывается специалистами на том же информационно-поисковом языке и создаётся поисковый образ запроса. Далее по соответствующим правилам в массиве поисковых образов документов отыскивается документ, в максимальной степени удовлетворяющий поисковому образу запроса.

При такой схеме решается задача однообразного описания общего содержания документа и запроса в интересах улучшения качества поиска. Наибольшее распространение в автоматизированных информационных системах получили *дескрипторные*

языки и семантические коды. Для сравнения семантических средств различных информационных систем использовались такие критерии, как полнота выдачи и точность (шум).

Исследования сравнительной эффективности различных информационно-поисковых систем проводились в разных странах. Так, в 1957–1966 годах английской ассоциацией специальных библиотек и информационных бюро был реализован Кренфильдский проект по сравнению эффективности различных ИПС [2]. Исследовались 33 информационно-поисковых языка и влияние семантических средств на полноту и точность систем.

Широко велись работы в области теории и практики внедрения ИПС в информационное обслуживание и в нашей стране. Были созданы многочисленные ИПС, среди которых системы «Реферат», «Асинит», «Ассистент» и многие другие. В результате исследований было показано, что существуют зависимости между показателями полноты и точности выдачи, а именно: чем больше полнота, тем меньше точность и наоборот.

Было показано также, что в документальных ИПС в принципе не могут быть достигнуты 100% показатели полноты или точности. Исследования, проведенные в Ленинградском государственном институте культуры в 1964–1966 годах, показали, что в традиционных ИПС, применяющих малую глубину и глубину индексирования и имеющих информационно-поисковые языки с малой семантической силой, потери информации и информационный шум составляют 50–60%. Кроме того, было установлено, что, чем больше затраты специалистов на ввод документов в систему, тем выше качество поиска.

Научно-исследовательские работы, проведенные в 60-е годы минувшего века, позволили создать прочную базу для дальнейшего развития направления, связанного с совершенствованием средств автоматизации поиска информации в интересах всех сфер деятельности общества. В эти годы информатика стала самостоятельной научной дисциплиной, включающей теорию информационно-поисковых систем. Представим себе, что специалист в области информатики уснул в 60-е годы прошлого столетия ледяным сном и проснулся в 2006 году. Он обнаруживает, что на дворе век информации и век информационного общества. В мире существует мировое информационное производство, новые технические средства: телекоммуникационные сети, гигантские возможности электронной памяти, совершенные компьютеры и при всем этом — всё ухудшающиеся возможно-

сти потребителей в удовлетворении их информационных потребностей.

Информация в условиях рынка является товаром, а отрасль информации в мире является одной из важнейших. Сегодня на коммерческой основе в мире работают в информационной отрасли свыше 8 тыс. крупных агентств. А «компаса», который бы определил правильное направление развития этой отрасли, к сожалению, нет. А где же наука об информации с созданными ранее заделами? Оказывается, в нашей стране, ранее занимавшей лидирующие позиции в мире в области информатики, теперь информатика — это не наука об информации, а прикладная дисциплина, посвящённая компьютерам, алгоритмическим языкам и пакетам прикладных программ. Скорее всего, она обслуживает направления проектирования программных средств информационных систем.

Особенно остро проблема поиска нужных сведений стоит при поиске деловых ресурсов в сети Интернет. Цель поиска — найти адрес сайта, где содержится нужная потребителю информация. Для этого, как правило, используется следующая схема. Исходная информация размещена на множестве сайтов, между потребителем и сайтами создан третий уровень — поисковые системы Интернет. Следовательно, оценивать полноту и точность выдачи информации необходимо, анализируя характеристики поисковых машин Интернет.

Анализ показывает, что в существующей схеме поиска не используется никаких средств для однообразного описания исходной информации и запроса, а сами поисковые машины находятся на уровне пятидесятилетней давности (до появления информационно-поисковых языков.). Как показали проведенные исследования, полнота в этом случае не может быть выше 50%. Учитывая, что по данным исследований компетентных в этой области организаций, лучшие информационные системы Интернет просматривают не более 30% сайтов Интернет и при этом на каждом сайте просматривают только видимую часть (1–5% объёма сайтов), полнота поиска в Интернет с помощью поисковых систем составит менее 1%.

