

На правах рукописи

Дударев Виктор Анатольевич

**ИНТЕГРИРОВАННАЯ СИСТЕМА БАЗ ДАННЫХ
ДЛЯ ИНФОРМАЦИОННОЙ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ
ПРИ ПРОГНОЗИРОВАНИИ СВОЙСТВ НЕОРГАНИЧЕСКИХ ВЕЩЕСТВ**

05.13.01 – Системный анализ, управление и обработка информации
(химическая технология)

А В Т О Р Е Ф Е Р А Т

диссертации на соискание ученой степени
кандидата технических наук



Москва – 2006

Работа выполнена на кафедре Информационных технологий государственного образовательного учреждения высшего профессионального образования "Московская государственная академия тонкой химической технологии им. М.В. Ломоносова".

Научный руководитель доктор технических наук, профессор,
заслуженный деятель науки и техники РФ
Корнюшко Валерий Федорович

Официальные оппоненты доктор технических наук, профессор
Кузин Рудольф Евгеньевич

доктор физико-математических наук, профессор
Бублик Владимир Тимофеевич

Ведущая организация Институт химических проблем микроэлектроники
Министерства образования РФ (ИХПМ)

Защита состоится "19" "декабря" 2006 года в 11.00 час. на заседании диссертационного совета Д 212.120.08 при Московской Государственной Академии тонкой химической технологии им. М.В. Ломоносова по адресу: 119571, г. Москва, просп. Вернадского, 86.

С диссертацией можно ознакомиться в библиотеке МИТХТ им. М.В. Ломоносова (119571, г. Москва, просп. Вернадского, 86).
Автореферат диссертации размещен на сайте <http://www.mitht.ru>.

Реферат разослан "17" "ноября" 2006 г.

Ученый секретарь
диссертационного совета,
доктор технических наук



Бурляева Е.В.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность работы

Обеспечение химиков-технологов достоверной информацией о свойствах и технологиях получения современных веществ является необходимым условием развития современной промышленности. На современном этапе качественная информационная поддержка специалистов невозможна без использования специализированных баз данных (БД). Разработка информационных систем (ИС) по свойствам веществ и процессам их получения ведется во всех промышленно развитых странах. Наиболее мощные информационные системы, основанные на современных СУБД, предлагают NIST (National Institute of Standards and Technology – Национальный институт стандартов и технологий, США) и STN (The Scientific and Technical Information Network – Международная сеть научно-технической информации). Как правило, БД по свойствам веществ разрабатываются в разных организациях и даже в разных странах. Полная интеграция таких систем невозможна из-за разного уровня качества данных, хранящихся в разных БД ИС. Обычно она связана и с организационными трудностями, т.к. большинство ИС используются в коммерческих целях или являются открытыми для доступа пользователей только определенных стран или организаций.

В последние годы наблюдается тенденция к кооперации в разработке ИС и к интеграции уже созданных ИС, как на национальном, так и на международном уровне. Актуальность решения этой задачи вызвана стремлением устранить необоснованное дублирование работ и уменьшить затраты на разработку и поддержку ИС. Кроме того, интеграция информации, содержащейся в ИС по свойствам веществ и технологиям их получения, позволяет применять методы компьютерного анализа для поиска взаимосвязей в данных. Использование найденных взаимосвязей позволяет проводить компьютерное конструирование новых перспективных соединений, обладающих заданными свойствами. Получаемая с помощью интегрированной ИС обобщенная информация может быть использована специалистами для поддержки принятия решений при выборе того или иного вещества и технологии его получения для использования в изделиях современной промышленности.

Цель работы

Целью работы является информационная поддержка принятия решений при прогнозировании свойств веществ на основе интеграции разнородных баз данных по свойствам веществ и технологиям их получения.

Для достижения цели работы были поставлены следующие задачи:

- провести анализ современных технологий интеграции разнородных информационных систем;
- осуществить выбор программной платформы для построения интегрированной информационной системы;
- проанализировать и систематизировать архитектуру современных информационных систем по свойствам веществ и технологиям их получения;
- разработать методику построения интегрированной информационной системы с учетом возможности ее использования конечными пользователями и системами поддержки принятия решений;
- разработать структуры данных для применения в интегрированной информационной системе по свойствам веществ;
- разработать интегрированную информационную систему в виде программного комплекса;
- разработать программное обеспечение баз данных по свойствам акустооптических, электрооптических и нелинейнооптических веществ “Кристалл” и по ширине запрещенной зоны неорганических веществ “BandGap”;
- применить созданную интегрированную информационную систему для прогнозирования свойств веществ, перспективных для использования в современной промышленности.

Для достижения этой цели было необходимо найти решение проблем интеграции информационных систем, удовлетворяющее следующим условиям. Решение должно быть:

- масштабируемым, т.е. обеспечивать возможность поэтапного добавления существующих информационных систем;
- достаточно простым для реализации, чтобы на основе предложенной методики любой участник мог самостоятельно разработать программные модули для включения своей информационной системы в интегрированную систему;
- гибким, чтобы учитывать различия в данных и информационных структурах ИС разных организаций;
- мощным, чтобы обеспечить сложные механизмы извлечения и манипулирования данными.

