

Особенности сводного электронного каталога БЕН РАН

Рассматриваются принципы построения и поисковые возможности сводного интернет-каталога книг и продолжающихся изданий, поступающих в фонды более 100 библиотек централизованной библиотечной системы, возглавляемой Библиотекой по естественным наукам (БЕН) РАН. Каталог позволяет обрабатывать практически неограниченный круг запросов, включающих в себя как любые элементы библиографических описаний, так и тематические рубрики, являющиеся некоторой «настройкой» над УДК. Библиографические записи, составляющие основу каталога, формируются в рамках интегрированной библиотечной системы БИБЛИОБУС и представляют собой в небольшой степени модифицированные стандартные библиографические описания.



**Николай Евгеньевич
Калёнов,**
*директор Библиотеки по
естественным наукам
Российской академии наук,
доктор технических наук,
профессор*

Ключевые слова: библиографический поиск, библиотечные каталоги, автоматизированные библиотечные системы, Интернет, БЕН РАН, ЦБС.

С начала 1990-х гг. в Библиотеке по естественным наукам (БЕН) РАН поддерживаются сводные интернет-каталоги, отражающие поступление изданий в фонды Центральной библиотеки (ЦБ) и библиотек, входящих в ее централизованную библиотечную систему (ЦБС) [3]. ЦБС БЕН РАН включает в себя более 120 библиотек, обслуживающих академические институты и научные центры московского региона. Основная часть фондов библиотек комплектуется централизованно ЦБ за счет средств, выделяемых БЕН Президиумом РАН по специальной целевой программе. Все издания, приобретаемые для фондов ЦБС (за исключением журналов, выписываемых непосредственно в адреса отдельных библиотек через подписные агентства), поступают в ЦБ, где проходят централизованную обработку (каталогизацию и систематизацию). Издания, получаемые в дар библиотеками, также передаются в ЦБ на обработку. Информация о журналах, поступающих децентрализованно в библиотеки через подписные агентства, вводится в сетевом режиме по мере поступления в обобщенную базу данных, поддерживаемую ЦБ. Благодаря такой технологии в ЦБ накапливается полная информация о фондах всех библиотек, входящих в ЦБС БЕН РАН, которая служит основой для создания сводных каталогов.

Исходя из результатов анализа информационных потребностей основных пользователей БЕН РАН — сотрудников академических институтов, при создании электронных каталогов (ЭК) Библиотека пошла путем, развивающим традиционный библиотечный подход — поддержку двух видов каталогов — журналов и книг.



**Светлана Александровна
Власова,**
*ведущий научный
сотрудник Библиотеки
по естественным наукам
Российской академии наук,
кандидат технических наук*

Такой подход обусловлен принципиальным отличием запросов на каталожную информацию по этим видам изданий. В подавляющем большинстве случаев пользователь, обращаясь к каталогу журналов, знает, какой именно журнал ему нужен, поскольку либо работает с определенным кругом журналов постоянно, либо ищет выпуск по библиографической ссылке. В редких случаях, когда ученый приступает к новой теме, его интересует, какие журналы по этой теме имеются в библиотеке. Поэтому БЕН РАН в своих журнальных каталогах предлагает пользователю возможность выбрать нужный журнал из алфавитного или тематического списков, а затем просмотреть, какие выпуски журнала имеются в фондах, и перейти к полным текстам интересующих статей (при наличии прав доступа к сетевой версии журнала). Что касается книг и продолжающихся изданий (материалы конференций, сборники трудов организаций и научных обществ), то пользователь во многих случаях заранее не знает, что конкретно ему нужно, какие новые книги поступили в библиотеку; он может знать только фамилию автора, приблизительное название, тематику и т. п., поэтому интерфейс поиска записей в этом каталоге должен быть существенно более развит, чем в каталоге журналов.

В настоящей статье рассматриваются особенности сводного интернет-каталога книг и продолжающихся изданий на сайте БЕН РАН (<http://www.benran.ru>). В каталоге отражены поступления отечественной литературы (с 1993 г.) и зарубежной (с 1995 г.).

