

Институт вычислительных технологий СО РАН
пр. Академика Лаврентьева, 6, Новосибирск, 630090, Россия

Новосибирский государственный университет
ул. Пирогова, 2, Новосибирск, 630090, Россия
E-mail: ¹ bar@ict.nsc.ru; ² fedotov@nsu.ru; ² fedotov@sbras.ru

ИССЛЕДОВАНИЕ ИНФОРМАЦИОННЫХ ПОТРЕБНОСТЕЙ НАУЧНОГО СООБЩЕСТВА ДЛЯ ПОСТРОЕНИЯ ИНФОРМАЦИОННОЙ МОДЕЛИ ОПИСАНИЯ ЕГО ДЕЯТЕЛЬНОСТИ *

В работе проведено исследование информационных потребностей научного сообщества, основанное на известной методике «5W+ 1H» и на анализе тематики совместных информационных проектов коллективных пользователей – организаций СО РАН. На основе этого исследования проведено обоснование информационной модели деятельности научного сообщества, которая (модель – В. Б., А. Ф.) легла в основу модели, описывающей программные системы информационного обеспечения научной деятельности, предназначенные для включения в научно-информационный процесс слабоструктурированных интернет-документов.

Ключевые слова: информационные потребности, информационная модель описания деятельности научного сообщества.

Введение

Бурное развитие высоких технологий в области передачи и обработки информации, в частности создание современных телекоммуникационных систем (прежде всего сети Интернет), привело к появлению принципиально новых возможностей организации практически всех этапов научно-информационного процесса, что, в свою очередь, обусловило качественный рост информационных потребностей научного сообщества. Кроме того, за указанный период в России произошла смена общественно-экономической формации, приведшая к изменению принципов функционирования и финансирования науки, что также не могло не сказаться на характере информационных потребностей ученых-исследователей.

Наиболее перспективным направлением развития информационного обеспечения научной деятельности являются электронные информационные технологии. В настоящее время научные сообщества наиболее развитых стран и регионов мира обладают достаточно мощными информационными системами, которые в той или иной степени удовлетворяют потребностям исследователей в информации, однако основными недостатками большинства систем являются несвоевременная актуализация информации (этот недостаток не относится к библиотечным системам) и ограниченность возможностей обеспечения интеграции ресурсов как внутри каждой из систем, так и с внешними системами (другими словами, низкая интероперабельность).

К тому же ограниченность возможностей классических информационно-поисковых систем во многом обусловлена тем обстоятельством, что в настоящее время наука об обработке данных, особенно в ее прикладном аспекте, несколько отстает от соответствующих аппаратно-программных средств. Аналогичное отставание наблюдается и в прикладных исследованиях, посвященных извлечению из информации знаний – адекватного отражения действительности в сознании человека в виде представлений, понятия, суждений, теорий.

Преодоление указанных проблем возможно путем создания «интеллектуальных» *информационных систем*, в качестве составных компонент которых являются, наряду с традиционной информационной системой, еще и «рассуждающая» информационная система (формали-

* Работа выполнена при частичной финансовой поддержке РФФИ (проекты № 06-07-89060, 06-07-89038, 07-07-00271, 08-07-00229), президентской программы «Ведущие научные школы РФ» (грант № НШ-931.2008.9) и интеграционных проектов СО РАН.

зующая правила логического вывода) и интеллектуальный интерфейс (диалог, графика и т. д.), благодаря которым компьютер в диалоговом режиме усиливает алгоритмическое мышление и логические возможности человека.

Развитие сети Интернет предоставило создателям интеллектуальных информационных систем новые возможности, связанные с одновременным доступом ко множеству разнородных источников данных, что открывает широкие перспективы в развитии более совершенных технологий получения знаний. Однако многие современные исследования в области интеллектуального поиска опираются на неявное предположение о возможности широкого распространения более или менее подробной стандартизации представления информации. Разумеется, реализация подобных проектов, прежде всего концепции Semantic Web¹ консорциума W3, позволила бы вывести работу с информацией на качественно новый уровень, но одна из основных особенностей сети Интернет как феномена цивилизации заключается в том, что развитие сети изначально носит децентрализованный характер, поэтому многие информационные ресурсы, содержащие важную информацию из той или иной предметной области, не соответствуют рекомендациям консорциума W3. В частности, на большинстве сайтов документы являются *слабоструктурированными*, т. е. значения атрибутов их метаданных, если они есть, носят достаточно произвольный характер, а не являются элементами заданных словарей (это относится как к содержательным, так и к структурным метаданным). Более того, реальная картина развития сети Интернет такова, что даже в рамках крупной научной корпорации не удастся административно утвердить единые стандарты хранения и представления информации.

Разработка интеллектуальных систем информационного обеспечения научной деятельности, интегрирующих разнородные информационные ресурсы, сталкивается со следующей проблемой: созданный еще в досетевую эпоху мощный аппарат для манипулирования *данными* недостаточно приспособлен для работы с *информацией* и *знаниями*, а результаты последних теоретических исследований в области автоматического извлечения знаний не позволяют обрабатывать огромные пласты информации, доступной через сеть Интернет.

