

**Обработка сведений о научных мероприятиях:
инструментальные и технологические средства**

**Hardware and Technological Tools
for Processing Data on Scientific Events**

**Обробка відомостей про наукові заходи:
інструментальні й технологічні засоби**

Пожидаев А. В.

*Всероссийский институт научной и технической информации РАН
(ВИНИТИ РАН), Москва, Россия*

A. Pozhidaev

*Russian Institute of Scientific and Technical Information of the Russian Academy of Sciences,
Moscow, Russia*

Пожидаєв А. В.

*Всеросійський інститут наукової і технічної інформації РАН
(ВІНІТІ РАН), Москва, Росія*

Представлена подсистема автоматизированной системы комплектования и регистрации входного потока научно-технической литературы ВИНТИ, предназначенная для сбора, обработки, систематизации и представления в выходных информационных продуктах ВИНТИ описаний научных мероприятий. Рассмотрены наиболее важные решения: формализация, структура информационного массива описаний научных мероприятий и разработанные средства автоматизации ручного труда пользователей подсистемы.

The report introduces a subsystem of the system of automated acquisition and registration of the inflow of the scientific and technical publications of the Institute of Scientific and Technical Information. This subsystem is used for collecting, processing, cataloguing and presenting descriptions of scientific events in the Institute's output information products. The most significant decisions concerning formalization, the structure of the information array of the descriptions of scientific events and automation of the manual labor of the users of this subsystem are presented.

Представлено підсистему автоматизованої системи комплектування та реєстрації вхідного потоку науково-технічної літератури ВІНІТІ, яка призначена для збирання, обробки, систематизації та представлення у вихідних інформаційних продуктах ВІНІТІ описів наукових заходів. Розглянуто найважливіші рішення: формалізація, структура інформаційного масиву описів наукових заходів та розроблені засоби автоматизації ручної праці користувачів підсистеми.

Число научных конференций, симпозиумов, семинаров, выставок, проводимых по всему миру, точно определить невозможно. Однако даже самые грубые подсчеты дают впечатляющую цифру в 4-5 тысяч научных мероприятий ежегодно. Примерно столько же мероприятий анонсируется в периодических изданиях, рассылками по почте, объявлениями в сети Internet. Можно утверждать, что любой ученый, аспирант, студент заинтересован если не в участии, то, по крайней мере, в получении информации, которая будет опубликована по результатам научных мероприятий, проводившихся по интересующей его тематике. Но в условиях значительного количества конференций задача отслеживания необходимых мероприятий становится весьма сложной. Дополнительную сложность в процессе сбора информации о научных мероприятиях представляет то, что каждый организатор использует свой собственный стиль оформления анонса мероприятия. Произвольное оформление анонсов вызвано не только разнообразием проводимых мероприятий, но и отсутствием четких стандартов на оформление анонса мероприятия, не только на международном уровне, но и на уровне отдельных государств.

До появления Интернета единственным источником сводных, систематизированных сведений о научных мероприятиях были печатные издания, наиболее ценными среди них – сводные обобщающие материалы. Практически все информационные центры издают или издавали подобные

информационные материалы, научные общества и организации рассылали списки планируемых мероприятий. В ВИНТИ таким изданием является, издающийся с 1960 г., «Бюллетень международных научных съездов, конференций, конгрессов, выставок».

Формирование данного информационного продукта осуществлялось в ВИНТИ в ручном режиме, вплоть до 2005 года, когда была внедрена в опытную эксплуатацию подсистема обработки сведений о научных мероприятиях автоматизированной системы комплектования и регистрации входного потока НТЛ ВИНТИ.

Основным информационным объектом созданной подсистемы является описание научного мероприятия. Следуя общепринятой терминологии, было принято, что: *Научное мероприятие – это научная конференция, симпозиум, семинар, выставка и т.п., проводившееся или планирующееся к проведению.*

В исходном описании любого научного мероприятия чаще всего содержатся лишь три обязательных элемента – оригинальное название, место и дата проведения. Однако для обеспечения эффективного поиска и систематизации такой ограниченный набор элементов недостаточен, поэтому было принято решение дополнить перечень элементов описания научного мероприятия, как элементами, несущими дополнительную информацию о мероприятии, так и элементами-классификаторами, определенными в процессе анализа существующих методологий описания мероприятий в печатных и электронных изданиях и необходимыми для реализации задач поиска и систематизации. Расширенный перечень элементов описания выглядит следующим образом:

