

**О ПРЕДМЕТНЫХ ИНФОРМАЦИОННЫХ СИСТЕМАХ
НАЦИОНАЛЬНОГО МАСШТАБА
(НА ПРИМЕРЕ ИНФОРМАЦИОННОЙ СИСТЕМЫ
ДЛЯ ФИЗИКОВ РОССИИ ИРИД)**

А.П. Ковалева¹, С.А. Крашаков², Л.В. Щур³, Л.Н. Щур²

¹Отделение физических наук РАН,

²Институт теоретической физики им. Л.Д. Ландау РАН,
142432 Московская обл., Ногинский р-н, г. Черноголовка, Россия

³Институт физики твердого тела РАН

Emails: alina@gpad.ac.ru, sakr@chg.ru, lvs@chg.ru, shchur@chg.ru

**ON THE INFORMATION SYSTEMS FOR THE NATIONAL
SCIENTIFIC SOCIETIES
(WITH THE EXAMPLE, INFORMATION SYSTEM
OF RUSSIAN PHYSICAL SOCIETY IRID)**

A.P. Kovaleva¹, S.A. Krashakov², L.V. Shchur³, L.N. Shchur²

¹Department of Physical Sciences RAS,

²Landau Institute for Theoretical Physics RAS,
Chernogolovka, Moscow region 142432 Russia

³Institute of Solid State Physics RAS

Emails: alina@gpad.ac.ru, sakr@chg.ru, lvs@chg.ru, shchur@chg.ru

We discuss how to create, develop and maintain information system of national scientific society. Our analysis is based on the experience with the creation of national wide networks in Russia and CEEnet society. We present some particular realisations of our ideas we have with the IRiD.

Введение

Создание информационных систем необходимо для поддержки фундаментальных и прикладных исследований в каждой из предметных областей науки. Назначение этих систем достаточно широко и определяется насущными задачами.

К числу основных задач можно отнести, например, предоставление информации о конференциях. Информация о конференциях доступна ведущим специалистам по данной области непосредственно от организаторов, и у них может создаться впечатление, что информационная система не требуется. Однако это не так. Во-первых, конференции важны также и для молодых специалистов, имена которых, как правило, не известны организаторам. Как показывает практика, молодые специалисты узнают о проводимых мероприятиях случайно. Во-вторых, в наше время появляется все

больше междисциплинарных конференций, информация о которых не может быть получена всеми специалистами в каждой из областей, по которой проводится конференция. Кроме того, часто конференции по смежным областям могут стать интересными для более широкого круга исследователей.

В докладе обсуждаются проблемы, которые возникают (или могут возникнуть) при создании, развитии и сопровождении информационных систем по предметным областям национального масштаба. При этом мы использовали опыт создания компьютерных сетей для науки и образования как в России, так и в странах Центральной и Восточной Европы. Этот опыт суммирован в недавнем докладе [1] Питера Раствла, директора компьютерного центра Венского Университета и создателя научной сети Австрии. Мы в значительной степени используем эту работу.

Для конкретизации мы будем говорить о создании информационной системы для физиков Российской федерации. Такая система ИРиД (Информационные Ресурсы и Данные) создается под эгидой Объединенного физического общества РФ и Отделения физических наук РАН [2]. Работа координируется Европейским физическим обществом в рамках создания национальных информационных систем по физике, и их объединения в единую информационную систему Европы. Хотя мы будем иметь в виду создание ИРиД, тем не менее, проблемы и задачи, которые мы рассматриваем, являются достаточно общими и применимыми и для других научных сообществ.