Необходимость разработки и теоретического обоснования моделей, описывающих системы информационного обеспечения научной деятельности, предназначенные для включения в научно-информационный процесс слабоструктурированных информационных ресурсов (документов), очевидна. В основе этих моделей должна лежать информационная модель описания деятельности научного сообщества, учитывающая современный уровень развития информационных потребностей исследователей. Исследованию информационных потребностей научного сообщества для построения информационной модели описания его деятельности и посвящена данная работа.

Основные характеристики информационных потребностей в сфере науки

Одним из наиболее распространенных способов исследования информационных потребностей является пришедшая из англоязычной литературы формула «5W + 1H», заключающаяся в постановке шести вопросов, которые предъявляются лицу, запросившему информацию, но, очевидно, могут быть поставлены и априорно при анализе информационных потребностей некоторой социальной группы [Арский и др., 1996]. Вот как эта формула, может быть адаптирована к сфере науки:

– WHERE (Где?) – где работает потенциальный потребитель информации – в научно-исследовательском учреждении, на предприятии, в правительственном ведомстве и т. д.?

– WHAT (Какая?) – какая информация может представлять интерес – об окружающей среде, о конкурентах, внутренняя? Какого типа – в исходном виде, ретроспективная, текущая, прогностическая?

– WHO (Кто?) – кто может выступать в качестве потенциального потребителя информации – исследователь, специалист (инженер, агроном и т. п.), управленец?

– WHY (Зачем?) – зачем требуется информация – для научного исследования, разработки, изучения, планирования или управления?

¹ <http://www.w3.org/2001/sw>

– WHEN (Когда?) – когда требуется информация – немедленно, регулярно, по мере необходимости?

– HOW (Как?) – в каких видах и в какой форме требуется информация – в виде оригиналов, в машиночитаемом виде, подвергнутая анализу или другой обработке?

Дав ответы на эти вопросы, мы сможем описать принципы организации информационного обеспечения научной (включая научно-организационную, инновационную и т. п.) деятельности. Следует отметить, что в постановке некоторых вопросов, например «Где?», «Кто?» и «Зачем?», наблюдается определенное сходство, поэтому ответы на них в значительной мере коррелируют (конкретные особенности корреляции будут описаны ниже).

Разумеется, наиболее важными и сложными являются вопросы «Какая?» (особенно в части, касающейся типа предоставляемой информации) и «Как?». Ответы на них, зависящие от вариантов ответов на вопросы «Где?», «Кто?» и «Зачем?», подразумевают, в том числе, подробное описание методов реализации предложенных решений.

При ответе на вопрос «Где?» необходимо учитывать, что информационное обеспечение научной деятельности предполагает не только удовлетворение специальных информационных потребностей научного сообщества, но и предоставление информации работникам других отраслей экономики и управления, могущих сотрудничать с научными учреждениями, ибо такое сотрудничество способствует, в конечном счете, повышению эффективности научной деятельности. Здесь речь идет, прежде всего, о промышленных предприятиях и различных финансовых институтах, заинтересованных в получении информации об инновационных разработках с целью их внедрения или коммерциализации, а также об органах государственной власти и местного самоуправления, которые заинтересованы в получении разнообразной информации научного и научно-организационного характера для принятия соответствующих управленческих решений. Способы удовлетворения соответствующих информационных потребностей будут рассмотрены при ответах на вопросы «Кто?», «Зачем?» и «Какая?».

В постановке вопроса «Кто?» наблюдается заметная корреляция с вопросом «Где?» (хотя и неполная, поскольку в организациях каждого из типов, перечисленных в вопросе «Где?», могут работать разные категории потребителей информации, причем информационные потребности, даже применительно к научно-технической информации, например, управленцев, представляющих науку, промышленность и органы государственной власти, несколько различны). Эта корреляция позволила разделить информационные потребности в сфере науки и техники на три основных вида:

- информационные потребности исследователей;
- информационные потребности специалистов;
- информационные потребности управляющих (руководителей).

Основные различия в выражении и удовлетворении информационных потребностей указанных групп потребителей информации сведены в таблицу 1.

Представленные в таблице характеристики «Четкость осознания» и «Полнота информации» призваны ответить на вопрос «Какая?», характеристика «Срочность» – на вопрос «Когда?», а характеристики «Форма получения», «Степень переработки» и «Виды предпочтительной информации» – на вопрос «Как?».