- наименование мероприятия – оригинальное;
- порядковый номер к наименованию;
- наименование мероприятия на русском языке (в том случае, если язык библиографического описания не русский);
- наименование мероприятия параллельное (для редких языков);
- дата проведения мероприятия: начало – год, месяц, день и окончание – год, месяц, день;
- страна проведения;
- место проведения мероприятия (населенный пункт);
- организатор (ы) мероприятия;
- тип мероприятия;
- форма проведения мероприятия;
- язык библиографического описания;
- язык параллельного библиографического описания;
- характер мероприятия;
- географический охват мероприятия;
- состав участников;
- адрес проведения мероприятия (не населенный пункт и (или) конкретное место в населенном пункте);
- рабочий язык;
- ключевые слова;
- краткое резюме;
- материалы мероприятия;
- тематики мероприятия;
- секции мероприятия;
- связи с описаниями других мероприятий;
- актуальность описания.

Таким образом, описание мероприятия содержит 24 элемента. Безусловно, заполнение всех элементов не обязательно и определяется в каждом конкретном случае полнотой и доступностью в источнике сведений о мероприятии.

Основополагающим разделением мероприятий был выбран признак серийности – мероприятие может быть представлено в массиве либо *разовым*, либо *серийным*.

Помимо разделения «разовое»/«серийное», в процессе обработки сведений о мероприятиях возникла необходимость в еще одном типе мероприятия – *обобщенное мероприятие*. Обобщенное мероприятие – это описание мероприятия, никогда не проводившегося в реальности, но необходимое для определения последовательности мероприятий.

Подробнее о формализации, структуре и технологической схеме информационного массива описаний научных мероприятий можно прочитать в статье «Систематизация и использование сведений о научных мероприятиях в автоматизированной технологии ВИНТИ» в журнале НТИ за май 2006 года, а в данном докладе мне бы хотелось осветить по возможности, более подробно разработанные подходы к автоматизации обработки информации о научных мероприятиях, ввиду того, что данная задача является весьма сложной, но очень перспективной, как любая технология, связанная с элементами искусственного интеллекта.

Основным инструментом формирования и поддержания массива научных мероприятий в подсистеме обработки информации о научных мероприятиях является программа взаимодействия с массивом мероприятий. Ее основными задачами являются создание, редактирование и удаление описаний мероприятий, однако ввиду большого количества информации, которое необходимо ввести в описание каждого мероприятия и значительной интеллектуальной работы при определении классификаторов, ручной ввод информации весьма затратен по времени (см. Рис.1).

