

УДК [002:51]:002.6 ВИНТИ

И. Ю. Никольская

Проблемы создания и перспективы развития информационной системы “Математика” в составе информационной системы ВИНТИ

Описывается специальная информационно-поисковая система по математическим наукам, основанная на кооперации ВИНТИ и нескольких математических центров.

В настоящее время информационное обеспечение исследований в области математических наук предоставляется пользователям в виде двух реферативных журналов (РЖ) — “Математика” и “Вычислительные науки” — и двух ежегодных указателей (“Авторско-библиографический” и “Систематическо-предметный”). С 2001 г. пользователям предоставляются две одноименные реферативные базы данных (РБД), являющиеся электронным аналогом названных реферативных журналов.

Обращаясь к РЖ и РБД, пользователь получает информацию об опубликованных в мире материалах (статьях, монографиях, тезисах докладов и т. д.) с библиографическими описаниями, кратким содержанием каждой публикации и классификационными индексами по иерархической классификации математики.

Однако полнота обрабатываемой информации не может быть гарантирована, так как не все публикации могут попасть и попадают в РЖ. Кроме того, нет механизма, с помощью которого можно было бы установить, какие именно научные источники отсутствуют. Например, как правило, в РЖ не попадают данные о препринтах, предоставляющих информацию о новых направлениях научных исследований.

Отсутствие определенной формализации изложения текста реферата лимитирует широту информационных возможностей, например, не позволяет получить сведения о методах, применяемых при исследовании одной и той же проблемы, о реализации внедрения полученных результатов, о применении методов, разработанных в определенной научной области, при решении проблем другой математической области (или другой науки) и др.

Нет также механизма выявления научной проблематики различных мировых математических исследовательских центров и коллективов, работающих над одной и той же проблемой. Не гарантирована полнота изложения результатов, полученных авторами и исследователями, в силу установленных ограничений на объем реферата.

Помимо резкого сокращения объема отражаемых публикаций и снижения качества рефератов РЖ уже не отвечает в полном объеме современным запросам профессиональных математиков и

специалистов смежных областей знания, так как представленная в нем информация до недавнего времени была доступна только в печатном виде. Это затрудняло и замедляло доступ к ней и не позволяло выполнять качественный поиск, тем более ретроспективной.

К 1991 г. БД генерировались почти по всей тематике РЖ ВИНТИ, кроме тематических фрагментов “Вычислительные науки” и “Математика” [1]. Специфика РЖ “Математика” не позволяла создавать БД в используемых на начальном этапе генерации форматах СДФ и МЕКОФ, в которых не было предусмотрено описание сложных математических формул, необходимых в текстах рефератов и в заглавиях документов. В 1995–1996 гг. Производственно-издательским комбинатом (ПИК) ВИНТИ совместно с отраслевым профилирующим отделом ВИНТИ был реализован проект “Математика” [2].

В результате работ, выполненных в указанный период, а также дополнительных работ (в 2001–2002 гг.) по приведению выпусков БД “Математика” в соответствие с требованиями нормативно-технического предписания (НТП 10–2004) “Представление элементов данных во внутрисистемном формате ВИНТИ” была обеспечена возможность генерации текущих математических выпусков БД [1].

Предоставляемая в настоящее время пользователям РБД имеет глубину ретроспективы с 1997 г., в то время как большинство РБД по другим областям знания имеют ретроспективу с 1981 г. или чуть более позднюю. В настоящее время РБД “Математика” доступна в режиме on-line с помощью браузера Explorer компании Microsoft, но сложносимвольная информация при этом остается мало читабельной. Для полного отображения данных дополнительно должен использоваться TEX Explorer или любая другая программа, поддерживающая данный формат файлов [1].

Из всего вышеизложенного становится очевидным, что наступило время пересмотра существующего состояния информационной деятельности по математическим наукам и создания новой, более совершенной и более содержательной в научном и

информационном смысле системы, базирующейся в технологическом плане на современных автоматизированных информационных технологиях и располагающей более богатым банком знаний

Для решения этих задач была разработана концепция электронной информационной системы ВИНТИ “Математика”. Согласно этой концепции на базе ВИНТИ как головного института была начата разработка специальной информационно-поисковой системы по математическим наукам. Создание такой системы в полной мере может быть осуществлено только при совместной работе, основанной на кооперации нескольких научных математических центров, таких как ВИНТИ, МИАП, МГУ и др., располагающих специалистами-математиками высокой квалификации, в том числе имеющими опыт разработки систем научной информации

Известно, что качество информационной системы в первую очередь определяется спектром информационных услуг, предоставляемых пользователю. При этом имеется в виду как удовлетворение запросов самого пользователя, так и информационный сервис, предлагаемый системой.