Нельзя не отметить, что столь резкое разграничение информационных потребностей исследователей и аналогичных потребностей специалистов и управленцев (которое еще сильнее выражено в [Михайлов и др., 1976]), более характерно для общественно-экономических условий 1960-х – 1980-х гг., когда наука и техника рассматривались как два существенно различающихся вида общественной деятельности. За последние 10–15 лет ситуация значительно изменилась. С одной стороны, благодаря господству высоких технологий размывается грань между наукой и производством. С другой стороны, в России произошла смена общественно-экономической формации, приведшая к изменению принципов функционирования и финансирования науки, вследствие чего ученые, используя результаты своих фундаментальных исследований, стали более активно заниматься опытно-конструкторскими работами и даже непосредственным производством уникальных наукоемких изделий. Кроме того, выполнение проектов Федерального агентства по науке и инновациям, грантов РФФИ и т. п. требует управленческих навыков, включая использование управленческой информации, не только от руководства научных учреждений, но и непосредственно от исследователей.

Таблица 1

Основные различия информационных потребностей ученых-исследователей, специалистов и управляющих*

Характеристики выражения и удовлетворения информационных потребностей	Исследователи	Специалисты	Управляющие
Четкость осознания и выражения	небольшая	большая	очень большая
Требуемая полнота информации	большая	не больше, чем нужно	самая нужная
Срочность удовлетворения	не важна	важна	очень важна
Форма получения информации	любая	удобная для использования	максимально удобная для восприятия
Степень переработки первичной информации	минимальная	большая	очень большая
Виды предпочтительной информации	первоисточники, библиография, численные данные	фактографическая информация	обзорно-аналитическая информация

* Таблица взята из: [Арский и др., 1996]

В частности, в работе [Баракнин, Федотов, 2004] на основании анализа результатов социологических опросов молодых ученых СО РАН [Гвоздева, Высоцкий, 2004] показано, что молодые исследователи, являющиеся наиболее активными пользователями сети Интернет, испытывают насущную потребность в разнообразной научно-организационной и управленческой информации.

Из сказанного выше следует, что ответ на вопрос «Зачем?» для специалистов и управляющих более или менее очевиден: информация им нужна соответственно для продвижения разработок и для осуществления планирования и управления. Что же касается ученых, то им информация требуется как для научных исследований, так и для выполнения функций разработчиков и управленцев.

Итак, можно констатировать, что табл. 1, приведенная в [Арский и др., 1996], в настоящее время применительно к исследователям отвечает на вопрос «Зачем?» в соответствии с тем аспектом деятельности научных работников, для осуществления которого требуется данная информация. К тому же некоторые положения таблицы нуждаются в уточнении. Так, утверждение, что «четкость осознания и выражения информационных потребностей ученых – небольшая», конечно же верно применительно к процессу научного поиска, но не совсем применимо к повседневому труду исследователя. Например, в [Пойа, 1959] показано, что процесс решения математической задачи начинается с распознавания ее элементов с использованием определений, т. е. в данном случае имеется четко выраженная информационная потребность. Аналогичная ситуация возникает, когда ученый ищет конкретную публикацию и т. п.

Таким образом, как отмечено в [Михайлов и др., 1976], существуют два типа информационных потребностей:

- 1) потребности в сведениях об источниках необходимой научной информации;
- 2) потребности в самой необходимой научной информации.

При этом ученым-исследователям свойственны оба вида информационных потребностей.

Однако наибольшее влияние современные высокие технологии оказали на решение вопроса «Когда?». Развитие телекоммуникационных систем, прежде всего, сети Интернет, привело к появлению принципиальной возможности практически немедленного удовлетворения возникающих информационных потребностей, а также значительно упростило регулярное

предоставление пользователю периодически обновляемой информации, например, посредством рассылки электронных почтовых сообщений.

Как же быть с приведенным в табл. 1 утверждением о том, что для ученого срочность удовлетворения информации не важна (по крайней мере когда речь идет об информации, необходимой непосредственно для поведения научных исследований)? Очевидно, что при прочих равных условиях предпочтительно скорейшее удовлетворение возникшей информационной (как, впрочем, и любой другой) потребности, поэтому вывод о «несрочности» потребности в научной информации был продиктован прежде всего экономическими соображениями. Это было вполне оправдано в тех условиях, когда отсутствовала возможность немедленного удовлетворения информационных потребностей, и различие между «срочным» и «несрочным» предоставлением информации носило сугубо количественный характер. В настоящее время немедленное удовлетворение информационных потребностей по-прежнему более затратно, чем предоставление информации по мере необходимости (например, создание веб-сайта научного журнала с полными текстами статей обходится дороже, чем размещение на сайте одних только аннотаций, предполагающее получение пользователем твердых копий заинтересовавших его публикаций по обычным библиотечным каналам). Однако здесь вступают в дело психологические соображения.