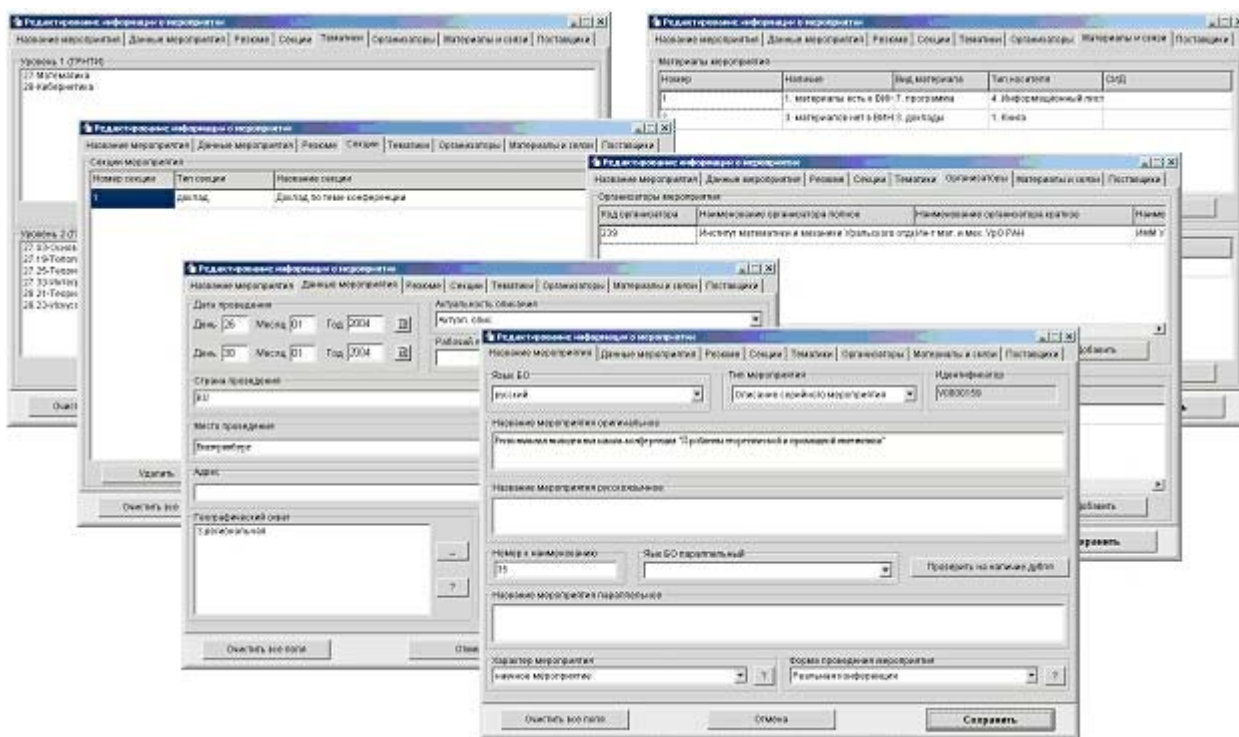


Рис. 1. Интерфейс пользователя при вводе сведений о мероприятии

Автоматизация процесса обработки информации о научных мероприятиях осуществлялась для всех этапов процесса:

- этап ввода сведений о мероприятии;
- этап обработки информации, получение дополнительной информации и определение классификаторов;
- этап формирования информационных продуктов на основе введенной ранее информации.

Для автоматизации процесса ввода сведений о мероприятиях, в частности ввода анонсов научных мероприятий, размещенных в сети Интернет, разработано программное обеспечение, позволяющее реализовать технологию *маркирования* – выделения информации в анонсах мероприятий

цвето-стилистическими схемами, что позволяет осуществлять автоматизированный ввод маркированной информации в массив мероприятий. После разметки соответствующих элементов информации о мероприятии, пользователь может проконтролировать результат распознавания разметки, дополнить и (или) исправить результат маркирования и нажатием одной кнопки передать информацию в массив мероприятий в базе данных ВИНИТИ (см. Рис.2).

