

Организация и использование информационных ресурсов

**КИСЕЛЕВА** Надежда Николаевна, доктор химических наук, и.о. заведующего лабораторией полупроводниковых материалов Института металлургии и материаловедения им. А.А.Байкова РАН

**МУРАТ** Дмитрий Павлович – аспирант Московского института электроники и математики (Технического университета)

**СТОЛЯРЕНКО** Андрей Владиславович – аспирант Московского института электроники и математики

**ДУДАРЕВ** Виктор Анатольевич – инженер Института металлургии и материаловедения им.А.А.Байкова РАН

**ПОДБЕЛЬСКИЙ** Вадим Валерьевич – доктор технических наук, профессор Московского института электроники и математики

**ЗЕМСКОВ** Виктор Сергеевич – доктор технических наук, профессор, главный научный сотрудник Института металлургии и материаловедения им.А.А.Байкова РАН

## БАЗА ДАННЫХ ПО СВОЙСТВАМ ТРОЙНЫХ НЕОРГАНИЧЕСКИХ СОЕДИНЕНИЙ «ФАЗЫ» В СЕТИ ИНТЕРНЕТ\*

В настоящее время известны миллионы химических соединений. Доля неорганических соединений среди них составляет около 10%. Однако речь идет о сотнях тысяч неорганических веществ, к которым ежегодно добавляются тысячи новых соединений. Для хранения и поиска обширной информации о неорганических веществах в промышленно развитых странах разрабатываются базы данных (БД). Помимо всемирно известной библиографической системы CAS (Chemical Abstract Service), наиболее широко используются БД ICSD (Inorganic Crystal Structure Database) [1], Pauling File [2], PDF (Powder Diffraction File) [3], CRYSTMET [4], ИВТАН-ТЕРМО [5], THERMO-CALC [6], Gmelin-Online Datensystem [7] и т.д., содержащие информацию о кристаллоструктурных, термохимических и теплофизических свойствах десятков тысяч неорганических веществ. В отличие от известных информационных систем по веществам и материалам [1-7] БД «Фазы» по свойствам неорганических соединений с самого начала создавалась как база данных для разрабатываемой информационно-аналитической системы, предназначенной для конструирования неорганических соединений [8-11]. Возможны два пути использования баз данных: локальный вариант на CD-ROM и доступ из глобальной сети Интернет. Второй путь является более сложным для программной реализации, но именно он обеспечивает доступ к ежедневно обновляемым базам данных и новым версиям программ конструирования неорганических соединений. Разработка версии БД «Фазы», доступной специалистам в сети Интернет, является первым шагом в создании информационно-прогнозирующей системы.

### 1. КОНЦЕПТУАЛЬНАЯ СТРУКТУРА БАЗЫ ДАННЫХ «ФАЗЫ»

Первая версия БД «Фазы» была разработана в начале восьмидесятых годов [8] для ЭВМ БЭСМ-6. Она содержала информацию о кристаллохимических, термохимических, сверхпроводящих, акустооптических, электрооптических и нелинейнооптических свойствах соединений. Позже последние три группы свойств были объединены в специальной БД «Кристалл» [12, 13], поэтому они не представлены в более поздних версиях БД «Фазы».

В девяностые годы был разработан локальный вариант БД «Фазы» для персональных компьютеров [10]. Информация БД обновляется ежедневно, поэтому использование локальной версии этой информационной системы, переносимой на дисках на компьютеры пользователей, как правило, 2-3 раза в год, не обеспечивало оперативности доступа к новой информации. Локальный вариант БД был основан на системе управления базами данных (СУБД) ДАТАРЕАЛ [14], которая не предоставляла возможностей работы в сети, поэтому в качестве программной основы новой интернет-версии БД «Фазы» была опробован тестовый вариант СУБД MS SQL Server 2000 [15]. При создании сетевой версии БД «Фазы» разработчики частично изменили структуру БД с учетом возможностей новой СУБД, разработали программы конвертирования и загрузки данных из локальной версии БД, программы ввода и редактирования информации, а также программные средства доступа к БД из сети Интернет.

В настоящее время БД по свойствам неорганических соединений «Фазы» содержит информацию об около 43 тыс. тройных соединений из более 18 тыс. систем. Информация извлечена из 14 тыс. публикаций. Объем БД превышает 460 Мбайт. БД «Фазы» не имеет аналогов по количеству тройных соединений, информа-

---

\* Работа выполнена при поддержке РФФИ, гранты №04-07-90086 и №05-03-39009 ГФЕНа. Авторы благодарят И.Н.Белокурова, И.В.Прокошева и В.В.Хорбенко за помощь в разработке баз данных.