Психологическое влияние своевременного удовлетворения информационных потребностей на производительность труда было отмечено еще в начале 1920-х гг. А. А. Богдановым (предвосхитившим в своей «Тектологии» не только общую теорию систем, но и некоторые основные принципы кибернетики): «...если грамотный, культурный рабочий лишается привычной уже для него газеты, чтения брошюр, книг, то падает его “настроение” и опять понижается рабочая сила», вследствие чего обстоятельства, «вынуждающие к сокращению таких потребностей, могут иметь серьезное значение для работоспособности» [Богданов, 1989]. Подчеркнем, что в процитированном фрагменте речь идет о влиянии на работоспособность *текущих* информационных потребностей (т. е. непосредственно не связанных с производственным процессом данного работника). Разумеется, неудовлетворенные *конкретные (специальные)* потребности (т. е. потребности в профессиональной информации), появившиеся с развитием сети Интернет, в частности невозможность оперативного получения нужной информации, тем более могут оказать негативное влияние на работоспособность. Таким образом, необходимо максимально срочное удовлетворение информационных потребностей научных работников.

Сказанное ни в коей мере не означает, что классические способы удовлетворения информационных потребностей посредством получения информации на бумажных носителях, общения на конференциях и т. п. ушли в прошлое, однако наиболее перспективным направлением развития информационного обеспечения научной деятельности являются все-таки электронные информационные технологии. Далее мы будем вести речь только о тех способах удовлетворения информационных потребностей научного сообщества, которые базируются на электронных технологиях. В рамках указанного подхода основным инструментом информационного обеспечения научной деятельности являются *информационные системы* (взгляд на это понятие представлен в [Барахнин, Федотов, 2004]). В настоящее время в подавляющем большинстве случаев подразумевается, что информационная система обладает удаленным доступом через сеть Интернет. Таким образом, возможность срочного удовлетворения информационных потребностей научного сообщества зависит от оперативности появления научной информации в сети Интернет, а также регистрации ее в каталоге той или иной информационной системы, ибо в противном случае соответствующий информационный ресурс практически не имеет шансов стать достоянием широкой научной общественности.

Исходя из сформулированных выше положений приступим к ответу на основные вопросы, характеризующие информационные потребности: «Какая?» и «Как?».

Исследование информационных потребностей коллективных пользователей – научных учреждений СО РАН

Перейдем к более конкретному анализу потребностей научного сообщества в информации, распространяемой с использованием электронных технологий. Изучение информацион-

ных потребностей ученых осложняется тем, что эти потребности зависят от множества разных факторов и носят в значительной мере персонифицированный характер [Арский и др., 1996]. При этом одним из более или менее объективных методов определения информационных потребностей является построение картины фактического использования исследователями разных видов источников информации, а также собственная оценка учеными относительной важности этих источников.

В качестве объекта исследования рассмотрим «внутренние» информационные потребности Сибирского отделения РАН (под «внутренними» мы подразумеваем потребности ученых СО РАН в научной информации из источников самого СО РАН). Репрезентативность подобной выборки доказывается следующими фактами.

Сибирское отделение РАН – это расположенные на территории трех федеральных округов почти 100 научно-исследовательских и конструкторско-технологических учреждений, в которых работает более 20 тысяч человек, в том числе (по состоянию на 1 января 2007 г.) 8 725 научных сотрудников, из них 126 членов РАН, 1 744 доктора и 4 748 кандидатов наук². Для удовлетворения информационных потребностей Сибирского отделения создана Сеть передачи данных СО РАН³, в которой зарегистрировано около 150 организаций-абонентов. Только в Новосибирске Сеть обслуживает более 40 000 пользователей и насчитывает свыше 12 000 подключенных компьютеров. Кроме того, в региональных научных центрах Отделения находится еще около 30 000 пользователей. Суммарный объем информации, получаемой и отправляемой по каналам Сети, составляет более 700 Гбайт в сутки, при этом 58 % общего объема составляет информация, получаемая абонентами из Сети, а 42 % – передаваемая ими во внешний мир.

Высокий уровень информатизации СО РАН (и, следовательно, развитые информационные потребности работающих в нем ученых) подтверждает рейтинг Webometrics Кибернетической лаборатории Национального исследовательского совета Испании. В этот рейтинг входят 500 ведущих сайтов университетов и научно-исследовательских центров всего мира, при его подсчете основное значение имеет число размещенных на сайте научных работ и количество ссылок на них. По состоянию на июль 2008 г. сайт Сибирского отделения РАН занимал наивысшее среди российских сайтов 54-е место (18-е в Европе)⁴.

К сожалению, построение картины фактического использования учеными СО РАН различных источников «внутренней» информации пока не осуществлено, поскольку система мониторинга Сети и сбора статистики передачи данных СО РАН [Шокин и др., 2007] создана совсем недавно и еще не накопила достаточного количества данных для проведения соответствующего анализа.

С другой стороны, имеются данные, позволяющие исследовать оценку научным сообществом СО РАН сравнительной важности тех или иных форм удовлетворения информационных потребностей коллективных пользователей. В качестве таких данных мы рассматриваем итоги трех конкурсов интеграционных проектов СО РАН, проводимых раз в 3 года. Проекты-победители определялись Постановлениям Президиума СО РАН⁵. По итогам конкурса 2000 г. победителями было признано 88 проектов, в 2003 г. – 180 проектов, в 2006 г. – 247 проектов. В каждом проекте принимали участие несколько институтов СО РАН (а иногда и других научных организаций), притом институты – участники проекта, представляли, как правило, разные направления наук. Описанная процедура проведения конкурсов позволяет сделать вывод о том, что проекты, так или иначе связанные с информатикой, адекватно отражают информационные потребности коллективных пользователей – научных учреждений СО РАН.