Научная новизна

- предложен комплексный подход к интеграции ИС, как на уровне пользовательских интерфейсов, так и на уровне источников данных;
- на основе теории множеств дано определение релевантной информации в контексте интегрированной ИС по свойствам неорганических веществ;
- разработаны схемы данных и алгоритмы разрешения конфликтов гетерогенности для интегрированной ИС по свойствам веществ и технологиям их получения;
- разработана методика применения интегрированной ИС в программном комплексе компьютерного конструирования химических соединений для прогнозирования свойств веществ.

Практическая значимость

Разработан и внедрен в Институте металлургии и материаловедения им. А.А. Байкова РАН (ИМЕТ РАН) программный комплекс, реализующий интегрированную ИС, объединяющий информационные системы по свойствам веществ и технологиям их получения. При помощи этого программного комплекса выполнена интеграция информационных систем, разработанных ИМЕТ РАН совместно с другими организациями России: БД по свойствам неорганических соединений “Фазы”, БД по фазовым диаграммам полупроводниковых систем “Диаграмма”, БД по свойствам акустооптических, электрооптических и нелинейнооптических веществ “Кристалл”, БД по ширине запрещенной зоны неорганических веществ “BandGap”, БД по свойствам химических элементов “Элементы” и информационной системы по свойствам полупроводникового кремния и процессам его получения и обработки “Кремний”. Полученный информационный комплекс не только позволяет конечным пользователям получать доступ ко всей информации и расчетным подсистемам в рамках интегрированной ИС, но и использовать ИС в качестве источника информации для программ компьютерного конструирования соединений и СППР.

Применение интегрированной ИС позволяет сократить время, затрачиваемое на поиск полной информации по свойствам и технологиям получения веществ.

Полученный программный комплекс используется в учебном процессе кафедры "Материалы микро-, опто- и наноэлектроники" МИТХТ при чтении курсов "Технология полупроводниковых материалов" и "Моделирование процессов полупроводниковой технологии".

Методы исследования

Структуризация и формализация предметной области выполнена на основе методов структурного системного анализа. Для определения релевантной информации в контексте интегрированной ИС и построения модели понятий предметной области использован математический аппарат теории множеств. При разработке интегрированной ИС использованы теория построения БД и Web-технологии. Для иллюстрации использования ИС в интеллектуальных системах использованы методы индуктивного вывода и компьютерного конструирования неорганических соединений, основанные на обучении ЭВМ распознаванию образов.

Апробация работы

Основные результаты работы докладывались и обсуждались на следующих научно-технических конференциях и семинарах: III-rd International Conference, "Information Research, Applications and Education – i.Tech", Bulgaria, Varna, 2005; Всероссийских научных конференциях "Научный сервис в сети Интернет", Новороссийск, 2002, 2005; Международной научной конференции "Гагаринские чтения", Москва, 2002; Научных конференциях молодых специалистов ИМЕТ им. А.А. Байкова РАН (проводятся в рамках Международных научных конференций "Теоретические основы создания металлических сплавов со специальными свойствами"), Москва, 2004, 2005; Первой научной конференции молодых ученых МИТХТ им. М.В. Ломоносова, Москва, 2005; Международных студенческих школах-семинарах "Новые информационные технологии", Украина, Судак, 2002, 2004; Всероссийском семинаре по кинетике физико-химических процессов в газовой динамике (Институт механики МГУ им. М.В. Ломоносова), 2006.

Публикации

Результаты диссертационной работы опубликованы в 14 печатных трудах, в том числе в 3 статьях в журналах, рекомендованных ВАК РФ для опубликования результатов диссертационных работ, 3 статьях, 8 публикациях в сборниках трудов и тезисов докладов конференций и семинаров.

Структура и объем диссертации

Диссертация состоит из введения, четырех глав и заключения, изложенных на 150 страницах, включая библиографию из 163 источников, 41 рисунка и 9 таблиц.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** кратко рассмотрены актуальность и практическая ценность работы. Сформулирована цель работы и поставлены задачи. Кратко рассмотрены методы исследований, научная новизна, результаты апробации и внедрения диссертационной работы. Приведена структура диссертации и краткое содержание основных разделов.

В **первой главе** кратко рассмотрены наиболее значимые БД по свойствам веществ и технологиям их получения, созданные в мире. На рис. 1 дано распределение БД по тематике содержащейся в них информации.