ция<sup>1</sup> о которых хранится в таблицах этой информационной системы. Сведения для формирования БД отбираются из периодических изданий, справочников, монографий, отчетов, а также реферативных журналов. БД содержит информацию о наиболее распространенных свойствах соединений. Анализ литературы показывает, что наиболее часто упоминаемая характеристика неорганических соединений – сингония. Следующей по степени изученности является пространственная группа и число формульных единиц в элементарной ячейке. Наиболее распространенная (по степени изученности) термохимическая характеристика – тип и температура плавления, далее – температура распада соединения в твердой или газообразной фазах. Значительно реже определяется температура кипения. Включение сверхпроводящих свойств фаз выполнено по заказу пользователей ИМЕТ РАН, разрабатывающих новые сверхпроводники. Информация обо всех этих свойствах соединений включена в БД «Фазы» (рис.1).

## 2. ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ИНТЕРНЕТ-ВЕРСИИ БАЗЫ ДАННЫХ

При разработке программного обеспечения (ПО) БД использовалась распределенная модель для хранения данных «клиент-сервер». Такая модель позволяет производить все ресурсоемкие операции на мощном сервере и тем самым уменьшить нагрузку на компьютер пользователя. Комплекс разработанного ПО состоит из двух подсистем: подсистемы удаленного администрирования, которая предоставляет разработчикам удобные средства обновления информации, и подсистемы удаленного доступа пользователей к БД из сети Интернет.

Подсистема доступа к БД «Фазы» через сеть Интернет использует трехкомпонентную модель: «Сервер БД» - «Web-сервер» – «Пользователь» [16, 17]. Основное преимущество этой модели заключается в разделении функций сервера БД и Web-сервера, что обеспечивает системе большую гибкость работы и устойчивость.

Сервер БД осуществляет хранение данных и управление данными. Под управлением СУБД Microsoft SQL Server 2000 хранятся таблицы данных о системах, свойствах фаз и таблицы литературных ссылок, а также вспомогательная информация, например, список пользователей системы или список литературных ссылок, для которых в БД хранятся полные тексты статей. Для интеграции Web-сервера и сервера БД использован объектный интерфейс ActiveX Data Objects (ADO) [17].

Web-сервер предназначен для обеспечения доступа к информации БД из сети Интернет и представления ее в удобном для пользователя виде, для чего применяются технологии Internet Server Application Programming Interface (ISAPI) и Active Server Pages (ASP). В качестве Web-сервера используется Microsoft Internet Information Server 4.0 на платформе Windows 2003 Server. На Web-сервере содержатся ASP-документы (страницы) и ISAPI-приложения, с помощью которых запросы пользователей пересылаются серверу БД. Данные от сервера БД оформляются в виде HTML-страниц и посылаются пользователям.

## 3. СИСТЕМА ПОИСКА ИНФОРМАЦИИ В БАЗЕ ДАННЫХ

Для удобства пользователей возможные запросы к БД запрограммированы, и для поиска информации нужно выбрать соответствующий пункт меню и ввести параметры запроса в определенную экранную форму. Например, реализованы такие запросы:

- выдать информацию обо всех тройных соединениях в заданной химической системе;
- выдать список всех соединений с определенным соотношением компонентов;
- выдать информацию обо всех соединениях определенного состава с заданным типом кристаллической структуры в системах с одним или двумя фиксированными элементами;
- выдать все соединения, температура конгруэнтного плавления которых выше (ниже) заданного значения, в системах с одним или двумя фиксированным элементом;
- выдать список всех тройных систем, соединения в которых изучены определенным автором
- и т.п.

Найденная информация выдается на экран в виде привычных для химиков и материаловедов таблиц. Для большинства англоязычных ссылок последних лет возможен просмотр полных текстов статей в pdf-формате.