Отсутствие публикаций результатов количественного анализа характеристик поиска информации в Интернет по полноте выдаваемой информации и информационному шуму вводит потребителя в заблуждение. Потребитель, как правило, не представляет, что объём не выданной, но удовлетворяющей потре-

бителя информации, то есть релевантной информации, на два порядка превышает объём выданной. Если вы получили в ответе на запрос 10 документов, вы должны знать, что 990 (!) релевантных документов осталось не выданными. Такие оценки представляются даже завышенными, так как половина документов в Интернете представлена на английском языке, а остальные документы — на языках других народов мира.

Изменить данные оценки в реальной практике не представляется возможным, так как многие владельцы сайтов в интересах безопасности сами закрывают доступ поисковых систем Интернет к своей информации.

В профессиональных базах полнота и точность поиска существенно выше, чем в Интернет, за счёт того, что разработчики профессиональных баз при проектировании заранее разрабатывают структуру хранилищ и закрепляют каждое хранилище за определённым видом информации. Однако в каждом хранилище, содержащем массив документированной информации, существуют те же проблемы, которые перечислены выше.

#### 4. Россия и мировые информационные ресурсы

Российское руководство, подписав Окинавскую Хартию вместе с другими странами «большой восьмёрки» в 2000 году, подтвердило понимание той большой роли, которую играет информация и знания в экономике и развитии общества в настоящее время. Что, по нашему мнению, необходимо сделать в России, чтобы мы на практике имели равный доступ к мировым информационным ресурсам наравне с развитыми странами.

1. Широкий круг специалистов, работающих с информацией, должен быть специально подготовлен. Рассмотрим опыт подготовки кадров передовых зарубежных стран, в частности, стран «большой восьмёрки», которые более тридцати лет работают с профессиональными базами и последние годы с деловыми ресурсами Интернет. Умение работать с мировыми информационными ресурсами предусматривает владение компьютерной и информационной грамотностью.

Под *компьютерной грамотностью* понимается умение работать с компьютером и владение навыками работы с прикладными программами, позволяющими осуществлять профессиональную деятельность.

*Информационная грамотность* предполагает умение выделить при решении стоящей задачи требования к информации, знание структуры мировых информационных ресурсов, стоимости информации, владение технологиями доступа к необходимым сведениям, а также умение находить необходимую информацию, её оценивать и использовать при решении конкретных задач. Информационная грамотность позволяет добыть нужную для принятия решений информацию и, используя её, подготовить обоснование принимаемых решений.

Компьютерная грамотность необходима современному специалисту, так как освоение информационной грамотности должно опираться на знание современных технических и программных средств и умение с ними работать.

Освоение компьютерной и информационной грамотности должно проходить последовательно: сначала компьютерная, а затем информационная грамотность. По существу, *информатика это наука об информации* и её прикладное значение — владение компьютерной и информационной грамотностью.

В российских школах и вузах даются знания и умения только в области компьютерной грамотности. Обучение в области информационной грамотности практически не ведётся. Исключение составляет специальность «Прикладная информатика (по областям)», где в дисциплине «Мировые информационные ресурсы» даются сведения о мировом рынке информации. Но и в этой специальности в требованиях к специалисту отсутствуют требования к умению работать с мировыми информационными ресурсами.

Анализ требований практики в области информационной грамотности и результаты исследований методики подготовки специалистов по данному направлению в ведущих университетах США: Harvard University, University of Massachusetts, California State University, University of Washington, State University of New York, а также Western Association of Schools and Colleges позволили выявить используемые в США стандарты в области информационной грамотности. Эти стандарты включают следующие требования к знаниям и умениям специалистов:

- знания и навыки, позволяющие определять информационные потребности для решения конкретной задачи или проблемы;
- знания и навыки, позволяющие работать на рынке информации с учётом существующих законодательных и этических норм;



- знания и навыки эффективного доступа к информационным ресурсам;
- знания и навыки оценки полученной информации;
- знания и навыки использования найденной в мировых ресурсах информации для решения конкретной задачи или проблемы.