Информационная система должна удовлетворять следующим основным требованиям:

— гарантированная полнота вводимого массива документов по объявленной тематике и данным, содержащимся в документе;

— высокая информативность рефератов и формализация изложения;

— всеаспектность отражения содержания документа как на естественном языке в реферате, так и на языке классификационных кодов и ключевых слов;

— наличие ряда дополнительных сведений: место выполнения работы, указание характера документа (оригинальная работа, обзор, персоналия и т. д.), вида документа (статья, диссертация, монография, препринт).

Внедрение современных технологий, использующих высокоорганизованную электронную технику, предоставляет новые возможности по части расширения спектра информационных услуг (быстрота поиска нужной информации, использование для поиска больших массивов документов, возможность вхождения в информационные банки данных и передачи информации по различным каналам связи и др.).

Однако автоматизацию процессов подготовки, хранения, поиска и выдачи информации можно рассматривать лишь как средство совершенствования обслуживания пользователей. Основой совершенствования обслуживания и расширения сервиса является обогащение информационного фонда данными, в первую очередь, получаемыми путем углубленной интеллектуальной обработки документа; обеспечение средств доступа к хранящейся в банке информации; возможность пополнения информации, содержащейся в базах данных новыми фактами, полученными путем применения экспертных систем и, в конечном итоге, — создание базы знаний.

Автоматизация информационных процессов должна обеспечить пользователю возможность не только получить ответ на поставленный вопрос, но и провести диалог с базой знаний. Поэтому наряду с документальными базами данных должен

быть создан ряд обслуживающих и справочных баз данных, коррелирующих с документальными базами. Все эти базы должны быть погружены в систему и формировать ее единый многоаспектный информационный фонд. При этом предполагается, что информация, предназначенная для ввода в базы данных системы, может предоставляться как на бумажных, так и на электронных носителях.

Таким образом, цели создания информационной системы (ИС) “Математика”:

А. Создание многоаспектного информационно-го фонда, содержащего:

— документальные базы первичных и вторичных информационных источников;

— обслуживающие базы, необходимые для формирования документальных баз данных (классификационные системы, многоязычные терминологические словари, списки сериальных изданий и др.);

— справочные базы (научных центров, тематических исследовательских планов и их реализации, научных конференций, регистрационной системы ученых-математиков, планов изданий математических издательств и др.)

Б. Организация оперативного и качественного информационного обслуживания индивидуальных и коллективных пользователей, которым в их деятельности необходима многоаспектная информация по спектру проблем и задач, исследованию и решению которых возможно с использованием информационных ресурсов ИС

Приведем описание содержания баз данных.

1. Документальные базы данных.

1.1 База первоисточников полнотекстовые копии документов на электронных носителях. Система обеспечивает получение копий документов в любом количестве экземпляров.

1.2. База вторичных документов содержит: библиографическое описание документа; реферат; формализованное изложение содержания работы; классификационные индексы, указывающие на область, к которой относится исследование, ключевые слова и ключевые фразы, указание метода исследования, приложений, внедрения, дополнительные сведения: место хранения оригинала и его идентификатор в базе первоисточников, место выполнения работы и др. При этом предполагается, что в эту базу данных ежемесячно должны вводиться около 30 тыс. новых документов, а также постепенно будет создаваться и ретроспективный фонд.

1.3. База цитированной литературы сведения о литературе, цитированной в любом документе, хранящемся в базе вторичных документов, а также о том, в каких документах, зарегистрированных в системе, цитируется определенная, интересующая пользователя, научная публикация

1.4. База “Итоги науки и техники. Серия математическая” предназначена для тематического поиска в многотомном и многосерийном издании “Современная математика”

2. Обслуживающие базы.

2.1 Классификационные системы, используемые в мировой практике классификации математической литературы.