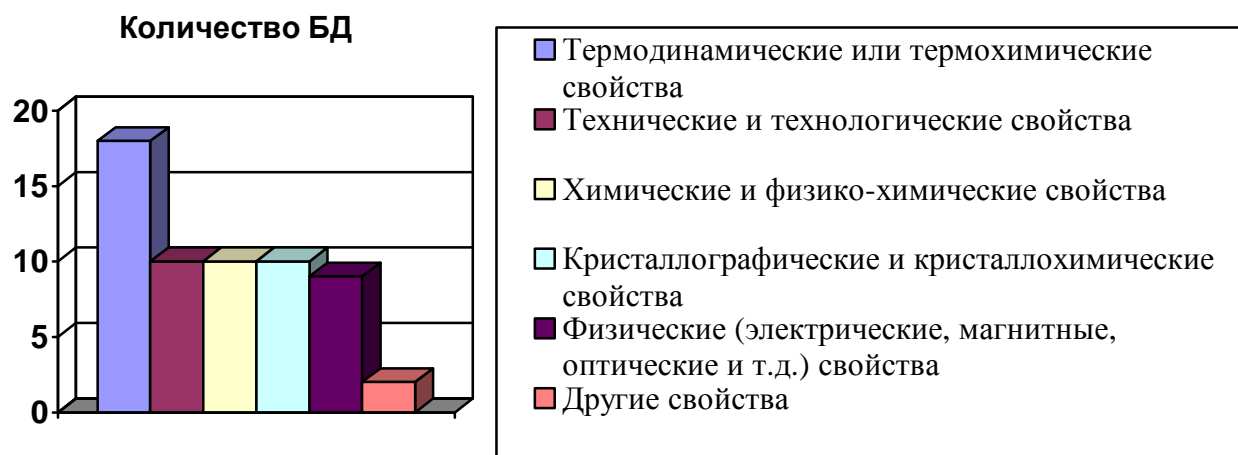


Рис. 1. Распределение БД по свойствам неорганических веществ по тематике.

Детально рассмотрены три подхода к интеграции: (1) интеграция корпоративных приложений (Enterprise Application Integration, EAI), (2) интеграция корпоративной информации (Enterprise Information Integration, EII) и (3) программное обеспечение для извлечения, преобразования и загрузки данных (Extract, Transform, Load – ETL), основанное на технологии хранилищ данных.

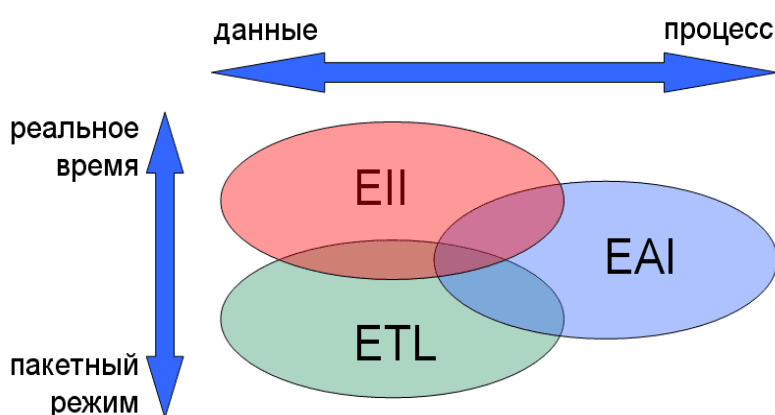


Рис. 2. Современные подходы к интеграции.

Указаны области применения, а также достоинства и недостатки, присущие этим подходам. Отмечено, что ни один из существующих подходов не способен решить все проблемы, возникающие при интеграции ИС, т.е. справиться с объединением информационных источников и прило-

жений ИС по свойствам веществ и технологиям их получения (рис. 2).

Проанализированы информационные потоки при разных методах интеграции ИС по свойствам веществ и технологиям их получения, а также связь интегрированной ИС с системами поддержки принятия решений (СППР) при исследовании и использовании химических веществ для современной промышленности (рис. 3).