Информация для сводного каталога формируется в процессе автоматизированной обработки изданий, поступающих в ЦБ. Комплексная автоматизация всех технологических процессов по «пути книги» для неперiodических изданий (предварительный заказ, регистрация и распределение поступлений по библиотекам ЦБС, каталогизация, систематизация, инвентаризация, шифровка, распечатка всех учетно-финансовых документов) обеспечивается разработанной специалистами БЕН РАН системой БИБЛИОБУС [2]. Эта система является развитием системы «SOLAR» [1], созданной и эксплуатируемой в БЕН РАН с 1993 по 2005 год. Хотя БИБЛИОБУС с точки зрения архитектуры системы и ее интерфейсов (она построена по технологии клиент-сервер на основе СУБД MS-SGL) принципиально отличается от SOLAR (которая представляет собой файловую систему, базирующуюся на структуре типа ISO-2709), технологические решения, связанные с каталогизацией поступающих в ЦБС БЕН РАН изданий, принятые для системы SOLAR, используются и в новой системе.

В вопросах машинной каталогизации БЕН пошла путем, отличным от большинства отечественных и зарубежных библиотек, в которых катало-

гизаторы вводят в компьютер отдельные элементы библиографического описания, соответствующие полям формата MARC (в России — RUSMARC [5]). Как показал проведенный нами анализ, внедрение такого ввода потребовало бы втрое увеличить штат каталогизаторов. Стремясь обеспечить максимальную эффективность внедрения автоматизации, БЕН решила задачу машинной каталогизации другим путем. Каталогизаторы вводят в компьютер библиографическую запись (БЗ) в соответствии со стандартами 7.1—2003, 7.80—2000 и др., дополненную сведениями обо всех коллективных, индивидуальных авторах и других лицах (редакторах, составителях и т. п.), связанных с изданием. Поиск в каталоге (см. ниже) возможен по всем словам, входящим в БЗ. Выделение отдельных элементов описания, предназначенных для уточнения области поиска в ЭК (индивидуальные и коллективные авторы, заглавие, год издания, ISBN и ISSN), а также, при необходимости, формирование различных видов MARC-записей (что требуется при передаче данных в другие библиотечные системы) осуществляется автоматически с помощью специальных программных средств, разработанных сотрудниками БЕН. Чтобы обеспечить однозначность распознавания элементов библиографической записи и исключить дублирование данных, при вводе соблюдаются три правила, несколько отличающие вводимую информацию от традиционной записи — все персоналии вводятся в именительном падеже, фамилия отделяется от имени (инициала) знаком подчеркивания (вместо «под ред. Ивановой А.А.» вводится «Иванова_А.А., ред.»), все разделители библиотечных областей («//», «/» «:», тире после точки в разделителе «. —») окружены пробелами.

Такой подход позволил перейти на машинную каталогизацию без увеличения штата каталогизаторов и без серьезного обучения их принципам каталогизации, основанным на формате типа MARC. Эти принципы заложены в программное обеспечение. Автоматизированная каталогизация всего потока отечественных изданий введена в промышленную эксплуатацию с начала 1993 года. За всю историю машинной каталогизации и электронного каталога не было выявлено ни одной серьезной проблемы, связанной с отказом от ввода библиографического описания по отдельным полям.

Для обеспечения доступа к сводному ЭК книг и продолжающихся изданий специалистами БЕН РАН создана специальная система программ, написанных с использованием компилятора C++ Builder и работающих в CGI-технологии [4]. Эта система позволяет проводить в каталоге развитый поиск и получать необходимую информацию на экране компьютера в формате стандартных библиографических описаний.

При создании каталога разработчики ориентировались на неподготовленного пользователя, незнакомого с библиотечной терминологией.

В то же время, с нашей точки зрения, каталог научной библиотеки как поисковая система должен существенно отличаться от общих поисковых машин типа Google, Yandex, Rambler и пр. Отличие состоит в том, что каталог должен точно отвечать на вопрос, имеются ли в фондах библиотеки именно те издания, которые ищет пользователь, в то время как поисковые машины выдают в качестве релевантной информацию, содержащую огромное количество «шума», не представляющего интерес для пользователя.

Поиск в ЭК осуществляется по следующим поисковым полям:

- Слова из библиографического описания;
- Автор;
- Название;
- Организация, конференция;
- Год издания;
- УДК.

Запрос на поиск может содержать от одной до трех строк терминов, соответствующих выбранным поисковым полям. Строки соединяются логическим «И», «ИЛИ», «НЕ» в соответствии с выбранным условием поиска. В каждую строку может быть введено несколько терминов, связанных логическим «И» или «ИЛИ» (рис. 1).

