

ПРОБЛЕМА ПРЕДСТАВЛЕНИЯ ТРУДОВ НАУЧНОГО КОЛЛЕКТИВА В ИНТЕРНЕТ С ИСПОЛЬЗОВАНИЕМ МИНИМАЛЬНЫХ СРЕДСТВ ЭЛЕКТРОННОЙ СРЕДЫ

Е.И. Моисеев, А.А. Муромский, Н.П. Тучкова

Проблеме информационного запроса в математике посвящено достаточно много работ [1, 2], однако она далека от разрешения. Впервые с ней столкнулись издатели научной сериальной литературы на заре эры компьютерных и ИНТЕРНЕТ-технологий. Оказалось, что размещение публикации в базе данных не обеспечивает ее эффективного нахождения при информационном запросе, если сама база данных не организована определенным образом. Далее, размещение базы данных в ИНТЕРНЕТ уменьшило релевантность документа запросу, т.к. изменились в худшую сторону параметры полноты и точности поискового массива (увеличился шум), также увеличилось время поиска.

Основатели библиографических баз данных (международные издательства, такие как McGraw-Hill, и др.) обратились к научному сообществу с проблемой информационного запроса. В частности, речь шла о запросе математической библиографии. Еще в 70-е годы XX в. появились работы об экспоненциальном характере роста математической литературы [3], также было замечено, что каждые 50 лет возникает около 200 новых математических терминов [4]. Были инициированы работы по созданию указателя для американского математического журнала (American Mathematical Monthly)[5], послужившего моделью для предметного тезауруса, который разрабатывался в рамках проекта, организованного коллективом журнала *Historia Mathematica* [6, 7]. Этот проект показал всю сложность поставленной задачи, поскольку кроме основной цели – применения тезауруса для поиска информации, возникла необходимость решения целого ряда связанных проблем, а именно: проблема форматов и стандартов для представления терминов и математических записей в электронном виде, проблема архивирования данных, проблема организации доступа к ним. Кроме прочего, возникла организационная проблема, связанная с необходимостью обработки огромного количества печатной информации и привлечению для этого квалифицированных высокооплачиваемых специалистов, так как первоначально предполагалось, что достаточно привлечь студентов соответствующих специальностей. Предварительные оценки затрат на проект оказались неверными и он был реализован только частично.

В настоящее время библиографическая информация размещается в ИНТЕРНЕТ, что способствует возведению очередной Вавилонской башни математической литературы и составляет проблему математической коммуникации [6, 8, 9]. В AMS (Американское Математическое Общество) затраты, необходимые для обеспечения задачи информационного запроса в математике, были оценены приблизительно в сто миллионов долларов США. Сама же эта задача определена как одна из глобальных информационных задач, так как математические знания со времен Эвклида, относятся к фундаментальным знаниям, которые передаются от поколения к поколению и должны быть достоверны [11]. Это породило к жизни проект AMS «World Mathematics Library», объединяющий усилия издательств, научных сообществ, университетских и школьных библиотек, направленных на решение проблемы оцифровки и распространение математической литературы [12].

Для отечественной научной библиографии такого глобального проекта нет. Те работы российских авторов, которые, опубликованы в международных издательствах и/или попали в соответствующие базы данных, будут найдены, а остальные - могут быть не найдены. Это распространяется и на так называемые "серые публикации" - отчеты и труды академических институтов, препринты и сборники университетов, труды российских конференций, которые, как известно, несут в себе большой научный потенциал и способствуют сохранению приоритетов в науке, т.к. сроки их опубликования в 2-3 раза меньше, чем в сериальных международных изданиях. К тому же, "серые публикации" имеют сложные смешанные авторские права, что не позволяет отдельному автору самостоятельно включить информацию о препринте, например, в e-library или в другие электронные библиографические ресурсы. Исключение этих и других математические публикации из общего мирового информационного потока, способствует снижению международного авторитета российской науки, а также изоляции отдельных научных коллективов. В то же время разработчики библиографических информационных ресурсов [13] (как печатных, так и электронных) признают, что без участия в той или иной мере фактически каждого автора, реализовать такие масштабные проекты невозможно.