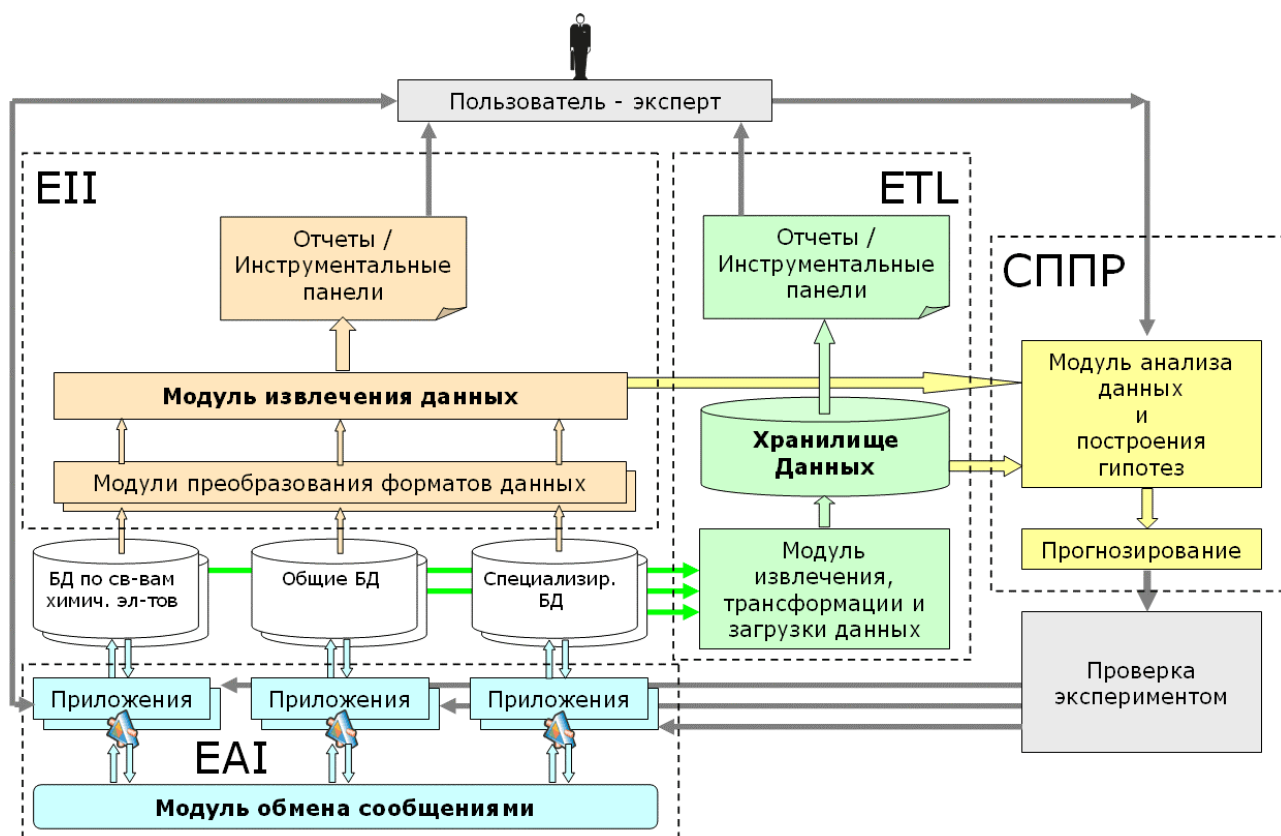


Рис. 3. Анализ информационных потоков при использовании разных подходов к интеграции и место интегрированной ИС в СППР.

Учитывая то, что интегрированную ИС планировалось использовать в качестве источника информации для СППР, был сделан выбор в пользу подхода ЕИ при интеграции на уровне данных. Этот подход, в отличие от ETL, позволяет получать актуальные данные из информационных источников, не требует промежуточного хранилища данных и является более гибким.

Рассмотрены основные подходы к интеграции информации средствами ЕИ: (1) Global-As-View (GAV – описывает глобальную схему предметной области в терминах представлений локальных схем predetermined источников данных) и (2) Local-As-View (LAV – рассматривает схемы локальных источников данных как материализованные представления в терминах общей глобальной схемы предметной области). После рассмотрения достоинств и не-

достатков, присущих GAV и LAV, осуществлен выбор подхода LAV для разработки интегрированной ИС, поскольку он позволяет построить масштабируемую интегрированную ИС с возможностью оперативного подключения новых источников информации. При LAV-интеграции программные адаптеры выполняют функцию преобразования данных из формата информационных источников к глобальной схеме X (рис. 4). Кратко рассмотрены и систематизированы три типа конфликтов гетерогенности, которые должны быть разрешены при разработке интегрированной ИС: (1) платформенные и системные, (2) синтаксические и структурные, (3) семантические.

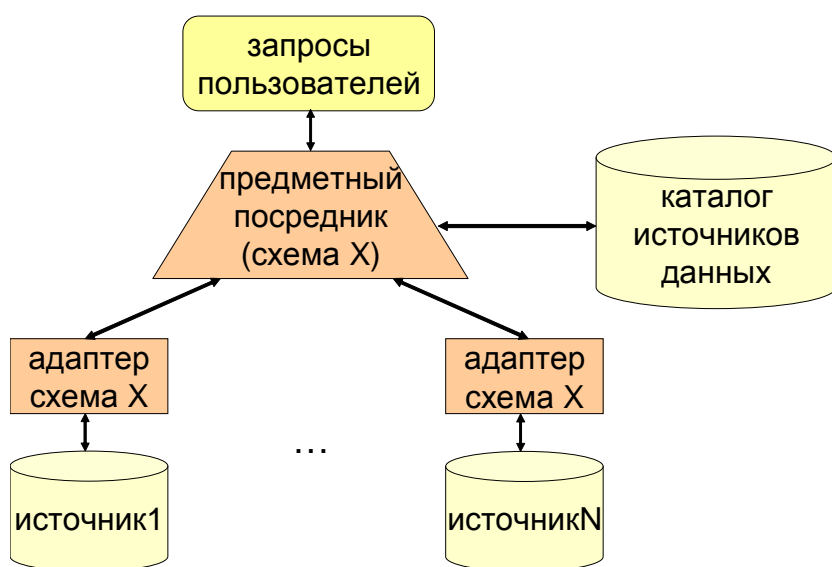


Рис. 4. Интеграция на основе принципа Local-As-View.

проведенного независимыми компаниями, сделан вывод, что в настоящее время Microsoft предлагает наиболее надежные и высокопроизводительные системы, при этом ССВ решений этой компании оказывается ниже, чем у основных конкурентов. Это обусловило выбор решений на платформе Microsoft для реализации интегрированной ИС.