В отличие от универсальных поисковых машин термины запроса в каждой строке по умолчанию соединяются связкой «ИЛИ». Сделано это для того, чтобы в некоторой степени компенсировать возможные ошибки пользователя при вводе запроса. При работе с универсальной поисковой машиной вероятность нулевой выдачи документов по всему мировому сетевому пространству чрезвычайно мала, поэтому пользователь, не получив ни одного документа на свой запрос, в первую очередь, будет искать ошибки в запросе. Вероятность же нулевой выдачи результатов поиска в каталоге отраслевой библиотеки — явление обычное (книга отсутствует в фонде), поэтому пользователь, допустив орфографическую ошибку в одном из терминов библиографического описания интересующего его издания при нулевой выдаче, не будет задумываться о правильности запроса.

Получив же, по умолчанию, большой массив документов из каталога, пользователь обратит на это внимание и обязательно проверит правильность запроса, а заодно поменяет «ИЛИ» на «И».

Система позволяет проводить поиск всех записей, содержащих начальный фрагмент слова (поиск

Рис. 1.
Пример
поискового
запроса

с правым усечением). Для этой цели фрагмент термина завершается символом * (звездочка).

При работе с каталогом имеется возможность просмотра, выбора и автоматического ввода в запрос терминов из словарей (алфавитных списков), соответствующих различным поисковым полям. Для просмотра словаря нужно в строку запроса ввести поисковый термин или его фрагмент, выбрать соответствующее поле и нажать кнопку «Словарь». Например, для введенного в строку запроса фрагмента «Аналитическая теория» (поисковое поле «Название») на экран будет выдан фрагмент словаря из 20 терминов. Следующие 20 терминов словаря можно получить, нажав «Вперед», предыдущие — «На-

зад». Чтобы перейти в любое другое место данного словаря, следует в соответствующую строку внизу страницы ввести новый термин и нажать кнопку «Словарь».

Просматривая словарь, пользователь отмечает интересующие его термины, а затем активизирует кнопку «Возврат на поиск». Выбранные термины автоматически помещаются в строку запроса и соединяются логической связкой «ИЛИ». При необходимости связку можно поменять на «И».

Рассмотрим особенности поиска по различным поисковым полям.

Поле «Слова из библиографического описания» включает в себя все слова, введенные каталогизатором в библиографическое описание издания, за исключением служебных и неинформативных слов (предлоги, союзы, артикли и т. п.), перечень которых определяется словарем «стоп-слов», используемым при формировании каталога.

Необходимо отметить, что каталогизаторы, формирующие библиографические описания, сокращают многие слова в соответствии с ГОСТом (например, «тез.» вместо тезисы, «учеб.» вместо «учебное» и т. п.). Если осуществлять поиск по элементам такого библиографического описания, пользователь, не подозревающий о сокращениях, рискует «недополучить» значительную часть документов, соответствующих его запросу (на запрос с термином «тезисы», документы, содержащие в описании «тез.», выданы не будут). Этот факт учтен при разработке электронного каталога БЕН РАН. При его формировании используется специальный словарь сокращений, что позволяет по запросу, содержащему несокращенные слова, выдавать записи, содержащие сокращения.

Данная система позволяет также проводить поиск по коротким словам, состоящим из двух букв, и терминам, содержащим специальные символы, что не реализовано во многих представленных в Интернете каталогах. Для нахождения книг, в названиях которых входят слова «НТ» или «С++», достаточно в запросе ввести эти слова, выбрав поисковое поле «Слова из библиографического описания». Поскольку год издания книги является одним из слов библиографического описания, поиск по году также можно задавать в этом поле (иногда это, как будет показано ниже, достаточно удобно). При этом возможно, хотя и крайне маловероятно, появление нескольких «шумовых» записей, если в описании книги встретится число, равное заданному году.

Поисковое поле «Автор» подразумевает не только собственно автора издания, но и любое лицо, несущее интеллектуальную ответственность за издание (редактор, составитель и т. п.), введенное библиографом в каталог.

Поиск по этому полю возможен по фамилии (например, Фролов), фамилии и одному инициалу (при этом после инициала вводится символ правого усечения — Фролов О*), фамилии и двум инициа-

лам (Фролов О.П.). Следует иметь в виду, что библиографы при описании издания вводят сведения об авторе так, как они представлены в издании, поэтому на запрос «Фролов О.П.» запись, автор которой введен как «Фролов Олег Павлович», выдана не будет. В связи с тем, что на разных изданиях «персона» может присутствовать в разных формах, рекомендуется выбирать автора из соответствующего словаря во всех имеющихся формах и использовать связку «ИЛИ» либо вводить в запрос фамилию и первую букву имени с символом правого усечения (например, Фролов О*).