Проблемы доступа к информации и ее поиска

В настоящее время, после этапа энтузиазма и восхищения возможностями быстрого доступа к сети Интернет стало понятно, что существует много сложностей по получению интересующей информации имеющимися средствами. Причин для этого несколько. Первая, это присущий сети Интернет динамизм: информация весьма динамична, как по содержанию, так и по месту ее происхождения (хранения). Вторая причина состоит в отсутствии систем для целенаправленного поиска профессионально ориентированной информации. Количество ссылок, генерируемых поисковыми системами общего пользования на один и тот же запрос практически утрачивается каждый год. Совершенно очевидно, что классические поисковые системы типа Google, Rambler и т.п. не в состоянии находить информацию в таких узкоспециализированных областях, как наука, и физика, в частности. В то же время поиск по имеющимся спискам ссылок на все физические институты в мире также мало продуктивен. И, наконец, централизованно создаваемые информационные системы (примером такой системы может служить ИСИР - Интегрированная Система Информационных Ресурсов РАН [3]) не в состоянии адекватно отражать все изменения и, не-

смотря на декларируемую динамичность, они по сути своей статичны и зачастую выдают уже устаревшую информацию.

Заметим, что затраты на наполнение и поддержание таких информационных систем значительно выше затрат на их создание. Ясно, что при экспоненциальном росте объема информации эта составляющая будет продолжать расти. Это означает, что необходимы специализированные системы для автоматизированного поиска, обработки, хранения и предоставления информации. Следует также учитывать и юридический аспект – авторские права на полнотекстовую информацию. Чаще всего, по формальным причинам, она не может быть скопирована на другой сервер.

Один из путей решения этой задачи состоит в создании предметно-ориентированных логических сетей, примером которых может служить распределенная информационно-поисковая система PhysNet, создаваемая в Ольденбургском университете под эгидой Европейского физического общества и при участии национальных физических обществ [4,5]. При этом сбор и поиск информации производится специальными системами и на вполне определенном и четко ограниченном пространстве в Интернете. Эффективность системы значительно возрастает, если при создании Web-страниц использовать стандартные форматы описания ресурсов (RDF), например, Дублинский мета-код [6]. Иными словами, мы создаем систему для интеграции распределенных ресурсов, для оптимизации потоков информации и для ее автоматизированной обработки.

Основные вопросы

При создании системы важно иметь ответы на следующие основные вопросы:

1. Что такое Информационная Система Научного Общества (ИСНО) для физиков (математиков, химиков, историков, ...)?
2. Кто финансирует ИСНО (ее создание, ее развитие, ее сопровождение)?
3. Какие сервисы предоставляет ИСНО?
4. Кто пользователи этой системы?
5. Как организовать обратную связь с пользователями системы?

Основные аспекты ИСНО

При планировании большой информационной системы необходимо определить пути реализации ее основных аспектов.

К числу основных аспектов ИСНО могут быть отнесены:

1. Организационная модель.

Это может быть инициативный проект, подаваемый в государственный, частный или международный научный фонд. Возможен проект по кооперации усилий организаций-членов проекта.

2. Модель финансирования.

Необходимо заранее определить источники финансирования (гранты фондов, взносы организаций-участников). Важно понимать происхождение возможных расходов и способы покрытия конкретных расходов из различных источников. Наше время также диктует необходимость учета и возможной коммерческой деятельности по мере выполнения проекта, хотя в этом направлении опыт научных организаций достаточно ограничен.

IRiD (ИРиД), История и задачи

В марте 1998 было учреждено Объединенное физическое общество Российской Федерации. В настоящее время ОФО РФ состоит из 47 региональных отделений и двух коллективных членов. Установлены контакты с Европейским физическим обществом, Американским физическим обществом, с физическими обществами ряда других стран, включая страны СНГ. Одной из важнейших задач является создание единого информационного пространства с использованием новых информационных технологий, и это определяет одно из основных направлений деятельности ОФО РФ на ближайшее время. Аналогичную задачу ставит перед собой и Европейское физическое общество, являющееся организатором создания информационной системы PhysNet, в создании которой принимают национальные физические общества ряда стран и ряд физических институтов.