² Визитная карточка Сибирского отделения Российской академии наук (<http://www.sbras.ru/cmnn/general.html>).

³ Сеть передачи данных Сибирского отделения РАН (<http://www.ac-tel.ru/mw/index.php/Введение>).

⁴ Top 100 R&D European Institutes (http://www.webometrics.info/top100_r&d_europe.asp).

⁵ Постановление Президиума СО РАН от 13.04.2000 № 137 «Об итогах конкурса интеграционных программ (проектов) СО РАН – 2000» (<http://www.sbras.ru/win/anonses/373.html>); от 21.02.2003 № 62 «Об итогах конкурса интеграционных проектов СО РАН – 2003 г.» (<http://www.sbras.ru/win/anonses/841.html>); от 26.01.2006 № 32 «Об интеграционных проектах, выполняемых по заказу Президиума СО РАН» (<http://www.sbras.ru/win/anonses/1334.html>); от 09.02.2006 № 54 «Об итогах конкурса комплексных интеграционных проектов СО РАН – 2006» (<http://www.sbras.ru/win/anonses/1341.html>); от 09.02.2006 № 55 «Об итогах конкурса междисциплинарных интеграционных проектов СО РАН – 2006» (<http://www.sbras.ru/win/anonses/1342.html>).

Подавляющее большинство проектов, которые предусматривали получение новых результатов в области информатики или / и использование методов информатики для создания программных систем информационного обеспечения научной деятельности на основе новых интернет-технологий, имело в качестве организаций-исполнителей один или несколько академических институтов, занимающихся исследованиями в области информатики: Институт математики СО РАН, Институт вычислительных технологий СО РАН, Институт систем информатики СО РАН, Институт вычислительной математики и математической геофизики СО РАН, Институт вычислительного моделирования СО РАН, Институт динамики систем и теории управления СО РАН, Институт автоматизации и процессов управления ДВО РАН и др., либо вуз аналогичного профиля, например Новосибирский государственный университет или Томский университет систем управления и радиоэлектроники.

Анализ тематики проектов (табл. 2) позволил выделить 5 основных типов задач из области информатики, решаемых в рамках этих проектов, причем некоторые проекты могут соответствовать сразу нескольким типам задач:

- 1) исследование и когнитивное моделирование интеллекта;
- 2) разработка средств анализа моделей информационных структур;
- 3) проведение компьютерного анализа большого массива данных в той или иной области с целью получения новых знаний;
- 4) разработка и создание с использованием интернет-технологий специализированных информационных систем на основе современных алгоритмов обработки данных;
- 5) исследование общих принципов организации телекоммуникационных систем.

Таблица 2

Анализ тематики интеграционных проектов СО РАН,
связанных с информационными технологиями

Год	Всего проектов	Моделирование интеллекта	Анализ моделей информационных структур	Телекоммуникационные системы	Анализ данных, извлечение знаний	Создание информационных систем	
						всего	из них с анализом данных
2000	5	0	0	1 (20 %)	2 (40 %)	4 (80 %)	1 (25 %)
2003	12	1 (8 %)	0	2 (17 %)	6 (50 %)	8 (67 %)	4 (50 %)
2006	11	1 (9 %)	1 (9 %)	3 (27 %)	6 (55 %)	5 (45 %)	4 (80 %)

Таким образом, научное сообщество испытывает все более растущую потребность не просто в информационных системах, но в системах, извлекающих из имеющихся данных новые знания, т. е. в интеллектуальных системах, включающих в себя наряду с информационной системой (понимаемой как множество связанных между собой ресурсов), логическую систему, интеллектуальный интерфейс, а также средства автоматического пополнения базы данных и извлечения данных из текстов [Арский и др., 1996].

Еще один важный вывод на основе анализа интеграционных проектов, в рамках которых были созданы программные системы информационного обеспечения какой-либо отрасли науки, например «Электронный атлас биоразнообразия животного и растительного мира Сибири» или «Электронная библиотека MathTree», состоит в том, что подобные системы могут развиваться лишь в случае актуализации информации самими пользователями этих систем (см., в частности, [Ершов и др., 2006; Федотов и др., 1998]). Более того, даже относительно систем научно-организационной направленности, создаваемых в рамках одной большой научной корпорации (Сибирского отделения РАН), в [Жижимов и др., 2006] сделан вывод о том, что «эффективная эксплуатация информационных ресурсов возможна только в том случае, когда они постоянно поддерживаются авторами».