Во **второй главе** приведены результаты разработки методики построения интегрированной ИС. Для этого рассмотрено текущее состояние и принципы построения ИС по свойствам и технологиям получения химических веществ, т.к. учет специфики информационных структур, содержащихся в БД, и их семантики является необходимым условием разработки интегрированной ИС. Кратко рассмотрены ИС ИМЕТ РАН, построенные на различных программно-аппаратных платформах с использованием разных подходов к хранению и обработке информации: ИС по свойствам неорганических соединений “Фазы”, ИС по фазовым диаграммам систем с полупроводниковыми фазами “Диаграм-

Проведен обзор современных программных платформ для разработки интегрированной ИС. Предложены критерии выбора платформ: производительность, безопасность, надежность, интероперабельность и совокупная стоимость владения (ССВ). Используя предложенные критерии и результаты тестирования,

ма”, ИС по свойствам полупроводникового кремния и процессам его получения и обработки “Кремний” и т.д.

Приведены краткие результаты разработки ИС по веществам с особыми акустооптическими, электрооптическими и нелинейнооптическими свойствами “Кристалл” и ИС по ширине запрещенной зоны неорганических соединений “BandGap”, проведенной в рамках диссертационной работы. Описана разработка универсального программного комплекса DBAdmin, позволяющего выполнять удаленное администрирование БД всех ИС ИМЕТ РАН с использованием единого пользовательского интерфейса. Особенности данного комплекса являются: 1) возможность удаленного (по локальной сети или через Интернет) взаимодействия с БД; 2) способность работать с БД произвольной структуры, поскольку структура данных считывается при подключении к информационному источнику; 3) возможность эффективного взаимодействия с БД под управлением разных СУБД (Microsoft SQL Server, Oracle и т.д.) за счет использования OLE DB и ODBC. Отмечается, что использование DBAdmin позволило стандартизировать процедуры администрирования всех БД ИС в рамках ИМЕТ РАН и дало возможность использования этого комплекса для единого управления всеми БД в рамках интегрированной ИС.

При исследовании ИС по свойствам веществ и процессам их получения отмечено присутствие во многих ИС расчетных подсистем, с помощью которых динамически рассчитываются значения тех или иных свойств заданного класса веществ по введенным пользователем параметрам или осуществляется визуализация рассчитанной по некоторым правилам информации. При этом исследователи активно используют такие расчетные подсистемы для получения информации, которая не может быть представлена в табличной форме (рисунки, динамические графики зависимостей и т.п.). Такие расчетные подсистемы используют данные из конкретных информационных источников, и, следовательно, их невозможно применять вне контекста оригинальных ИС, т.к. Web-приложения ИС являются естественным интерфейсом к расчетным подпрограммам. Учитывая важность расчетных подсистем, отмечена необходимость их включения в интегрированную ИС.

Необходимость как можно более тесной интеграции ИС по свойствам веществ и технологиям их получения обуславливает архитектуру построения современных ИС с доступом пользователей через Интернет. Отмечено, что серверная часть ИС разделена на две составляющие: (1) база данных информационной системы (БД ИС); (2) Web-приложение информационной системы (Web-

приложение ИС). Основываясь на архитектурном разделении ИС, выделены два класса подходов к интеграции ИС. Первый класс подходов заключается в интеграции информационных ресурсов на уровне виртуального объединения их гетерогенных источников информации (ЕИ). Второй класс подходов позволяет объединить пользовательские интерфейсы, из которых осуществляется доступ к информационно-расчетным подсистемам, т.е. Web-приложения соответствующих ИС (ЕАИ).

Предложен комплексный подход к интеграции, сочетающий в себе интеграцию на уровне данных и пользовательских интерфейсов (ЕИ+ЕАИ). В рамках предлагаемого подхода предоставляется как доступ к текущим пользовательским интерфейсам ИС и свободное перемещение пользователей между ними (ЕАИ), так и богатые возможности по сбору и агрегации информации, полученной из разнородных распределенных источников данных по свойствам веществ, согласно общей разработанной информационной схеме (ЕИ).

При разработке методов интеграции Web-приложений ИС ставились задачи поиска релевантной информации в интегрируемых ИС и обеспечения прозрачного перехода пользователей между Web-приложениями ИС с соблюдением безопасности. Для реализации механизмов поиска релевантной информации предложено использование базы метаданных (далее метабаза) – специальной БД, содержащей справочные сведения об интегрируемых ИС – и разработана ее структура.

В метабазае содержится информация по интегрируемым информационным системам (множество D), химическим системам (множество S) и их свойствам (множество P). Для описания взаимосвязи между элементами множеств D , S и P определено тернарное отношение W на множестве $U = D \times S \times P$. Принадлежность элемента (d, s, p) отношению W , где $d \in D, s \in S, p \in P$, интерпретируется следующим образом: “в интегрируемой информационной системе d содержится информация по свойству p химической системы s ”.