Поиск по полю «Название» предусматривает поиск по начальному фрагменту заглавия издания. Фрагмент может состоять из одного или нескольких слов, последнее из которых может иметь правое усечение. Если пользователь не уверен, что заглавие начинается с данных слов, лучше использовать поиск по полю «Слова из библиографического описания», соединяя термины связкой «И».

Запрос по полю «Организация, конференция» позволяет найти издания, авторами которых выступают не отдельные лица, а коллективы. К таким изданиям относятся труды институтов, материалы конференций, симпозиумов и т. п. Наименования организаций и коллективов представлены в каталоге так, как в издании. Поскольку названия организаций и конференций могут иметь неодинаковые написания в различных изданиях, рекомендуется пользоваться при формулировке запроса соответствующим словарем или вводить в запрос разные варианты названия.

Поиск по полю «УДК» — это поиск по заданным индексам Универсальной десятичной классификации. Этим видом поиска пользуются, прежде всего, библиотечные специалисты, которые знакомы с системой УДК. Наряду с данным поиском в каталоге реализован поиск библиографических описаний по тематическим разделам («Тематический поиск»), ориентированный на специалистов в конкретной предметной области. Такой поиск имеет специфические особенности, которые приведены ниже.

При нажатии кнопки «Тематический поиск» (рис. 1) на экране раскрывается список наименований рубрик верхнего уровня, позволяющий осуществлять переход на рубрики более низких уровней (рис. 2).

Для осуществления поиска по необходимым рубрикам нужно их отметить и нажать «Поиск». При необходимости уточнения запроса, например, по году или виду издания (справочник, учебник и т. п.) следует нажать кнопку «Уточнение запроса», после чего система перейдет на поисковую страницу, в первую строку запроса которой будут внесены индексы выбранных рубрик и указано поисковое поле «УДК». Во вторую и третью строки можно добавить термины, уточняющие запрос, а затем выполнить запрос, нажав «Поиск». Рубрики,

предлагаемые каталогом для выбора пользователям, являются некоторой надстройкой над индексами УДК — под каждой из них в системе подразумевается определенная комбинация индексов УДК. Она и преобразуется в запрос, отправляемый на выполнение после нажатия кнопки «Поиск» (или в первую строку запроса после нажатия «Уточнение запроса»). Надстройка над индексами УДК, обеспечивающая поиск по тематическому рубрикатору, разрабатывается специалистами-систематизаторами БЕН РАН, являющимися членами Методического совета по УДК при ВИНТИ.

Перечисленные возможности каталога позволяют обрабатывать достаточно сложные запросы, например запрос на поиск издания, не являющегося учебником, изданного в 2010 г., связанного с применением нано-

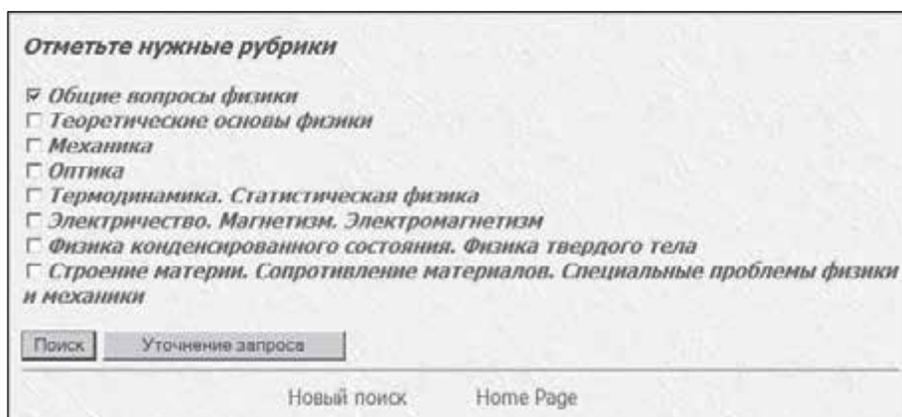


Рис. 2.
Список под-
рубрик рубри-
ки «Физика.
Механика»