К тому же в интеллектуальных информационных системах компьютер в диалоговом режиме усиливает комбинаторное мышление и логические возможности человека, при этом происходит автоматизированное пополнение базы данных. В силу указанных обстоятельств

при работе с интеллектуальными информационными системами многих пользователей, возможности систем резко возрастают. Поскольку пользователи, принимающие участие в актуализации информации, могут находиться в разных регионах России и даже мира, постольку практическое взаимодействие таких программных систем с внешним миром в плане занесения в них новых данных целесообразно организовывать преимущественно (или даже почти исключительно) через веб-интерфейс.

Информационная модель описания деятельности научного сообщества

Прежде всего, сформулируем, информация о каких сущностях (точнее, классах сущностей) требуется при описании той или иной отрасли человеческой деятельности.

Любая деятельность человека предполагает определенное противопоставление субъекта и объекта деятельности, причем в качестве субъекта деятельности могут выступать как отдельные люди, так и группы (коллективы) людей. В условиях современного общества производственно-технические отношения между людьми возникают, как правило, посредством вхождения этих людей в одну группу, а характер этих отношений определяется функциями конкретного человека в группе. В свою очередь, группы также могут вступать между собой в те или иные общественные отношения (подчиненности, учредительства и т. п.). Таким образом, процесс деятельности организации может быть охарактеризован описаниями следующих сущностей.

1. Субъекты деятельности:

- а) группы;
- б) отдельные лица.

2. Объекты деятельности:

- а) продукты деятельности;
- б) акты деятельности.

Между этими сущностями устанавливаются связи.

1. Отношения между субъектами и объектами деятельности:

- а) группа – объект деятельности;
- б) персона – объект деятельности.

2. Отношения между субъектами деятельности – связи типа:

- а) группа – группа;
- б) персона – группа;
- в) персона – персона.

Что же касается связей между объектами деятельности, то ввиду сложности соответствующих моделей, а также их большой специфичности для каждой конкретной сферы деятельности в рамках данной статьи этот вопрос не рассматривается.

Выбор конкретного множества описаний из приведенного выше списка определяется родом деятельности организации. В работе [Барахнин и др., 2006] был проведен сравнительный анализ информационных моделей описания деятельности научного сообщества и других видов человеческой деятельности, сходных с ней в том или ином аспекте (производство, искусство, законодательная и общественная деятельность), а также анализ отражения этой деятельности на сайтах российского сектора сети Интернет (отечественная корпоративная культура, влияющая, в частности, на особенности представления организациями информации о своей деятельности, во многом обусловлена особенностями исторического развития России и весьма отличается от корпоративной культуры западных стран). Результаты анализа приведены в табл. 3 (знак «+» означает актуальность представления соответствующего аспекта, «±» – актуальность при отсутствии в большинстве случаев практической реализации, «–» – неактуальность).

Из таблицы 3 видно, что информационная модель описания научной деятельности обладает следующими особенностями:

- 1) необходимость включения подробной информации о персонах, связи персон сохраняют актуальность;
- 2) необходимость включения подробной информации о структуре групп;
- 3) возможность вхождения персоны сразу в несколько групп;

- 4) максимально подробное представление информации о предмете деятельности;
- 5) наличие связей между персонами и предметом деятельности.

Заметим, что к информационной модели описания деятельности научного сообщества наиболее близка модель описания деятельности органов законодательной власти, однако она реализована далеко не в полной мере даже на сайте Государственной думы (каждому созыву соответствует «новая» биография персоны, даже если эта персона была депутатом и других созывов; отсутствует привязка законодательных актов непосредственно к описанию персоны – автора соответствующей законодательной инициативы).

Дополнительно перечислим основные отличительные особенности информационного обеспечения научно-инновационной деятельности, отмеченные в [Шокин и др., 2005]:

- 1) субъекты деятельности – преимущественно организации;
- 2) возможность утраты отдельными источниками актуальности;
- 3) наличие большого количества «внешних» источников (например, нормативно-правовых актов), которые могут быть непосредственно связаны с объектами деятельности.

Отличительной особенностью предложенной информационной модели описания деятельности научного сообщества является четкое выделение субъекта и объекта деятельности. Предложенная модель эффективна при описании как научной деятельности в той или иной предметной области (когда основные субъекты деятельности – персоны), так и деятельности крупных научных корпораций (когда в качестве основных субъектов деятельности наряду с персонами выступают организации). Для сравнения заметим, что модель, использованная при создании Единого научного информационного пространства РАН [Бездушный и др., 2005] излишне персоноцентрична, например «организации» рассматриваются в одном ряду с «проектами». Другой отличительной особенностью предложенной модели является неиерархичность структуры субъектов деятельности, возникающая из-за возможности вхождения персоны сразу в несколько групп. В связи с этим требует решения проблема работы с персональными данными, которые могут одновременно принадлежать к разным ветвям иерархического дерева и вместе с тем должны однозначно определять персону, поскольку предполагаемая возможность извлечения из содержащихся в информационной системе данных новой информации и знаний влечет за собой необходимость наличия связи имен собственных (как элементов библиографического описания и т. п.) с информацией о конкретных носителях этих имен, ибо в противном случае имя несет лишь назывную, но не информационную функцию [Михайлов и др., 1976].