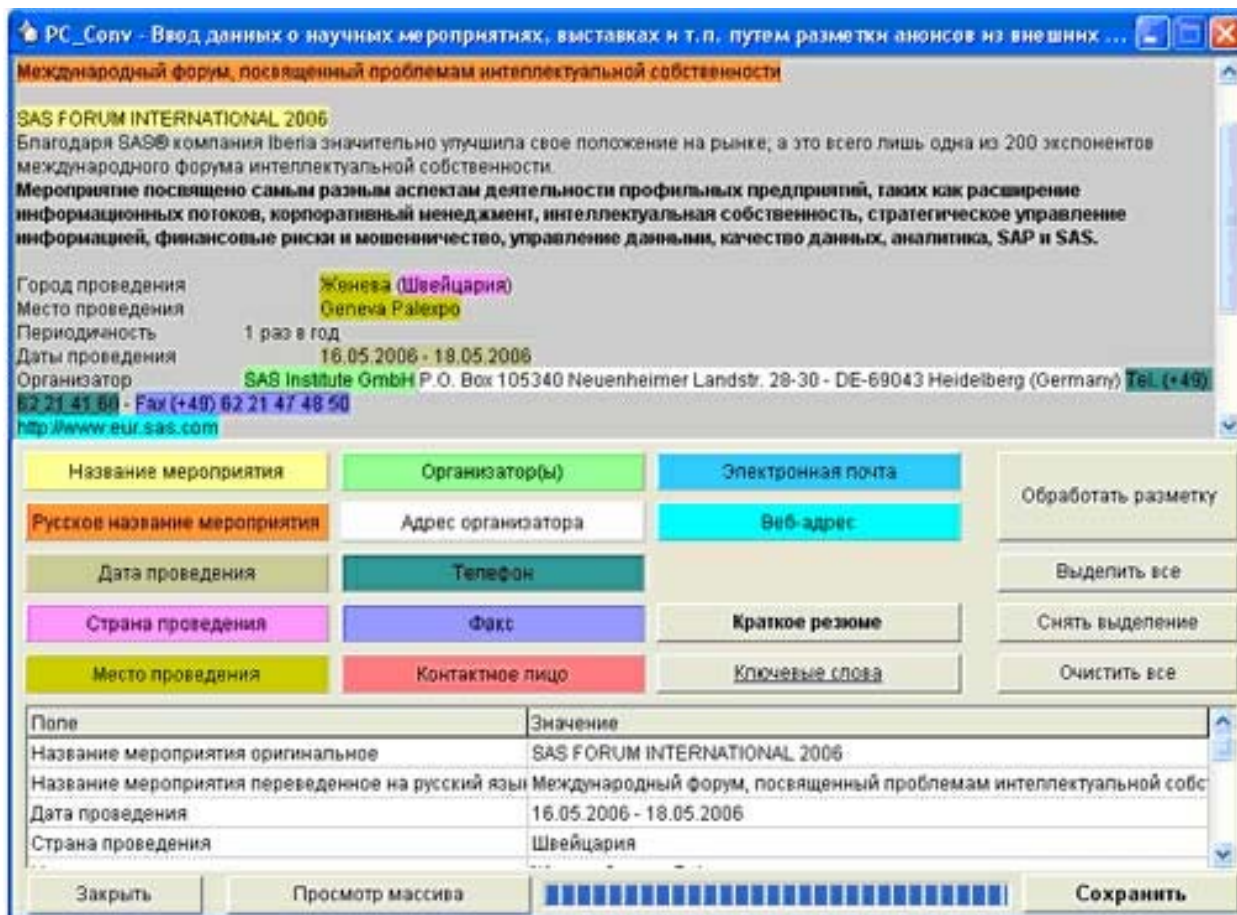


Рис. 2. Интерфейс пользователя при маркировании информации о мероприятии

Для автоматизации обработки информации и определения классификаторов используется обширный массив словарей (26 штук), содержащий не только основные классификаторы, но и элементы названия, указывающие на них, а также перевод этих элементов на основные европейские языки, например, если в названии конференции упоминается «Annual meeting», то автоматически будет определено, что данное мероприятие является серийным. Большой объем подстановок, которые необходимо провести при таком подходе компенсируется быстродействием современных компьютеров – таким образом, пользователь практически не ощущает задержки. Создание подобных словарей является одной из основных задач ВИНИТИ, и работа по наполнению этих словарей продолжается до сих пор.

Для автоматизации процесса формирования информационных продуктов на основе введенной ранее информации, разработано программное обеспечение, позволяющее реализовать выгрузку информации в файлы Microsoft Word на основе имеющихся шаблонов (см. Рис.3) с заданием оформления и простановкой символов иных языковых раскладок.

Европейский научный открытый форум (Euroscience Open Forum (ESOF-2006))		
69	<i>15 - 19 июля</i>	Wissenschaft im Dialog GmbH
	Мюнхен Hotel Bayerischer Hof	Siemin Beyersdorf Tel: +49 (0)30 20 67 29 94; fax + 49 (0)30 20 64 92 05; e-mail: siemin.beyersdorf@w-i-d.de http://www.esof2006.org
3 Европейский симпозиум по ЯМР-спектроскопии в почвоведении, геонауках и экологии (3 European Symposium on Nuclear Magnetic Resonance (NMR) Spectroscopy in Soil, Geo and Environmental Sciences)		
70	<i>6 - 9 августа</i>	Technischen Universität München
	Фрайзинг Kardinal-Dopfner-Haus	Brigitte Eberle Fax +49 8161 714466; e-mail: nmr06@wzw.tum.de http://www.wzw.tum.de/bk/nmr06/
Греция		

Рис. 3. Фрагмент оригинал-макета «Бюллетеня международных научных съездов, конференций, конгрессов, выставок» в файле MS Word

Использование Microsoft Word позволило отказаться в нашем случае от использования издательских систем типа Ventura и сделало возможным формирование оригинал-макетов информационных продуктов на любом компьютере практически любым пользователем.

Разработанный технологический процесс «Маркирование информации – автоматизированная обработка – контроль со стороны пользователя – задание шаблона – выгрузка сведений в автоматическом режиме» после небольших доработок может быть применен для самых различных задач, где требуется обработка больших объемов информации.

Литература

1. Шапкин А. В. Автоматизированная система комплектования и регистрации входного потока ВИНТИ. Ч.1 // НТИ. Сер. 1. – 2005. – №3. – С.8-19.
2. Егоров В. С., Пожидаев А. В., Чернобровская Т. Н. Систематизация и использование сведений о научных мероприятиях в автоматизированной технологии ВИНТИ. // НТИ. Сер. 1. – 2006. – №4. – С.17-23.
3. Батюшко А. А., Глобачева Э. Я., Денисова Л.А., Егоров В.С., Омерда В.В., Шапкин А. В. Регистрация материалов научных мероприятий при обработке входного потока ВИНТИ: программно-технологические средства и новые информационные продукты. // НТИ-2002. Информационное общество. Интеллектуальная обработка информации. Информационные технологии. Материалы 6-й международной конференции (Москва, 16-18 октября 2002 г.).– М.: ВИНТИ, 2002.– С. 385-388.
4. Бюллетень международных научных съездов, конференций, конгрессов, выставок. – М.: ВИНТИ, 2005.