В этой связи уместно обратить внимание на библиометрические показатели: число научных публикаций, их цитируемость и др., которые используются во многих странах мира, и прежде всего в странах Европейского сообщества, для оценки состояния и развития науки. В 90-е годы XX в. стали востребованы и уже активно использовались в развитых странах библиографические базы данных (БД), такие как NATIONAL SCIENCE INDICATORS (распространяется Институтом научной информации ISI,

Пенсильвания, США), SCIENCE CITATION INDEX, SOCIAL SCIENCE CITATION INDEX, ARTS & HUMANITIES CITATION INDEX или JOURNAL CITATION REPORTS и др. В БД ISI National Science Indicators, собраны данные ряда стран начиная с 1981. В 1997 г. в России не было ни одной из перечисленных БД, что безусловно отразилось на отечественных библиометрических показателях. В работе [14] указывается среднее число ссылок на одну публикацию за период 1981-1992 гг. по странам (см. табл. 1). По данным этой таблицы, СССР занимало последнее место. В этом факте прежде всего заметно слабое, иначе не скажешь, представление информации об отечественных работах. В табл. 2 приведены некоторые данные по исследовательской активности стран в период 1996-2000 гг. из БД NSIOD (ISI Национальные показатели науки, National Science Indicators on Diskette) [15]. Приведенные данные показывают, как изменился научный мир на пороге XXI в. Произошло перераспределение в пользу развитых европейских стран, Россия переместилась на 8-е место, но значение индекса цитирования российских публикаций выросло всего на 0.15.

Таблица 1. Среднее число ссылок на одну публикацию за период 1981-1992 по странам

Страна	Impact (Среднее число ссылок на публикации)
США	10.68
Швейцария	10.65
.....
Китай	1.83
Румыния	1.83
Египет	1.76
Нигерия	1.71
.....
СССР	1.43

Таблица 2. Исследовательская активность стран в период 1996-2000 гг. Общее число публикаций в БД - 3 570 733 (По данным базы NSIOD работы во всех областях естественных, социальных и гуманитарных наук.)

Ранг	Страна	Процент публикаций	Общее число публикаций в стране	Процент цитируемости публикаций	Impact (среднее число ссылок на публикации)
1.	USA	35,32	1 261 111	63.01	5.58
2.	UK	9,47	338 277	61.51	4.64
3.	JAPAN	9,17	327 465	57.82	3.47
4.	GERMANY	8,64	308 595	60.73	4.42
5.	FRANCE	6,36	226 934	59.41	4.19
6.	CANADA	4,68	167 107	61.31	4.52
7.	ITALY	3,98	142 274	60.60	4.12
8.	RUSSIA	3,52	125 530	37.75	1.58
...

Новая эра ГРИД-технологий ставит задачу информационного запроса в математике в другую плоскость, она может быть реализована как максимальное полное представление трудов отечественных ученых на основе однородных данных и однородных вычислительных систем и организация доступа к ним на основе унифицированных ГРИД-порталов.

В данной работе обсуждаются вопросы использования имеющихся у российского научного сообщества средств для включения публикаций российских авторов в мировые научные библиографические информационные ресурсы. Ставится цель – предоставить каждому автору возможность размещения своих публикаций в открытом доступе, при соблюдении авторских, интеллектуальных и других прав.

Предлагается вариант решения ряда основных проблем, связанных с представлением публикаций в электронной среде, а именно, задание вида и состава вторичного документа, размещения библиографии, каталогизации, способа индексирования и др.

ВТОРИЧНЫЕ ДОКУМЕНТЫ

Представление научной публикации в специальном виде, позволяющем получить наиболее полную информацию о ее ценности и способности удовлетворить информационную потребность, должен давать

вторичный документ. Большинство научных издательств и библиотек, имеющих представительства в ИНТЕРНЕТ, решает этот вопрос, размещая каталог публикаций, куда входят

Обязательные поля (например, ГПНТБ, <http://library.gpntb.ru>):

- ◆ авторы
- ◆ название
- ◆ коды классификаторов (УДК, MSC),
- ◆ тематические рубрики.

Дополнительные поля со свободным доступом (например, IET Digital Library, <http://www.ietdl.org/>):

- ◆ аннотации
- ◆ индекс цитирования

Дополнительные поля при авторизованном (платном или бесплатном для определенного сообщества) доступе:

- ◆ ключевые слова
- ◆ содержание (оглавление)
- ◆ полный текст

Вопрос о составе вторичного документа и о правах доступа к обязательным и дополнительным полям в каждом отдельном случае решается правообладателями.

Для научных публикаций ВЦ РАН разработан состав вторичного документа [16], состав указателя трудов ВЦ РАН.

Вторичный документ включает следующие поля: фамилии авторов, названия, аннотации, ключевые слова, содержание (оглавление), тип, идентификатор, УДК, список библиографии, каталожную карточку библиотеки ВЦ РАН, наименование раздела тематического указателя.

