Автоматизированная система формирования и поддержки базы данных результатов научной деятельности сотрудников академических организаций

С. А. Власова

МСЦ РАН

Аннотация. В статье описывается автоматизированная система создания и поддержки базы данных научных трудов сотрудников академических учреждений, разработанная специалистами МСЦ РАН. Информационная база системы содержит данные о следующих объектах: авторах, связанных с ними организациях (местах их работы), публикациях на аналитическом и монографическом уровнях, источниках (изданиях на сводном уровне — журналах, сборниках), докладах, сделанных на научных конференциях, симпозиумах, семинарах. В системе реализован административный модуль, предназначенный для ввода и редактирования данных. Пользовательский модуль системы представляет собой специальный поисковый аппарат, осуществляющий поиск информации о публикациях, источниках, докладах, мероприятиях, авторах путем обработки поисковых запросов. Отличительной особенностью системы является введенное понятие «эквивалентных» объектов. Эквивалентными считаются объекты, представленные в системе различными метаданными, но относящимися к одной физической сущности. Такими объектами являются «персоны», соответствующие одному автору с различными написаниями фамилии в библиографических описаниях публикаций; организации, имеющие различные варианты названий; статьи, опубликованные без изменений на различных языках.

Ключевые слова: научные труды, библиографические описания, автоматизированная система, база данных, поисковые запросы.

Automated system for creating and supporting a database of research results of academic organizations' employees

S. A. Vlasova

JSCC RAS

Abstract. The article describes the automated system for creating and maintaining a database of scientific works of academic institution's employees, developed by specialists of the Joint Supercomputer Center RAS. The system's information base contains data about objects: the authors, related organizations (places of their work), publications at the analytical and monographic levels, sources (publications at the summary level – journals, collections), reports made at scientific conferences, symposia, seminars. The system has an administrative module designed to enter and edit data. The user's module of the system is a special search engine that searches for information about publications, sources, reports, events, authors by processing search queries. A distinctive feature of the system is the introduced concept of «equivalent» objects. Such objects are «persons» corresponding to the same author with different spellings of the last name in the bibliographic descriptions of publications; organizations with different versions of names; articles which are published without changes in different languages.

Keywords: scientific works, bibliographic descriptions, automated system, database, search queries.

В настоящее время перед каждой научной организацией стоит задача учета научных трудов ее сотрудников (публикаций, докладов, авторских свидетельств и т. п.). В организациях Российской академии наук выполнение этой задачи обеспечивали информационно-библиотечные подразделения, которые вели картотеки трудов сотрудников, содержащие данные об их научных трудах. С развитием вычислительной техники на смену традиционным картотекам пришли информационно-аналитические системы, поддерживающие базы данных (БД) трудов сотрудников и обеспечивающие к ним доступ пользователей через Интернет [1–6].

В 2017 г. в Библиотеке по естественным наукам РАН была разработана и внедрена в техническую эксплуатацию типовая автоматизированная система формирования и поддержки базы данных публикаций [7]. Реализация системы была основана на технологии Microsoft ASP.NET 4 на платформе Microsoft .NET Framework в среде разработки Microsoft Visual Studio 2018 с использованием языка программирования C#. БД системы поддерживаются Microsoft SQL Server.

В 2019–2020 гг. система поддержки базы данных научных трудов получила свое дальнейшее развитие в МСЦ РАН. Система обеспечивает создание и поддержку следующих взаимосвязанных объектов:

- публикации на аналитическом и монографическом уровнях;
- источники (издания на сводном уровне, в которых опубликованы статьи);
- доклады;
- научные мероприятия, на которых сделаны доклады;
- персоны (авторы публикаций и докладов);
- организации (места работы персон).

Система состоит из двух модулей: административного и пользовательского.

В административном модуле (<http://dirsmc.ru/bd/adm.aspx>) осуществляются следующие процессы:

- ✓ ввод и редактирование данных об операторах, работающих с системой;
- ✓ ввод новых записей публикаций, докладов, мероприятий, персон, организаций;
- ✓ редактирование метаданных всех объектов;
- ✓ поиск и просмотр зарегистрированных в системе объектов;
- ✓ создание групп эквивалентных записей.