2.2. Многоязычные математические терминологические словари с привязкой к классификационным системам

2.3. Серийные издания, т. е. периодические и продолжающиеся математические издания с указанием полного и сокращенного названий, тематической направленности, списка членов редколлегии и др.

3. Справочные базы данных.

3.1. База данных "Научные центры" (институты, кафедры, общества, семинары) содержит названия центров, адреса, указание тематической направленности, наименование издаваемых научных трудов и др.

3.2. База данных "Тематические планы" содержит перечень разрабатываемых плановых тем, диссертационных работ с указанием руководителей, места выполнения и другие сведения.

3.3. База данных "Регистрационная система ученых" содержит фамилии, имена, отчества, основные даты биографии, места работы, тематическую направленность исследований, списки научных работ, защитившихся аспирантов, научных наград и др.

3.4. База данных "Научные форумы" содержит наименование мероприятий, сведения об организаторах, числе участников, месте проведения, подготовке к изданию трудов и др.

3.5. База данных "Издательские планы" предназначена для обеспечения полноты документальной базы, а также для информирования пользователей.

Система предусматривает возможность расширения числа баз данных и организации взаимодействия между ними.

Сервис, предоставляемый системой.

ИС предназначена для оперативного обслуживания своих постоянных пользователей, обращающихся с запросами. Система может выдавать любую информацию, содержащуюся в ее информационном фонде. Поиск информации осуществляется в режиме диалога пользователя с системой.

Обращение к системе происходит на естественном языке предметной области. Система обеспечивает "дружелюбное" отношение к пользователю и рассчитана на доступ к информационному фонду пользователей с разным уровнем подготовки.

В качестве рабочего естественного языка первоначально будет принят русский язык, а затем – английский и другие языки. Английский язык будет использоваться как язык-посредник для дальнейшего расширения набора рабочих языков.

Сведения могут выдаваться в любой заказанной компоновке информационных данных, содержащихся в системе.

Пользователь может стать постоянным абонентом издания, скомпонованного по его личному заказу (избирательное распространение информации

(ИРИ)). В зависимости от набора заказанных данных (информационных полей) такие издания могут носить и характер экспресс-оповещения. Например, литература, поступившая за определенный период в библиотеки страны по данной теме; подборка литературы, опубликованной определенным издательством за некоторый период; список научных центров, ведущих исследования по данной тематике; список трудов определенного ученого, опубликованных за указанный период; имена (и другие данные) ученых, работающих по определенной теме, т. е. образующих "незримые коллективы" и т. д.

Система обеспечивает получение информационных изданий в электронном виде, допускающем тиражирование этих изданий как на электронных, так и на бумажных носителях. Так, могут компоноваться и выдаваться на электронных носителях:

– реферативный журнал (русская версия, в перспективе и английская версия);

– сигнальная информация (библиографический сборник тематически упорядоченных библиографических описаний новых документов, поступивших в систему);

– указатели к реферативному журналу (к каждому номеру реферативного журнала), а также кумулятивные годовые указатели (авторские, систематические, указатели журналов, конференций, монографий, персоналий и т. д.)

В режиме диалога, используя также экспертную систему, пользователь может последовательно устанавливать интересующую его многоплановую информацию. Эта форма сервиса основана на том, что все базы данных, включенные в систему (и те, которые могут быть включены в дальнейшем), образуют Единый информационный фонд.

Математическое обеспечение системы будет давать возможность выдавать комплексный ответ на запрос, который компоуется путем обращения к разным базам данных. Банк данных обеспечивает взаимодействие всех баз, погруженных в Единый информационный фонд.

СПИСОК ЛИТЕРАТУРЫ

1. Арский Ю. М., Леонтьева Т. М., Никольская И. Ю., Шогин А. П. Банк данных ВИНТИ Состояние и перспективы развития. М.: ВИНТИ, 2006

2. Остиану П. М., Панюта Ю. И., Пробст М. А. и др. Новая база данных ВИНТИ – Математика и Вычислительные науки. НТИ-96 // Международная конф. "Информационные продукты и технологии", 20-21 ноября 1996 г., Москва. – М.: ВИНТИ, 1996. С. 184-185.

Материал поступил в редакцию 18.09.06