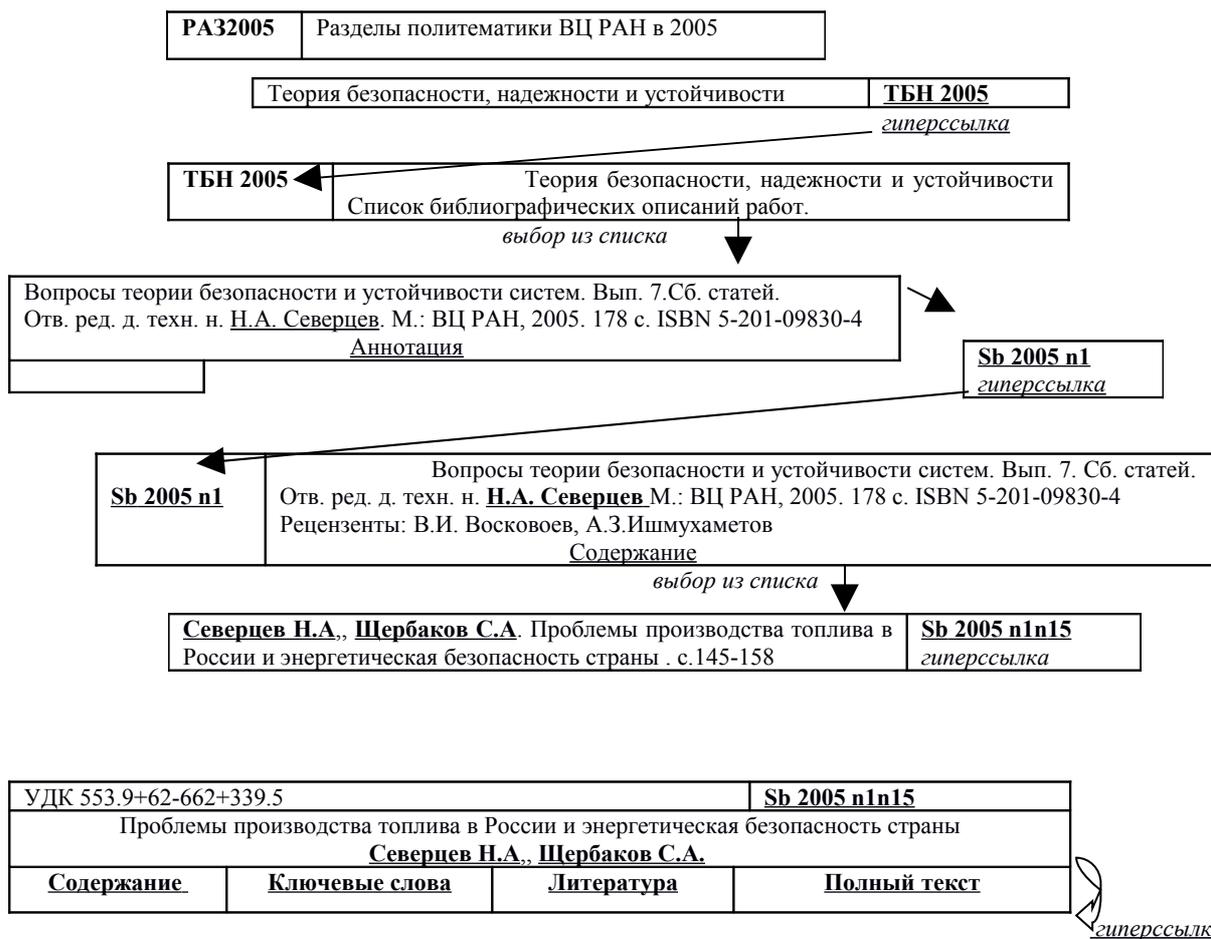
Пример 1. Вторичный документ ВЦ РАН, описание позиций и содержание.

<i>наименование раздела (гиперссылки на тематический указатель)</i>	Тезаурусы. Дескрипторные словари. Указатели	
<i>тип публикации (гиперссылки в указателе соответствующего раздела публикаций) и идентификатор публикации</i>	<u>Монографии ВЦ РАН</u>	M2005n10
<i>УДК</i>	УДК 025.43:517.91/.938	
<i>фамилии авторов</i>	Е.И.Моисеев, А.А.Муромский, Н.П.Тучкова	
<i>название публикации</i>	Тезаурус информационно-поисковый по предметной области: обыкновенные дифференциальные уравнения. Отв.ред. академик РАН Е.И. Моисеев. М.: МАКС Пресс, 2005. 116 с. ISBN 5-317-01489-1	
<i>аннотация (гиперссылка)</i>	<u>Аннотация</u>	
<i>ключевые слова (гиперссылки в указателе ключевых слов)</i>	<u>Ключевые слова:</u> <u>обыкновенные дифференциальные уравнения, дескрипторные словари,</u> <u>тезаурус, лексико-семантические указатели, иерархические указатели,</u> <u>контролируемая лексика, индексирование документов,</u> <u>парадигматические отношения</u>	
<i>содержание (гиперссылка)</i>	<u>Содержание</u>	
<i>литература (гиперссылка)</i>	<u>Литература</u>	
<i>каталожная карточка библиотеки ВЦ РАН (гиперссылка)</i>	<u>Каталожная карточка</u> ...	
<i>Соответствующие тематические ссылки (гиперссылки на тематический указатель)</i>	Указатель тематических ссылок: <u>Дифференциальные уравнение. Интегральные и другие</u> <u>функциональные уравнения</u> <u>Интегральные уравнения</u> <u>Функциональные уравнения</u>	