Между объектами в системе реализованы связи типа «эквивалентные записи». Связанные таким образом объекты система воспринимает как одинаковые. Необходимость ввода эквивалентных персон обусловлена тем, что написание фа-

миллии и имени одного автора в разных библиографических описаниях может отличаться. Например, в англоязычных публикациях персоны «Каленов Н. Е.» можно встретить следующие написания данного автора: «N. Kalenov», «N. E. Kalenov», «Nikolay Kalenov». Эквивалентность публикаций возникает тогда, когда в базе данных отражена статья на языке оригинала и ее полностью переведенные на другой язык версии. Связь эквивалентности для организаций устанавливается, когда организация меняет свое название (например, ИМГ АН СССР и ИМГ РАН).

Рассмотрим процесс ввода в систему публикаций и докладов. Ввод данных новой публикации начинается с ввода ее авторов в том порядке, который представлен в публикации. Для каждого автора проверяется его наличие в системе, в случае отсутствия происходит процесс регистрации новой персоны. В систему вводятся данные автора (фамилия и инициалы, дополнительная информация), и он привязывается к организации, которая либо уже существует в системе, либо выполняется ее регистрации.

После окончания ввода авторов система предоставит форму для ввода метаданных публикации: название публикации, вид издания (статья, монография), источник, год издания, том, номер, страницы, адрес полного текста, идентификаторы во внешних базах данных (см. *рис. 1*). При вводе названия публикации система показывает уже зарегистрированные публикации, в которых находит совпадения в названиях, по первым словам, вводимого названия публикации и которые имеют тех же авторов (см. *рис. 1*). Для привязки к вводимой публикации источника (издания, где опубликована статья), его нужно найти по фрагментам названия, а в случае отсутствия — зарегистрировать в системе (ввести название и дополнительную информацию). После окончания ввода всех необходимых метаданных публикация будет зарегистрирована в системе.

Ввод новой публикации

Власова С.А.

Каленов Н.Е.

Опыт

Название публикации

Власова С.А., Каленов Н.Е. Опыт автоматизации технологических процессов МБА // Информационное обеспечение науки: новые технологии: Сб. науч. тр. / Каленов Н.Е. (ред). - М.: Научный мир, 2009. - 342 с. , 2009. - С. 208-217.

Отказаться от ввода публикации

Вид публикации

статья в журнале

Источник

Поиск

Год

2009

Том

Номер

Страницы

Язык публикации

русский

Идентификатор во внешних базах данных

Адрес полного текста

Ввод

Рис. 1. Регистрация публикации

Регистрация нового доклада так же, как и ввод новой публикации начинается с ввода авторов. Затем система предоставляет форму для ввода метаданных доклада: название доклада, вид доклада (пленарный, секционный, стендовый, приглашенный), мероприятие, дополнительная информация. К докладу нужно привязать мероприятие, на котором был сделан доклад. По фрагментам названия мероприятия определяется его наличие в системе. В случае его отсутствия предоставляется форма для ввода следующих метаданных: название мероприятия, вид мероприятия (конференция, семинар, симпозиум, совещание), место проведения мероприятия (город и страна), даты проведения мероприятия, ссылка на сайт мероприятия, ссылка на материалы (см. рис. 2).

212

Ввод нового мероприятия

Название мероприятия

8-я Международная научно-практическая конференция "Научное издание международного уровня – 2019: стратегия и тактика управления и развития"

Вид мероприятия

конференция ▾

Место проведения (страна)

Россия

Место проведения (город)

Москва

Даты мероприятия

год 2019 с 24 . 04 по 26 . 04 (число.месяц)

Ссылка на сайт мероприятия

https://conf.rasep.ru/WCSP/WCSP2019

Ссылка на материалы

Дополнительная информация

Ввод

Рис. 2. Регистрация мероприятия

Следует отметить, что метаданные организаций в системе представлены в виде иерархической структуры: организация может включать подразделения, в которых есть отделы, которые, в свою очередь, могут включать лаборатории и т. д. Администратор системы вводит название организации, затем названия ее подразделений, далее к каждому подразделению привязывает названия его отделов и т. д. На рис. 3 показан интерфейс для редактирования организаций на примере организации «Научно-исследовательский институт системных исследований (НИИСИ РАН)». Здесь можно корректировать названия организации и подразделений, добавлять (или удалять) подразделения на любом уровне.