Московский Государственный Университет экономики, статистики и информатики последние 10 лет ведёт работу по подготовке специалистов в области умения работать с мировыми информационными ресурсами в соответствии с мировыми стандартами, используемыми развитыми странами. За это время подготовлены необходимые учебно-методические материалы, в том числе, для обучения в среде e-learning. В 2004 году в Издательстве Питер вышло учебное пособие «Мировые информационные ресурсы», в сентябре 2006 года в издательстве Финансы и статистика выпущен учебник «Управление информационными ресурсами». Таким образом, подготовлена база для обучения российских специалистов в различных областях по использованию мировых информационных ресурсов.

2. Помимо квалифицированных специалистов, умеющих работать с мировыми информационными ресурсами, созданными мировым сообществом, требуется наличие средств доступа к мировой электронной информации.

Большинство стран, подписавших Окинавскую Хартию, уделяют особое внимание внедрению современных информационных технологий в экономику и социальное развитие, в том числе средств доступа к мировой электронной информации. В табл. 1 приведены данные о числе персональных компьютеров и числе подключений к Интернет в 2000 году по странам, подписавшим Хартию. Следует отметить, что при 20 процентном ежегодном приросте парка компьютеров в России мы не компенсируем даже ежегодное моральное старение трети парка этих средств.

Приведённые в табл. 1 данные позволяют сделать следующий вывод: не готовя специалистов в области работы с мировыми информационными ресурсами и имея отставание от развитых стран «большой восьмёрки» в десятки раз по числу персональных компьютеров и числу подключений к Интернет, Россия не может практически реализовать своё право на равный доступ к накопленным человечеством знаниям и информации и не может быть равной в конкурентной борьбе на мировом экономическом пространстве.

Таблица 1

Страны	Число персональных компьютеров на 1000 чел	Число подключений к Интернет на 1000 человек
Великобритания	338	301
Германия	336	292
Италия	180	230
Канада	390	413
Россия	43	7
США	585	347
Франция	304	144
Япония	425	300

Источники: Тенденции в странах Европы и Северной Америки 2003 (статистический сборник ЕЭК ООН). — М.: Госкомитет РФ по статистике, 2004. — 320 с.

## 5. Заключение

Подводя итог рассмотрению проблем поиска в мировых информационных ресурсах, можно сделать следующие выводы:

1. В настоящее время имеет место информационный кризис, который заключается в том, что объёмы информации, хранимой в мировых информационных ресурсах и ресурсах знаний, всё увеличиваются, а степень удовлетворения информационных потребностей пользователей по полноте выдаваемой информации и информационному шуму всё ухудшается. Особенно отрицательно информационный кризис проявляется в области информации общего обеспечения экономики, где требования к информации по полноте и достоверности выше, чем в других отраслях.

2. Постоянный рост за последние десятилетия доли текстовой информации в мировых информационных ресурсах практически исключает возможность использования мирового парка компьютеров для интеграции информации и её обработки в интересах пользователей. Компьютеры, обладающие колоссальными возможностями по массовой обработке формализованной информации, её интеграции и сопоставлению, практически исключены из информационных процессов обеспечения пользования. Нарушены естественные соотношения: анализ и выявление смысла из текста за человеком, массовые операции над формализованной

информацией по заданному человеку алгоритмам, в том числе интеграция информации, поиск и обработка — за компьютером.

3. При построении деловых ресурсов Интернет практически не используются теоретические разработки и проверенные практикой технологии в области информатики. При построении профессиональных баз используются отдельные элементы эффективных информационных технологий. Однако, основной объём профессиональных баз — это тексты и, следовательно, перечисленные проблемы касаются и профессиональных баз.

Предварительное распределение информации по базам даёт определённые преимущества, но необходимо учитывать влияние закона рассеивания информации, по которому информация будет попадать не только в те базы, которые для неё выделены, но и в другие.

4. Отмечая недостатки, имеющие место в построении информационных ресурсов и организации поиска информации в них, мы должны понимать, что существующее мировое информационное пространство содержит колоссальный объём знаний, который каждая страна должна сегодня уметь извлекать в интересах своей экономики и развития общества.