Поиск релевантной информации по конкретной химической системе s сводится к определению отношения R , являющегося подмножеством декартова произведения $S \times S$ (иными словами, $R \subset S^2$). Таким образом, о любой паре $(s_1, s_2) \in R$ можно сказать, что система s_2 является релевантной системе s_1 . Т.е., чтобы решить задачу поиска релевантной информации в интегрируемых информационных системах, необходимо определить отношение R . Предлагаются следующие правила для построения R :

1) Для любых множеств $s_1 \in S, s_2 \in S$, состоящих из химических элементов e_{ij} , $s_1 = \{e_{11}, e_{12}, \dots, e_{1n}\}, s_2 = \{e_{21}, e_{22}, \dots, e_{2m}\}$ верно, что если $s_1 \subseteq s_2$ (то есть, все химические элементы из системы s_1 содержатся в системе s_2), то $(s_1, s_2) \in R$.

2) Отношение R симметрично. Иными словами, для любых $s_1 \in S, s_2 \in S$ верно, что если $(s_1, s_2) \in R$, то и $(s_2, s_1) \in R$.

В работе приведены и другие правила построения отношения R . Отмечено, что ни одно из определений не является подходящим для решения всех задач по определению релевантной информации в распределенных ИС, и на практике предложены несколько разных отношений релевантности R , которые названы классами релевантности. Отмечается возможность более четкого определения релевантной информации при использовании отношений R вида:

$$R \subset (d_1, s_1, p_1) \times (d_2, s_2, p_2), \text{ где } d_1, d_2 \in D; s_1, s_2 \in S; p_1, p_2 \in P.$$

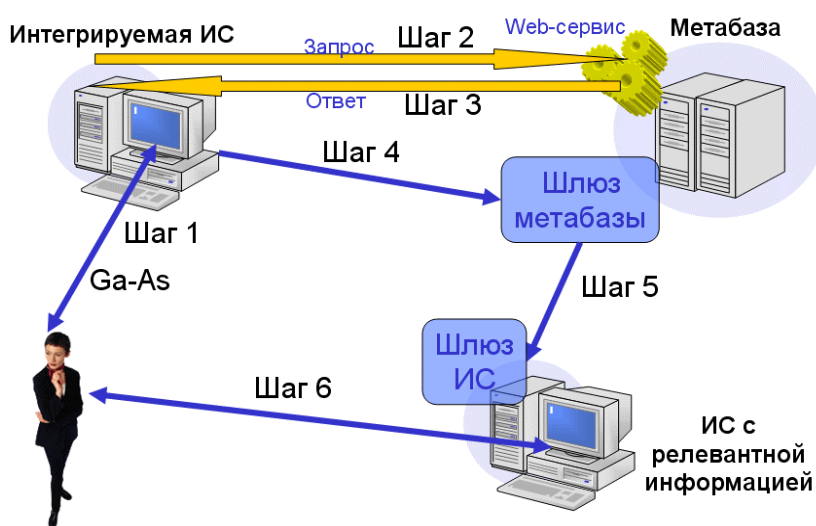


Рис. 5. Интеграция Web-приложений ИС с поиском релевантной информации в метабазе.

Для обеспечения безопасности при переходах пользователей между Web-приложениями ИС предложено использовать систему шлюзов безопасности, санкционирующих переходы пользователей между ИС и обеспечивающих отображение релевантной информации

(рис. 5). “Шлюз метабазы” санкционирует переход между ИС, а шлюз ИС выполняет сопряжение централизованной системы безопасности с системой безопасности интегрируемой ИС.

При разработке методики интеграции разнородных источников данных ИС решены задачи по разработке общей схемы предметной области и разработке путей разрешения конфликтов гетерогенности.

Описание сущностей и их свойств в разных ИС по свойствам веществ происходит с разной степенью детализации. Отмечено, что значения свойств, хранимые в разных информационных источниках, определяются, в первую очередь, составом неорганических веществ (набором образующих их химических элементов и соотношением). В свою очередь, физические свойства веществ во

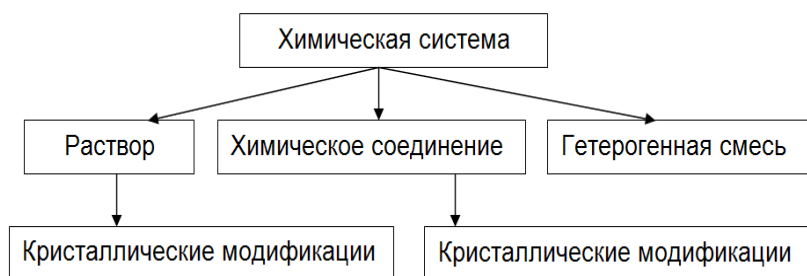


Рис. 6. Иерархия химических понятий.

многим зависят от кристаллической структуры. Поскольку интегрируемые ИС тесно связаны с химической технологией, то сущности в них также могут быть описаны с помощью иерархии понятий в

виде дерева (рис. 6). Обозначив сущности второго уровня общим термином “вещество” получаем трехуровневую иерархию химических объектов: система, вещество и кристаллическая модификация. Вся информация о свойствах химических сущностей, описываемых в интегрируемых информационных источниках, может быть представлена на одном из этих трех уровней. Для детального описания объектов каждого уровня использован математический аппарат теории множеств. Предложенная иерархия объектов используется в контексте интегрированной ИС.