В сентябре 2000 г. ОФО РФ и ООФА РАН совместно с НЦЧ РАН была подана заявка в РФФИ, которая была принята как проект 01-07-90395 “Информационные ресурсы для физиков ИРиД” (Информационные Ресурсы и Данные). Проект предусматривает создание системы для поиска, обработки, хранения и предоставления информации, необходимой для проведения научных исследований в области физики. Система IRiD видится как технически децентрализованная система серверов научных учреждений РФ, занимающихся исследованиями в области физики. При этом организационно ведущая роль отводится серверам Отделения физики РАН и Объединенного физического общества (ОФО РФ), как координаторам и организаторам исследований в области физики в целом. Система не ограничивается предоставлением только научной информации в области физики, или только организационной. Система видится как интегральное целое для удовлетворения потребностей научных работников и молодежи во всех аспектах научной деятельности.

3. Техническая модель.

При реализации проекта важно проанализировать технические модели реализации проекта на различных этапах и пути плавной, постепенной модификации технических решений. От одного центра на начальном

этапе к распределенной системе, а, возможно, и специально организованной сети серверов.

Некоторые детали технической реализации ИРИД.

При разработке ИРИД учитывались основные принципы создания PhysNet и, прежде всего, распределенность, децентрализованность, открытость архитектуры и программного обеспечения, зафиксированные в Хартии PhysNet [5]. Технически предполагается создание распределенной сети сборщиков информации (gatherer's) и обработчиков информации (broker's). Назначение сборщиков информации - создание индексных файлов, содержащих указатели на Web-страницы. Файлы создаются путем поиска на Web-страницах по заранее определенному списку Web-узлов. При этом большое внимание уделяется метаданным, содержащимся в документе, на основании которых, в основном, в дальнейшем и будет производиться поиск. Далее, по согласованному расписанию происходит обмен индексными файлами с остальными участниками распределенной информационной системы. Тем самым, центральный сервер (например, PhysNet, расположенный в Ольденбурге) не должен опрашивать серверы по всему миру. Это приводит к значительной экономии ресурсов, как центрального сервера, так и коммуникационных. Кроме того, автоматически решается проблема поддержания целостности зеркал. Назначение обработчиков запросов - предоставление пользователю доступа к необходимой информации путем поиска в индексных файлах. Предполагается равномерное распределение сборщиков и обработчиков по научным центрам и университетам. Для ускорения обработки запросов предполагается также обмен индексными файлами между физическими обществами, институтами и факультетами. В качестве программного обеспечения сборщиков и обработчиков информации выбрана свободнораспространяемая система Harvest [7] под управлением ОС FreeBSD.

Сервисы ИСНО

Возможные сервисы, которые ИСНО может предоставлять пользователям:

1. Сервис для организаций-членов национального общества
2. Сервисы, которые координируются на национальном уровне (архивы, зеркальные серверы, поисковые системы, кэш-серверы, вэб-порталы, информационное наполнение)
3. Сервисы для конечного пользователя (= для члена национального общества)
4. Информация о конференциях и справочно-навигационная система
5. Поддержка пользователей (система справок и помощи, консультационные сервисы)

6. Тренировка и обучение
7. Внедрение новых технологий
8. Исследование информационных систем и их развитие

Полагаем целесообразным организовать в последующих конференциях секцию по созданию и развитию предметно-ориентированных информационных систем национального масштаба. Было бы желательно организовать рабочую группу, которая проводила бы параллельный (или сателлитный) рабочий семинар и, возможно, школу по вопросам ИСНО. Рабочая группа может стать естественным проводником новых технологий и ядром по формулировке приоритетных исследований в этом направлении.