ОРГАНИЗАЦИИ

Научно-исследовательский институт системных исследований (НИИСИ РАН)

Межведомственный суперкомпьютерный центр Российской академии наук (МСЦ РАН)

Отдел информационных ресурсов и систем (ОИРС)

Санкт-Петербургское отделение (СПО)

Новое подразделение

Редактировать

Выход

Рис. 3. Административный блок. Редактирование организации

Пользовательский блок системы (<http://dirsmc.ru/bd/>) был создан с учетом опыта предыдущих разработок автора [8, 9]. Он представляет собой поисковый аппарат, осуществляющий поиск информации путем обработки поисковых запросов любой сложности. Для формирования запроса в поисковую форму вводятся термины (слова без знаков препинания) в одну, две или три строки (см. *рис. 4*). В терминах возможно использование правого усечения (символ *). По умолчанию слова связываются оператором «И», который можно поменять на «ИЛИ». Строки по умолчанию связаны оператором «И», который можно поменять на «ИЛИ» и «И НЕ».

Для каждой строки введенных терминов в поисковой форме выбирается наименование области (поля), в которой должен быть осуществлен поиск:

- фамилия персоны,
- название публикации,
- название журнала / сборника,
- название доклада,
- мероприятие (название, страна, город),
- организация,
- идентификатор во внешних базах данных.

- публикации,
- доклады,
- журналы / сборники,
- мероприятия,
- персоны,
- организации.

По умолчанию в поисковой форме выбрана опция «Публикации и доклады». Результат выполнения поисковых запросов выдается на экран порциями, размер которых задается в выпадающем списке «Показывать ... по ...» (по умолчанию — 20 документов на странице).

В том случае, если все поисковые поля формы оставить пустые, после проведения поиска (нажать на кнопку «Поиск») система выдаст все зарегистрированные записи в соответствии с выбранной опцией «Показывать».

На *рис. 4* приведен пример запроса на поиск совместных публикаций авторов «Савин» и «Сотников» за 2010–2016 гг., названия которых содержат термины «научное» и «наследие». По этому запросу система находит 6 публикаций, которые выдаются в виде стандартных библиографических описаний. В описаниях публикаций авторы и названия журналов (сборников) являются активными ссылками. Название публикации будет являться активной ссылкой, в том случае если метаданные публикации содержат адрес на ее полный текст.

Ссылка на названии источника (журнала, сборника) позволяет получить список всех статей, зарегистрированных в системе и опубликованных в данном источнике. При переходе по ссылке от фамилии автора пользователь получит следующую информацию: место работы (название организации, подразделение, отдел) персоны, все эквивалентные записи персоны, дополнительные данные (идентификаторы в различных БД), а также полный список всех публикаций данного автора (см. *рис. 5*).

Ссылка от названия организации (отделения, отдела) позволяет перейти на статьи всех персон, относящихся к данной организации (отделения, отдела).

Соболевская И.Н.

Научно-исследовательский институт (НИИСИ РАН). / Межведомственный суперкомпьютерный центр Российской академии наук (МСЦ РАН). / Отдел информационных ресурсов и систем (ОИРС)

Публикации: 26

Эквивалентные записи:
I. Sobolevskaya
I.N. Sobolevskaya
Irina Sobolevskaya
Sobolevskaya I.N.
И.Н. Соболевская

Идентификаторы в БД Scopus: 57204961984 РИНЦ: 970897

1 2

☐ N. Kalenov, I. Sobolevskaya, A. Sotnikov, S. Kirillov The use of 3D visualization technology web-collections for the formation of virtual exhibitions // CEUR Proceedings of the 21st Conference on Scientific Services & Internet (SSI-2019), Novorossiysk-Abrau, Russia, September 23-28, 2019., 2020. - Vol. 2543. - P. 382-390.

☐ Н.Е. Каленов, С.А. Кириллов, И.Н. Соболевская, А.Н. Сотников Визуализация цифровых 3D-объектов при формировании виртуальных выставок // Электронные библиотеки, 2020. - Т. 23, - № 3. - С. 418-432.

☐ Н.Е. Каленов, И.Н. Соболевская, А.Н. Сотников Формализация процессов формирования пользовательских коллекций в цифровом пространстве научных знаний // Электронные библиотеки, 2020. - Т. 23, - № 3. - С. 433-450.