Рассмотрены три типа конфликтов гетерогенности: (1) платформенные и системные, (2) синтаксические и структурные, (3) семантические, и предложены пути их разрешения на основе технологии Web-сервисов, языка XML и введения тезаурусов.

В **третьей главе** описана программная разработка интегрированной ИС в рамках предложенного комплексного подхода, сочетающего в себе интеграцию на уровне пользовательских интерфейсов и источников данных. Рассмотрена реализация подсистемы интеграции пользовательских интерфейсов. Для объединения Web-приложений ИС по свойствам неорганических веществ и технологиям их получения использована следующая структура метабазы (рис. 7). Назначение таблиц: DBInfo – корневая таблица, содержащая информацию об интегрируемых ИС; UsersInfo, UsersAccess – таблицы, содержащие информацию о пользователях интегрированных ИС и их правах доступа к другим интегрированным ресурсам; SystemInfo, PropertiesInfo, DBContent – таблицы, в которых описывается содержимое интегрируемых ресурсов (какая информация по химическим системам и их свойствам содержится в интегрируемых ИС); CompatibilityClasses, Compatibility, Systems2ConsiderInCompatibility – таблицы, содержащие информацию о доступных в метабазе классах релевантности и определяющие релевантные химические системы.

Реализована загрузка информации в метабазу с использованием языка XML и технологии Web-сервисов. Разработка велась в среде Microsoft Visual Studio 2003. Интегрируемые ИС формируют XML-документы, содержащие сведения об информационных изменениях, произошедших в их состоянии. Затем эти данные передаются Web-сервису обновления метабазы, взаимодействие с которым осуществляется по протоколу SOAP, являющемуся стандартным для взаимодействия разнородных программных модулей посредством сети Интернет. После каждого сеанса передачи данных происходит инкрементальное перестроение классов релевантности. Таким образом, в метабазе поддерживается актуальная информация о содержимом интегрируемых ресурсов, способах доступа к ним, пользователях и их правах.