КАТАЛОГИЗАЦИЯ, РАЗМЕЩЕНИЕ БИБЛИОГРАФИИ, ОРГАНИЗАЦИЯ ИНФОРМАЦИОННО-ГО ЗАПРОСА

Представление библиографии в ИНТЕРНЕТ сопровождается каталогизацией: составлением списков, словарей, тезаурусов, указателей научной библиографии. Проблема доступа к библиографической информации рассматривается в контексте использования политематического указателя и тезауруса. Для доступа к массиву публикаций ВЦ РАН предлагается вход через указатели: авторский, ключевых слов, политематический. Научные публикации хранятся на локальных серверах институтов РАН, авторизованный доступ к которым осуществляется через порталы ГРИД-среды[17]. Для доступа открыты вторичные документы научных публикаций. Полные тексты открываются при согласии правообладателей. В информационном запросе могут использоваться все поля вторичных документов.

Пример 2. Функциональная схема доступа к публикации через электронный Указатель-2005.



Работа выполнена при поддержке РФФИ (проект 07-07-00082) и программы фундаментальных исследований Президиума РАН (проект П-15).

ЛИТЕРАТУРА:

1. Новые технологии в информационном обеспечении науки: Сб. научн. тр. Каленов Н.Е. (ред.). М.: Научный Мир, 2007. 310 с.
2. Научный сервис в сети ИНТЕРНЕТ: многоядерный компьютерный мир. 15 лет РФФИ: Труды Всероссийской научной конф. (24-29 сентября 2007 г., г. Новороссийск). М.: Изд-во МГУ, 2007. 407 с.
3. Михайлов А.И., Черный А.И., Гиляревский Р.С. Основы информатики. Изд. 2-е, перераб. М.: Наука, 1968. 756 с.
4. May, K.O. Growth and quality of the mathematical literature. 1968. Isis 59, P. 363-371.
5. May K.O. Index of the American Mathematical Monthly, Volumes 1 through 80 (1894-1973)., Mathematical Association of America, 1977. Washington, DC.
6. Enros P.C. May Dictionary and Thesaurus Projects, Scope and Status Report., Institute for the History and Philosophy of Science and Technology, University of Toronto, Toronto. 1980. 62 p.

7. Lewis A.C. Kenneth O. May and information retrieval in mathematics. // *Hist. Math.*. 2004, V. 31, P. 186-195.
8. Петров А.А. Методы оценки влияний информационных и телекоммуникационных технологий на макропоказатели эффективности роста экономики. М.: ВЦ РАН, 2005. 225 с.
9. Price D.J. and de Solla. *Science Since Babylon.* , Yale Univ. Press, New Haven, 1961. СТ.
10. Second UCSU-UNESCO International Conference on Electronic Publishing in Science, which took place in February 2001. The Proceedings will appear at <http://Associnst.ox.ac.uk/~icsuinfo/> in the future.
11. Second UCSU/UNESCO International Conf. on. Electronic Publishing in Science, Paris, February 20-23, 2001
12. Ewing, J. Twenty centuries of mathematics: digitizing and disseminating the past mathematical literature. // *Notices of AMS.* 49, 2002. P. 771-777.
13. Черный А.И. Введение в теорию информационного поиска. М.: Наука, 1975. 238 с.
14. Маршакова-Шайкевич И.В. Вклад России в развитие науки: библиометрический анализ. М.: ТОО «Янус», 1995. 248 с. ISBN 5-88929-005-3
15. Маршаков-Шайкевич И.В. Вклад России в развитие мировой науки: библиометрическая оценка // *Отечественные записки*, 2002, №7
16. УКАЗАТЕЛЬ-2004. Указатель трудов Вычислительного центра им. А.А.Дородницына Российской академии наук за 2004 год. Составители: Г.М.Михайлов, А.А.Муромский, Е.В.Панфилова, Н.П.Петрова, Н.П.Тучкова М.:ВЦ РАН, 2006.296 с.
17. Технологии ГРИД. Том 1,2. М.: ИПМ им. М.В.Келдыша, 2006. 380 с. ISBN 5-98354-003-3.