## 4. РАСПРЕДЕЛЕННАЯ СИСТЕМА БАЗ ДАННЫХ ИМЕТ РАН

Основные идеи, на основе которых проектировались базы данных, разработанные в ИМЕТ РАН, следующие. Базы данных разбиваются на два типа: БД, содержащие наиболее распространенные сведения о неорганических объектах (веществах и физико-химических системах), и БД, в которых содержится самая подробная информация, оцененная высококвалифицированными экспертами, о практически важных веществах. База данных первого типа – это БД «Фазы» по свойствам неорганических соединений, содержащая информацию о термохимических и кристаллохимических параметрах соединений. БД второго типа – это, например, БД по фазо-

<sup>1</sup> Тройные соединения – неорганические соединения, образованные тремя химическими элементами. Тройные соединения – неорганические соединения, образованные тремя химическими элементами.

вым диаграммам систем с полупроводниковыми соединениями «Диаграмма» [18, 19] или БД по свойствам акустооптических, электрооптических и нелинейнооптических веществ «Кристалл» [12, 13].

В настоящее время БД «Фазы» содержит краткую информацию о наиболее распространенных свойствах тройных соединений и систем (рис.1). Более подробные данные о соединениях, обладающих полупроводниковыми, акустооптическими, электрооптическими или нелинейнооптическими свойствами, содержатся в разработанных в ИМЕТ РАН базах данных по фазовым диаграммам полупроводниковых систем «Диаграмма» [18, 19], по свойствам акустооптических, электрооптических и нелинейнооптических веществ «Кристалл» [12, 13] и ширине запрещенной зоны полупроводников «Bandgap». Например, в БД «Фазы» содержится краткая информация о ниобате лития. Более детальные сведения хранятся в БД «Кристалл». Краткие сведения о соединении  $\text{Bi}_2\text{Te}_2\text{Se}$ , включенные в БД «Фазы», могут быть дополнены подробной информацией из БД «Диаграмма» о системе Bi-Se-Te и о свойствах этого соединения, достоверность которой оценена высококвалифицированными российскими специалистами, а также о ширине запрещенной зоны этого соединения из БД «Bandgap». Для того чтобы предоставлять пользователю суммарную информацию обо всей совокупности свойств соединения, хранящейся в разных БД, была проведена интеграция разработанных в ИМЕТ баз данных. Интеграция выполнена на основе использования специальной метабазы [20]. Теперь пользователь может просмотреть весь массив данных о конкретном веществе или материале, хранящийся в разных базах данных виртуальной объединенной системы.

В результате интеграции была создана распределенная неоднородная система баз данных, в которой отдельные БД находятся на разных серверах и под управлением различных СУБД (Oracle (OC Solaris-2.5) или MS SQL Server (OC MS Windows 2003 Server)). Разработанная система является одним из первых опытов интеграции БД по свойствам неорганических веществ и материалов в России. Интегрированная система дает пользователю значительно более полное представление о свойствах веществ и материалов и их зависимостях от различных факторов, чем отдельные базы данных. Дальнейшее развитие распределенной информационной системы возможно по пути интеграции БД ИМЕТ с другими российскими и зарубежными БД.

Сейчас в БД «Фазы» вводится информация о тысячах четверных соединений. Начаты работы по оснащению разработанной БД системой конструирования новых неорганических соединений. Для решения последней задачи будет создана информационно-аналитическая система [11], основанная на методах искусственного интеллекта.

В настоящее время разработана новая версия БД по свойствам неорганических соединений «Фазы», которая доступна зарегистрированным пользователям глобальной сети Интернет (адрес БД в системе Интернет – <http://imet-db.ru/phases>). БД содержит информацию о термохимических, кристаллохимических и других свойствах соединений. В настоящее время собрана, систематизирована и введена в БД информация о свойствах около 43 тыс. тройных соединений из более 18 тыс. систем. Информация, обновляемая ежедневно, извлечена из 14 тыс. публикаций, полные тексты которых доступны для большинства англоязычных статей последних лет. Объем БД превышает 460 Мбайт. БД оснащена электронной библиотекой, в которой хранятся около 2 тыс. оригинальных публикаций в pdf-формате. БД «Фазы» интегрирована с базами данных по фазовым диаграммам полупроводниковых систем «Диаграмма», по свойствам акустооптических, электрооптических и нелинейнооптических веществ «Кристалл» и ширине запрещенной зоны полупроводников «Bandgap», разработанными в ИМЕТ РАН. Планируется интеграция БД ИМЕТ РАН с другими российскими информационными системами и создание распределенной системы БД по свойствам неорганических веществ и материалов. Начаты работы по вводу в БД «Фазы» массива информации по четверным соединениям и объединению БД с системой конструирования неорганических соединений, основанной на использовании методов искусственного интеллекта, что позволит использовать хранящуюся информацию для компьютерного конструирования новых неорганических соединений.