5. Выход из информационного кризиса видится в формализации наиболее важной для данной предметной области информации в тексте. Формализованная часть текста, удобная как для человека, так и для возможностей компьютерной обработки, будет являться дополнением к тексту. Поиск по формализованной информации сразу даст увеличение полноты поиска на два порядка за счёт того, что однообразно будет описываться информация, предоставляемая источником, и запрос, формируемый пользователем.

Затраты на формализацию в Интернете должен нести владелец сайта. Эти затраты могут быть разовыми или при создании сайта, или же при изменении наиболее важной информации о фирме. Выигрыш очевиден, сайт делается для рекламы, и формализация наиболее важной информации увеличит количество потребителей, которые найдут этот сайт, в десятки раз. В профессиональных базах формализацию могут проводить или Агентства — Генераторы или Поставщики, перенеся расходы на увеличение стоимости информации, выдаваемой с существенно большей полнотой.

Формализованный язык разрабатывается под задачи конкретной предметной области. При этом определяются объекты, свой-

ства, наиболее важные по каждому объекту, и области возможных значений этих свойств [3]. Например, это можно сделать в интересах экономики для деловых ресурсов Интернет.

По своей структуре формализованный язык соответствует восприятию человеком реального мира, то есть любой предметной области. Он построен на строгой математической основе — логике высказываний. Базы данных, содержащие информацию, описанную на таком языке, представляют совокупность значений истинности высказываний о данной предметной области. Формализованный язык, построенный на логике высказываний, однозначно воспринимается как человеком, так и компьютером, что позволяет из элементарных высказываний автоматически формировать наиболее сложные, то есть из отдельных фрагментов отображения реального мира (исходных сообщений от источников) формировать интегрированное описание реального мира. Язык запроса в такой системе строится на основе логики предикатов, то есть логике, свойственной как мышлению человека, так и логике высказываний, на которой построены электронные базы.

Используемые термины для описания классов объектов, их свойств и областей возможных значений этих свойств берутся из терминологических словарей, имеющихся в каждой предметной области, и классификаторов (общероссийских и отраслевых), что позволяет однообразно описывать сообщения от источников и запросы от потребителей.

Поскольку формализации подлежат наиболее часто используемые при запросах сведения, то существует необходимость находить исходные тексты документов. При предоставляемой концепции такой поиск осуществляется через фактографические базы и, используя связи между формализованной информацией и исходной, выдаётся искомый текст. В случае сохранения индексирования может использоваться и традиционный поиск.

К программному обеспечению, реализующую предлагаемую концепцию, предъявляется единственное требование — возможность работать с базами, содержащими сведения об отдельных объектах, и связями между объектами различных баз. Такими возможностями обладают многие из программных средств, существующих в нашей стране и за рубежом, например программные средства Справочных правовых систем (СПС).

Предлагаемая технология построения систем поиска в текстах с формализацией наиболее важной смысловой части

текста прошла апробацию в российской практике, которая подтвердила её высокую эффективность, и может быть использована для совершенствования мировых информационных ресурсов: деловых ресурсов Интернет и профессиональных баз данных и знаний.

### Список литературы

1. Федеральный закон Российской Федерации «Об информации, информатизации и защите информации» от 20 февраля 1995 г., № 24-ФЗ.
2. *Захаров В.П.* Информационные системы (документальный поиск). — СПб.: 2002. — 188 с.
3. *Овчаров Л.А., Селетков С.Н.* Автоматизированные банки данных. — М.: Финансы и статистика, 1982. — 262 с.
4. *Соколов А.В.* Информационно-поисковые системы: Учебное пособие для вузов / Под ред. А.Б. Рябова. — М.: Радио и связь, 1981. — 181 с.
5. *Хорошилов А.В., Селетков С.Н.* Мировые информационные ресурсы. — СПб.: Питер, 2004. — 176 с.
6. *Хорошилов А.В., Селетков С.Н., Днепровская Н.В.* Управление информационными ресурсами. — М.: Финансы и статистика, 2006. — 272 с.