Таблица 3

Сравнительный анализ
информационных моделей описания различных сфер деятельности

Тип организации	Подробная информация о персонах	Отображение включения персоны в несколько групп	Сохранение неактуальных связей между персонами и группами	Связи между субъектами и объектами деятельности	Отображение подробной информации о предмете деятельности
Производственные организации	–	–	–	–	+
Общественные организации	+	–	–	–	–
Органы законодательной власти	+	±	+	±	+
Творческие коллективы	+	–	±	+	–
Научные организации	+	±	+	+	+

Отметим, что коммерческие пакеты управления бизнес-процессами (например, «1С», специализированные стереотипы UML) моделируют прежде всего производственно-экономические отношения (безотносительно самого процесса производства) и к тому же требуют серьезной адаптации к особенностям конкретной сферы деятельности.

Построение модели информационной системы

На основании проведенного анализа можно сделать следующие выводы.

С одной стороны, программная система информационного обеспечения научной деятельности должна объединять разнородные коллекции ресурсов, притом структура некоторых коллекций, в частности «Организации» и «Персоны», весьма сложна и отлична от традиционной иерархической, поскольку допускает возможность включения персоны сразу в несколько групп.

С другой стороны, требование автоматизированного получения из данных, с которыми работает система, новой информации и знаний влечет за собой необходимость того, чтобы информационные ресурсы, образующие систему, были снабжены метаданными, причем значения атрибутов этих метаданных, вообще говоря, не являются элементами заданных словарей (в отличие от подхода, принятого при разработке концепции Semantic Web). Отсюда вытекает, что основными структурными элементами информационной системы должны являться *документы* (целостные информационные объекты), понимаемые как информационные ресурсы (обладающие по определению уникальным идентификатором), снабженные метаданными [Шокин, Федотов, 2000]. Это принципиально отличает наш подход от модели RDF⁶ консорциума W3, рассматривающей в качестве элементов системы отдельные ресурсы, представляющие и сущности, и их характеристики, при этом ресурсы, могут, вообще говоря, и не сопровождаться метаданными.

Выбор документа в качестве основного структурного элемента информационной системы дает возможность задавать связи между сущностями, описываемыми системой, посредством установления связей между соответствующими документами, при этом один документ может являться частью другого полностью или частично, в том числе и в виде гиперссылки.

Исходя из сказанного, модель информационной системы, подробно описанная в [Барахнин, Федотов, 2007а], строится посредством задания классов K_i , определяемых соответствующими множествами элементов метаданных M_i , и типов возможных связей между классами $M^j < K_i, K_i >$ с указанием элементов метаданных $M_{i,i'}^j$, описывающих атрибуты соответствующих связей. Тем самым любой документ d_i системы представляется как $d_i = < m_i^{j,k} >$, где $m_i^{j,k}$ – значения элементов метаданных M^j , k – количество значений (с учетом повторений) соответствующего элемента метаданных в описании документа.

Связи между документами задаются посредством модели направленных связей [Барахнин, Леонова, 2005]: если документ d_i входит в качестве значения элемента M^j метаданных документа d_i , то можно говорить о связи между этими документами вида $M^j < d_i, d_i, m_{i,i'}^{l,k} >$, где $m_{i,i'}^{l,k}$ – атрибуты этой связи, являющиеся значениями соответствующих элементов метаданных. Таким образом, выстраиваемые отношения фактически переносятся на уровень элементов, определяющих структуру документов.

В работе [Барахнин, Федотов, 2007б] показано, что при построении модели соблюдены основные системные принципы: целостность, структурность, иерархичность, множественность описания, взаимозависимость системы и среды.

Нетрудно видеть, что принципы построения модели вобрали в себя черты, свойственные как для традиционного объектно-ориентированного подхода, так и для используемого в Semantic Web языка RDFS. В частности, мы описываем классы в терминах их структуры, как это принято в ООП, а не определяем свойства в терминах классов, что характерно для RDFS. Такой выбор связан с тем, что задание базовой структуры создаваемой системы, опирающее-

⁶ Resource Description Framework (RDF) Model and Syntax Specification. W3C Recommendation 22 February 1999. <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>.

ся на разработанную модель предметной области, носит централизованный характер. С другой стороны, ограничения, накладываемые моделью на свойства классов, носят менее жесткий характер, чем при объектно-ориентированном подходе (например, может быть объявлено произвольное, в том числе нулевое, количество значений некоторого элемента метаданных), что сближает нашу модель с подходом RDFS.