Категории пользователей

Пользователей системы можно разделить на такие категории:

1. Исследовательские институты, и университеты (организации-члены общества) и их соответствующие подразделения по информационным системам, библиотеки, отделы Ученого секретаря, ...
2. Индивидуальные пользователи (члены общества); индивидуальные представители сообществ пользователей, ...
3. Школы: учителя, школьники, их родители
4. Граждане общества

Конечно, для охвата такого широкого круга пользователей необходимы существенные материальные затраты, как финансовых средств, так и рабочих рук. Представляется целесообразным, что финансирование подразделов будет осуществляться заинтересованными в этом конкретном подразделе институтами нашего общества, включая федеральные и местные структуры, министерства и ведомства. Это вполне может быть сделано при координации с программой «Электронная Россия».

Обратная связь с пользователями

Важным фактором успешного функционирования системы национального масштаба является эффективность организации связи с пользователями системы. Почему нужно прислушиваться к мнению пользователей информационной системы?

1. В конечном счете, ИСНО создается именно для пользователя
2. Фактически, пользователи финансируют создание ИСНО
3. Нужно быть уверенным, что мы понимаем нужды членов общества
4. Нужно быть уверенным, что наши приоритеты совпадают с ожиданиями пользователей
5. Различные категории пользователей могут иметь разные требования

6. Необходимо поддерживать осведомленность пользователей о ходе проекта

7. Нужно быть уверенным, что пользователи понимают наши проблемы

Возможные пути стимулирования участия пользователей можно осуществлять посредством организации и проведения следующих мероприятий:

1. Технические семинары, встречи по обсуждению развития
2. Конференции пользователей, обучающие семинары
3. ИСНО вэб-сайт, Годовой отчет, Электронный журнал
4. Контактные лица, раздел помощи
5. Специальные сервисы

Начало реализации системы ИРИД

В качестве первого шага, в январе 2001 было установлено зеркало PhysNet в Научном центре в Черноголовке (<http://www.chg.ru/PhysNet/>) [2]. Следующим шагом явился запуск пилотной версии системы обработки информации с первичных Web-серверов (пока только по организациям ИЦЧ РАН). Согласован формат и периодичность обмена данными с национальными физическими обществами ЕФО и центральным сервером PhysNet. В 3 квартале 2002 планируется установить сервер в Отделении физических наук РАН и открыть его для использования.

Основные проблемы - это быстрое устаревание информации, содержащейся на сайтах институтов, и практически полное отсутствие метаданных, позволяющих более целенаправленно вести поиск. Поэтому в настоящее время ОФО РФ и ОФН РАН ведут работу по разработке унифицированной системы предоставления материалов (на основе Дублинского метакода) – персоналии, научные советы, конференции, диссертации, статьи, проекты и т.д., и разработка методических рекомендаций для их размещения на серверах институтов и учреждений ОФО РАН.

Заключение

Чтобы быть уверенным в правильности пути, надо задавать себе такие вопросы, как:

1. Используете ли Вы сами те сервисы, которые предоставляете рядовому пользователю, без каких-либо привилегий?
2. Понимают ли пользователи, что их обслуживание стоит реальных денег?
3. Имеют ли Ваши пользователи какие-либо альтернативы вашему «монопольному» сервису? Если да, то могут ли они получить что-то большее от использования альтернативных систем?

Литература

- [1]. Peter Rastl, NRENs and their user community, CEENet Workshop, Issyk-Kul, Kyrgyzstan, 20-25 September 2001.
- [2]. S.A. Krashakov and L.N. Shchur, Russia mirror of PhysNet: one year experience and some statistics, SINN01, Oldenburg, December, 2001.
- [3]. Интегрированная система информационных ресурсов (ИСИР) РАН, <http://isir.ras.ru/>
- [4]. PhysNet: The Physics Departments and Document Network, <http://physics-network.org/Physnet/>
- [5]. Physics Network (PhysNet) Charter, <http://www.physics-network.org/PhysNet/charter.html>
- [6]. Dublin Core Metadata Initiative, <http://dublincore.org/>
- [7]. Harvest: A Distributed Search System, <http://harvest.sourceforge.net/>