технологий в органической химии. Для формулировки данного запроса пользователь должен войти в «Тематический поиск», раскрыть рубрику верхнего уровня «Химия», из рубрик следующего уровня отметить рубрику «Органическая химия» и нажать кнопку «Уточнение запроса». Система перейдет на поисковую страницу, в первую строку запроса которой будет внесен индекс выбранной рубрики и указано поисковое поле «УДК». Во вторую строку запроса для поискового поля «Слова из библиографического описания» нужно ввести термины «нано*» и «2010», соединив их логическим «И». В третью строку запроса также для поискового поля «Слова из библиографического описания» необходимо ввести термин «учеб*» и перед данной строкой указать логическое «НЕ».

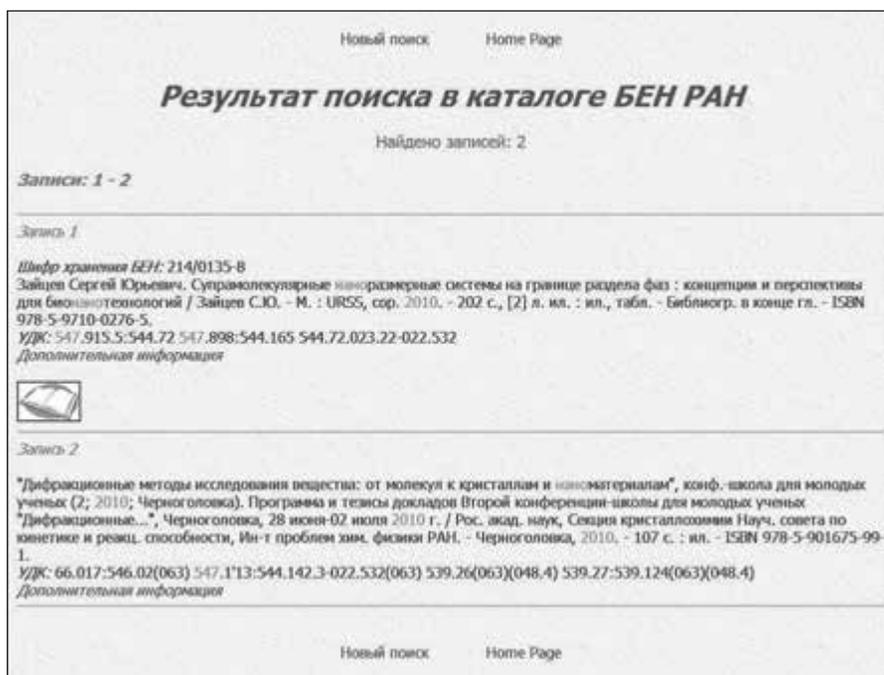


Рис. 3.
Результат
выполнения
поискового
запроса

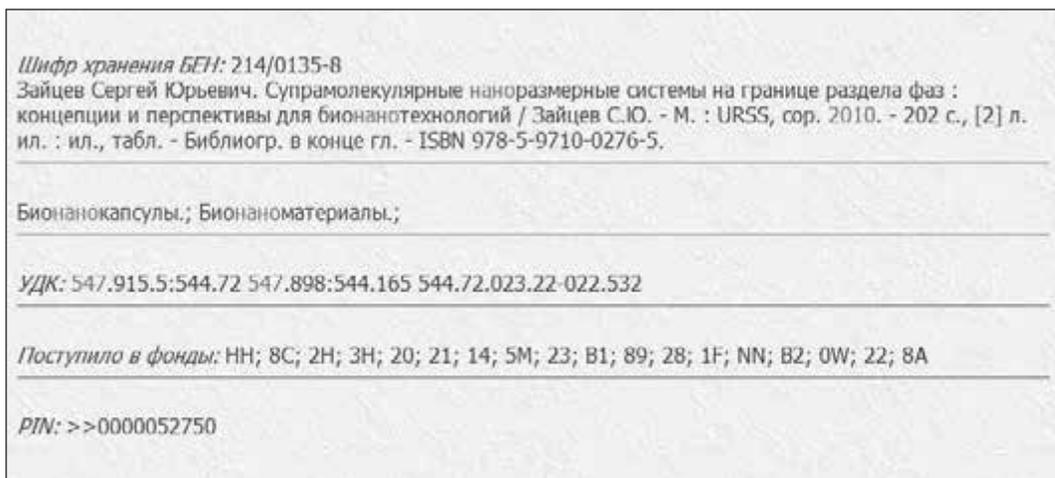


Рис. 4. Уточняющая информация для найденного в каталоге издания

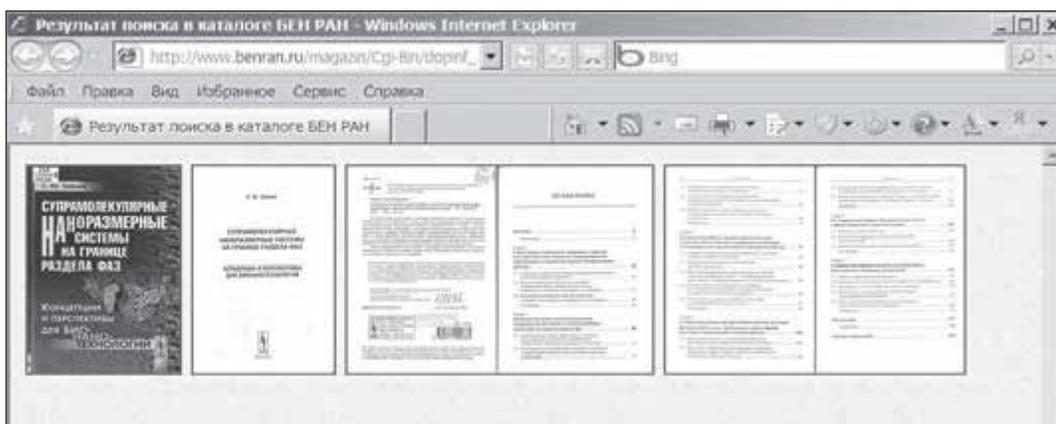


Рис. 5. Отсканированные страницы издания

Результатом выполнения поисковых запросов является информация о найденных в каталоге изданиях, которая выдается на экран порциями (размер которых задается пользователем). Каждая запись представляет собой стандартное библиографическое описание издания (рис. 3). Поисковые термины в найденных записях выделяются красным цветом. Если данное издание находится в фонде ЦБ, то указывается его шифр (шифры) хранения. Кроме того, если у издания в электронном каталоге присутствует информация, уточняющая отдельные элементы библиографического описания (раскрытие инициалов индивидуальных авторов, наименование коллективного автора, данные сводного уровня описания и т. п.), то ее можно получить, перейдя по гиперссылке «Дополнительная информация» (рис. 4). Здесь также перечисляются коды библиотек, в фонды которых поступило издание.