Заключение

Предложенная на основе проведенного анализа информационных потребностей научного сообщества информационная модель описания его деятельности, в которой четко выделены субъекты и объекты деятельности, легла в основу модели информационной системы, показавшей свою эффективность как при описании научной деятельности в той или иной предметной области (когда основные субъекты деятельности – персоны; см., например, [Барахнин, Федотов, 2007б]), так и деятельности крупных научных корпораций (когда в качестве основных субъектов деятельности наряду с персонами выступают организации; см., например, [Федотов и др., 2007]). В частности, модель информационной системы позволила решить нетривиальную задачу отнесения персоны одновременно к разным ветвям организационной структуры, не вводя «дублирующие» концепты, а разделив информацию о персоне на две части: личную – связанную с самой персоной, и служебную – связанную с должностью (должностями) персоны, причем каждой должности соответствует новая служебная запись.

Список литературы

- Арский Ю. М., Гиляревский Р. С., Туров И. С., Черный А. И.* Инфосфера: информационные структуры, системы и процессы в науке и обществе. М.: ВИНТИ, 1996.
- Барахнин В. Б., Леонова Ю. В.* Информационная модель отношений между документами в информационной системе // Вычислительные технологии. 2005. Т. 10. Специальный выпуск. С. 129–137.
- Барахнин В. Б., Леонова Ю. В., Федотов А. М.* К вопросу о формулировке требований для построения информационных систем научно-организационной направленности // Вычислительные технологии. 2006. Т. 11. Специальный выпуск. С. 52–58.
- Барахнин В. Б., Федотов А. М.* Информационная система: взгляд на понятие // Вестн. Новосиб. гос. ун-та. Серия: Информационные технологии. 2007а. Т. 5, вып. 2. С. 12–19.
- Барахнин В. Б., Федотов А. М.* Методика построения информационно-справочной системы по истории математической науки // Электронные библиотеки. 2007. Т. 10, вып. 1. [http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2007/linebreak part1/BF](http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2007/linebreak%20part1/BF).
- Барахнин В. Б., Федотов А. М.* Принципы структурирования сайтов информационной системы научного сообщества (на примере сайта Совета научной молодежи СО РАН) // Вычислительные технологии. Т. 9; Вестник КазНУ им. аль-Фараби. Серия: математика, механика, информатика. № 3 (42). Совместный выпуск. 2004. Часть 1. С. 254–259.
- Бездушный А. Н., Кулагин М. В., Серебряков В. А. и др.* Предложения по наборам метаданных для научных информационных ресурсов // Вычислительные технологии. 2005. Т. 10. Специальный выпуск. С. 29–48.
- Богданов А. А.* Тектология: всеобщая организационная наука: В 2 кн. М.: Экономика, 1989.
- Гвоздева Е. С., Высоцкий Е. М.* Сегодняшний день будущего российской науки. Новосибирск: Изд-во СО РАН, 2004.
- Еришов Ю. Л., Клименко О. А., Матвеева И. И. и др.* Древовидный каталог математических интернет-ресурсов // Информационные ресурсы России. 2006. № 1. С. 5–8.
- Жижимов О. Л., Турпанов А. А., Федотов А. М.* Корпоративный каталог СО РАН // Труды Восьмой Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» (RCDL'2006). Ярославль, 2006. С. 226–230.
- Михайлов А. И., Черный А. И., Гиляревский Р. С.* Научные коммуникации и информатика. М.: Наука, 1976.
- Пойа Д.* Как решать задачу / Пер. с англ. М.: Учпедгиз, 1959.

Федотов А. М., Артемов И. А., Ермаков Н. Б. и др. Электронный атлас «Биоразнообразие растительного мира Сибири» // Вычислительные технологии. 1998. Т. 3. № 5. С. 68–78.

Федотов А. М., Барахнин В. Б., Гуськов А. Е. и др. Построение информационной системы научного сообщества на основе интеграции разнородных коллекций ресурсов // Сборник тезисов постерных докладов Девятой Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» (RCDL'2007). Переславль-Залесский, 2007. С. 111–117.

Шокин Ю. И., Белов С. Д., Чубаров Л. Б. Предварительные результаты тестирования создаваемой системы мониторинга и сбора статистики СПД СО РАН // Вычислительные технологии. 2007. Т. 12, № 5. С. 126–134.

Шокин Ю. И., Федотов А. М. Информационная система Сибирского отделения РАН // Труды Второй всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» (RCDL'2000). Протвино, 2000. С. 6–14.

Шокин Ю. И., Федотов А. М., Барахнин В. Б. Особенности организации системы управления веб-контентом сайтов информационной поддержки инновационной деятельности // Вычислительные технологии. 2005. Т. 10. Специальный выпуск. С. 122–128.

Материал поступил в редколлегию 28.08.2008

V. B. Barakhnin, A. M. Fedotov

**Studying the Information Needs of Scientific Community
for Constructing the Information Model of Its Activity**

The work investigates the information needs of the scientific community based on the well-known «5W+ 1H» approach and on the subject analysis of joint information projects of collective users, i.e. SB RAS organizations. On the basis of this study a justification for the scientific community information model has been performed. This model in turn served as a foundation for the model describing software systems for the information support of scientific activity intended for the inclusion of semistructured internet documents in the scientific information process.

Keywords: information needs, information model of activity of scientific community.