☐ N. Kalenov, I. Sobolevskaya, A. Sotnikov 1. Mathematical modeling of the processes of interdisciplinary collections formation in the digital libraries environment // CEUR Proceedings of the 21st Conference on Scientific Services & Internet (SSI-2019), Novorossiysk-Abrau, Russia, September 23-28, 2019., 2020. - Vol. 2543. - P. 391-398.

☐ Nikolay Kalenov, Irina Sobolevskaya, Alexander Sotnikov Hierarchical Representation of Information Objects in a Digital Library Environment // Communications in Computer and Information Science, 2019. - Vol. 1093. - P. 93-104.

Рис. 5. Пример полученной информации о персоне

Система обеспечивает возможность выгрузки необходимых пользователю библиографических описаний публикаций в текстовый файл или в структурированный файл формата CSV. Для этого пользователю нужно отметить интересующие его публикации и воспользоваться кнопкой «Выгрузка записей» внизу страницы (см. рис. 4).

Помимо рассмотренного выше поиска публикаций в пользовательском блоке системы аналогичным образом реализован поиск и просмотр докладов. На рис. 6 представлен запрос на поиск докладов, названия которых содержат термин «3D», сделанных в 2018–2019 гг. сотрудниками МСЦ РАН.

В результате проведенного поиска система выдает список записей найденных докладов. Каждая запись содержит

перечень авторов доклада, название доклада и описание мероприятия (название, даты мероприятия, место проведения мероприятия). Фамилии авторов представляют собой активные ссылки, переход по которым обеспечит выдачу информации об авторе и всех докладах данного автора, зарегистрированных в системе. Ссылка от названия мероприятия обеспечивает переход на сайт мероприятия. Название доклада будет являться активной ссылкой, в том случае если метаданные доклада содержат адрес на публикацию доклада или его презентацию.

БД научных трудов сотрудников МСЦ РАН

Организация: МСЦ РАН

И: И

Название публикации: И

И: И

Название доклада: 3D

И: И

Год: от 2018 по 2019

Показывать: Доклады по 20

Сортировка: Год в обратном порядке

Поиск Сброс

Найдено записей: 3

- ☐ I.N. Sobolevskaya, A.N. Solntikov Principles of 3D Web-collections Visualization, 3rd International Conference on Computer-Human Interaction Research and Applications - CHIRA 2019, 20.09 - 21.09.2019 г., Австрия, г. Вена
- ☐ Сотников А.Н., Соболевская И.Н., Кириллов С.А., Каленов Н.Е. Применение технологии визуализации 3D WEB-коллекций для формирования виртуальных выставок. XX Всероссийская научная конференция "Научный сервис в сети Интернет". 17.09 - 22.09.2018 г., Россия, г. Новороссийск
- ☐ Сотников А.Н., Соболевская И.Н., Кириллов С.А., Чередищенко И.Н. Анализ методов визуализации 3D web-коллекций. XXII научно-практический семинар «Информационное обеспечение науки: новые технологии». 15.10 - 19.10.2018 г., Россия, г. Кунгур

Выгрузка записей: Текст

Рис. 6. Пример запроса на поиск докладов

Так же, как и в случае с публикациями, система позволяет выгружать выбранные записи докладов в текстовом формате и в формате CSV.

Представленная автоматизированная система поддержки базы данных научных трудов в настоящее время функционирует в технологическом режиме в МСЦ РАН. В системе зарегистрированы 325 персон из 46 организаций;

739 статей, опубликованные в 348 изданиях. Кроме того, в этом году начался ввод данных о сделанных докладах на научных мероприятиях. Было зарегистрировано 24 мероприятия и 38 докладов, сделанных в 2018–2020 гг.

Работа выполнена в МСЦ РАН – филиале ФГУ ФНЦ НИИСИ РАН в рамках государственного задания № 0580-2021-0014.