С недавнего времени в БЕН РАН внедрена технология сканирования обложек (и нескольких первых страниц) книг, поступающих в фонд ЦБ. Наряду с обложкой, как правило, сканируются титульный лист, редакционная статья, список рецензентов, оглавление. Полученные образы загру-

жаются на сервер БЕН РАН и доступны пользователям, работающим с интернет-каталогом. Под библиографическим описанием изданий, имеющих образы страниц, находится ссылка в виде раскрытой книги. Нажав на нее, пользователь увидит все отсканированные страницы данного издания (рис. 5). Картинка каждой страницы является активной ссылкой, нажав на которую, пользователь может увеличить ее до необходимого размера, а также последовательно переходить от страницы к странице.

Список источников

1. Васильев А.В. Автоматизированный комплекс SOLAR // Применение ЭВМ в информационно-библиотечной технологии: сб. науч. тр. — М.: БЕН РАН, 1995. — С. 35—38.
2. Он же. Функциональные особенности АИБС «Библиобус» // Новые технологии в информационно-библиотечном обеспечении научных исследований: сб. науч. тр. — Екатеринбург, 2010. — С. 95—99.
3. Власова С.А. Интернет-каталоги БЕН РАН / С.А. Власова [и др.] // Информационные ресурсы России. — 2003. — № 2. — С. 30—34.
4. Она же. Каталог книг и продолжающихся изданий ЦБС РАН в ИНТЕРНЕТ / С.А. Власова, Н.Е. Калёнов // Современные технологии в информационно-библиотечном обеспечении научных исследований: сб. науч. тр. — М.: БЕН РАН, 1999. — С. 100—102.
5. Российский коммуникативный формат [Электронный ресурс]. — Режим доступа: <http://www.rba.ru/rus-marс>

Иллюстративный материал предоставлен авторами статьи.