#### Литература:

1. Belsky A., Hellenbrandt M., Karen V.L., Luksch P. New developments in the Inorganic Crystal Structure Database (ICSD): accessibility in support of materials research and design // Acta Crystallogr. - 2002. - V.B58. - №3. - P.364-369.
2. <http://www.crystalimpact.com/pauling>.
3. Faber J., Fawcett T. The Powder Diffraction File: present and future // Acta Crystallogr. - 2002. - V.B58. - №3. - P.325-332.
4. Wood G.H., Rodgers J.R., Gough S.R., Villars P. CRYSTMET – The NRCC metals crystallographic data file // J.Res.NIST. - 1996. - V.101. - №3. - P.205-215.
5. Гурвич Л.В. ИВТАНТЕРМО – автоматизированная система данных о термодинамических свойствах веществ // Вестн. АН СССР. - 1983. - №3. - С.54-65.
6. Andersson J.-O., Jansson B., Sundman B. THERMO-CALC: a data bank for equilibria and phase diagram calculations // CODATA Bull. - 1985. - №58. - P.31-35.

7. Kloffer M. Gmelin-Online Datensystem. Ablauf der dezentralen Datenerfassung für das Gmelin-Online-System von der Diskette zur Datenbank // Software-Entwickl. Chem.1: Proc.Workshops Comput.Chem., Hochfilzen (Tirol), 19-21 Nov., 1986, Berlin e.a., 1987. - P.99-102.
8. Савицкий Е.М., Киселева Н.Н., Пищик Б.Н. и др. Разработка автоматизированного банка данных по свойствам тройных неорганических фаз // Докл. АН СССР. - 1984. - Т.279. - №3. - С.627-629.
9. Киселева Н.Н., Кравченко Н.В. Банк данных по свойствам тройных неорганических соединений как основа для компьютерного конструирования новых веществ // Журн.неорган.химии. - 1992. - Т.37. - №3. - С.698-702.
10. Киселева Н.Н., Кравченко Н.В., Петухов В.В. Банк данных по свойствам тройных неорганических соединений (вариант для IBM PC) // Неорган.материалы. 1996. - Т.32. - №5. - С.636-640.
11. Киселева Н.Н.. Компьютерное конструирование неорганических соединений. Использование баз данных и методов искусственного интеллекта. - М., Наука. - 2005. - 288 с.
12. Юдина Н.В., Петухов В.В., Черемушкин Е.А. и др. Банк данных по свойствам акустооптических, электрооптических и нелинейнооптических веществ // Кристаллография. - 1996. - Т.41. - №2. - С.490-495.
13. Киселева Н.Н., Прокошев И.В., Дударев В.А. и др. Система баз данных по материалам для электроники в сети Интернет // Неорганические материалы. - 2004. - Т.42. - №3. - С.380-384.
14. Киселев Н.Н. Система управления реляционными базами данных ДАТАРЕАЛ-1810 // Управляющие системы и машины. - 1991. - №6. - С.105-109.
15. Тихомиров Ю. Microsoft SQL Server 7.0. Мощнейшая система управления реляционными базами данных. - СПб.: БХВ-Петербург. - 2001. - 702 с.
16. Дунаев С.Б. Доступ к БД и техника работы в сети. Практические приемы современного программирования. - М.: Диалог МИФИ. - 2000. - 416 с.
17. Фролов А.В., Фролов Г.В. Базы данных в интернете: практическое руководство по созданию Web-приложений с базами данных. - М.: ИТД "Русская Редакция". - 2000. - 416 с.
18. Земсков В.С., Киселева Н.Н., Киселев Н.Н. и др. Банк данных по фазовым диаграммам полупроводниковых систем "ДИАГРАММА" // Неорган. материалы. - 1995. - Т.31. - №9. - С.1198-1203.
19. Христофоров Ю.И., Хорбенко В.В., Киселева Н.Н. и др. База данных по фазовым диаграммам полупроводниковых систем с доступом из Интернет // Изв. ВУЗов. Материалы электронной техники. - 2001. - №4. - С.50-55.
20. Korniyushko V., Dudarev V. Software development for distributed system of Russian databases on electronics materials // Proceedings of the Third International Conference "Information Research, Applications and Education - i.TECH 2005", Varna, Bulgaria. – Sofia: FOI-COMMERCE. - 2005. – P.27-32.

Рис.1