Литература

1. Захарова С. С., Гуреева Ю. А. Научные публикации: от картотеки трудов до библиографических профилей // Библиосфера. 2017. Т. 2. С. 85–89.
2. Мазов Н. А., Гуреев В. Н. Создание базы данных трудов сотрудников организации как основы для достоверного библиометрического анализа // Труды ГПНТБ СО РАН. 2015. № 9. С. 103–109.
3. Василенко Е. А., Панфилов В. И., Жуков Д. Ю., Сивуха Д. В. Разработка наукометрической базы данных публикационной активности ученых РХТУ им. Д. И. Менделеева // Научно-техническая информация. Сер. 1. 2013. № 11. С. 26–30.
4. Столяров Р. А., Чугреев В. Л. Автоматизированная система учета результатов интеллектуальной деятельности в научной организации // Вопросы территориального развития. 2015. № 6 (26). С. 4.
5. Кочергина Т. А., Исхакова Л. Д., Каленов Н. Е., Якшин М. М. База данных «Труды сотрудников НЦВО РАН» как часть информационно-поисковой системы «Волоконная оптика» // Системы и средства информатики. 2013. Т. 23. № 2. С. 284–296.
6. Левченко О. И., Соловьев А. В. Формирование базы данных публикаций сотрудников Института физики твердого тела РАН // Информационное обеспечение науки: новые технологии: Сборник научных трудов / Каленов Н. Е., Цветкова В. А. (ред.). М.: БЕН РАН, 2015. С. 215–221.
7. Власова С. А. Автоматизированная система поддержки корпоративной базы данных научных публикаций // Программные продукты, системы и алгоритмы. 2018. Вып. 2. С. 42–46.
8. Власова С. А., Каленов Н. Е. Информатика в академической библиотеке // Системы и средства информатики, 2016. Т. 26. № 3. С. 162–178.
9. Власова С. А., Каленов Н. Е. Интернет-каталог Библиотеки по естественным наукам Российской академии наук как специальная информационно-поисковая система, ориентированная на квалифицированного пользователя // Системы и средства информатики, 2019. Т. 29. № 1. С. 86–95.

References

1. Zakharova S. S., Gureeva Yu. A. Nauchnye publikatsii: ot kartoteki trudov do bibliograficheskikh profiley // Bibliosfera. 2017. Vol. 2. P. 85–89.
2. Mazov N. A., Gureev V. N. Sozdanie bazy dannykh trudov sotrudnikov organizatsii kak osnovy dlya dostovernogo bibliometricheskogo analiza // Trudy GPNTB SO RAN. 2015. № 9. P. 103–109.
3. Vasilenko E. A., Panfilov V. I., Zhukov D. Yu., Sivukha D. V. Razrabotka naukometricheskoy bazy dannykh publikatsionnoy aktivnosti uchenykh RKhTU im. D. I. Mendeleeva // Nauchno-tehnicheskaya informatsiya. Ser. 1. 2013. № 11. P. 26–30.
4. Stolyarov R. A., Chugreev V. L. Avtomatizirovannaya sistema ucheta rezul'tatov intellektual'noy deyatel'nosti v nauchnoy organizatsii // Voprosy territorial'nogo razvitiya. 2015. № 6 (26). P. 4.
5. Kochergina T. A., Iskhakova L. D., Kalenov N. E., Yakshin M. M. Baza dannykh «Trudy sotrudnikov NTsVO RAN» kak chast' informatsionno-poiskovoy sistemy «Volokonnaya optika» // Sistemy i sredstva informatiki. 2013. Vol. 23, № 2. P. 284–296.
6. Levchenko O. I., Solov'ev A. V. Formirovanie bazy dannykh publikatsiy sotrudnikov Instituta fiziki tverdogo tela RAN // Informatsionnoe obespechenie nauki: novye tekhnologii: Sbornik nauchnykh trudov / Kalenov N. E., Tsvetkova V. A. (red.). M.: BEN RAN, 2015. P. 215–221.
7. Vlasova S. A. Avtomatizirovannaya sistema podderzhki korporativnoy bazy dannykh nauchnykh publikatsiy // Programmnye produkty, sistemy i algoritmy. 2018. Iss. 2. P. 42–46.
8. Vlasova S. A., Kalenov N. E. Informatika v akademicheskoy biblioteke // Sistemy i sredstva informatiki, 2016. Vol. 26, № 3. P. 162–178.
9. Vlasova S. A., Kalenov N. E. Internet-katalog Biblioteki po estestvennym naukam Rossiyskoy akademii nauk kak spetsial'naya informatsionno-poiskovaya sistema, orientirovannaya na kvalifitsirovannogo pol'zovatelya // Sistemy i sredstva informatiki, 2019. Vol. 29. № 1. P. 86–95.