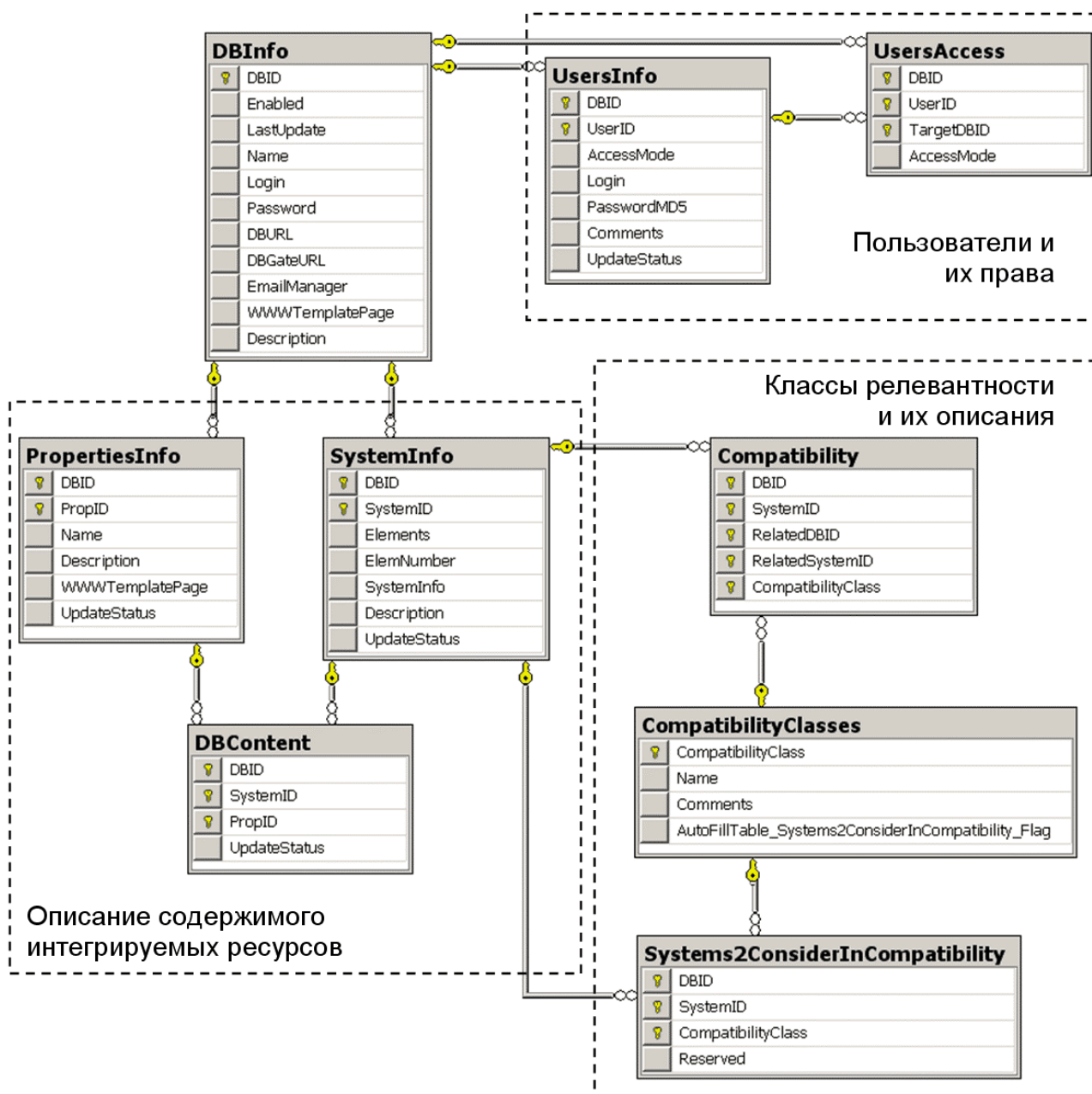


Рис. 7. Структура метабазы для интеграции Web-приложений ИС.

Разработан Web-сервис поиска релевантной информации, который на основе содержимого метабазы предоставляет пользователям интегрируемых Web-приложений ИС ответ на запрос релевантной информации (согласно выбранному классу релевантности). Запрос может инициировать только пользователь, зарегистрированный на уровне метабазы, а ответ, передаваемый в формате XML, содержит список найденной релевантной информации с учетом прав доступа пользователя. Формат этого XML-документа жестко фиксирован с помощью специально разработанного документа XML-schema.

Для облегчения подключения новых информационных систем к интегрированной ИС были разработаны программные компоненты, выполняющие роль посредников при взаимодействии с указанными выше Web-сервисами. Эти посредники управляют всеми аспектами сетевого взаимодействия и обеспечивают безопасность передаваемых данных.

Использование Web-сервиса поиска релевантной информации позволило получать список релевантной информации в XML-документе, который для пользователей Web-приложений, как правило, выводится в виде специальных гиперссылок на интегрированные информационные ресурсы. Для обеспечения прозрачного и безопасного перехода пользователя из контекста одного Web-приложения в контекст другого использована система шлюзов. Переход осуществляется через шлюз метабазы, выполняющий роль диспетчера безопасности, который перенаправляет пользователя (в случае успешной проверки) в шлюз интегрируемой ИС. Задача последнего заключается в выполнении прозрачной авторизации пользователя и предоставлении запрашиваемой информации.

Разработанная методика интегрирования, основанная на использовании шлюзов безопасности, является достаточно гибкой, так как сама политика “фильтрации” содержимого интегрируемого информационного ресурса задается настройками безопасности метабазы, а применяется на шлюзовой странице целевой ИС в контексте конкретного Web-приложения. Соответственно, учитываются модели безопасности, применяемые для информационной защиты конкретных ресурсов. В настоящее время интеграция на уровне Web-приложений проведена для ИС “BandGap”, “Диаграмма”, “Кристалл”, “Фазы”, “Элементы” и “Кремний”. Планируется функциональное расширение интегрированной ИС за счет подключения новых ИС по свойствам веществ и технологиям их получения и введения новых классов релевантности.

Описана реализация подсистемы интеграции источников данных. Для объединения источников данных по свойствам неорганических веществ пред-

ложена следующая структура метабазы: Meta_DBInfo – корневая таблица, содержащая информацию об интегрируемых ИС; Meta_ExpertInfo, Meta_UserInfo, Meta_UserExpert – таблицы, содержащие информацию о пользователях и экспертах ИС и их правах доступа; Meta_SystemInfo, Meta_SubstanceInfo, Meta_ModificationInfo, Meta_ModificationRegisrty, Meta_PropertyInfo – таблицы, предназначенные для разрешения семантических конфликтов на уровне химических систем, веществ, модификаций и свойств; Meta_DBExpert, Meta_PropertyExpert, Meta_SystemExpert, Meta_SubstanceExpert, Meta_ModificationExpert – таблицы, содержащие экспертные оценки качества информации в интегрируемых источниках данных ИС.

В связи с тем, что интегрируемые источники данных могут пересекаться по набору свойств веществ, а качество информации (достоверность и полнота) в каждой ИС отличается для разных свойств, разработан механизм, поддерживающий экспертные оценки интегрируемых данных. Экспертиза проводится высококвалифицированными специалистами, которые выставляют оценки, характеризующие качество данных в разных интегрируемых ИС. Таким образом, при наличии информации по какому-либо физико-химическому свойству в нескольких интегрируемых БД, интегрированная ИС может выдавать не только сами данные, но и степень их достоверности, рассчитанную на основе экспертных оценок. Заметим, что ИС может функционировать и без пользовательских рейтингов и экспертных оценок.

Далее описаны требования к реализации программных адаптеров, которые согласно выбранному подходу Local-As-View осуществляют трансформацию структур данных из внутреннего представления интегрируемого источника к разработанной общей информационной схеме. Согласно разработанной методике, программные адаптеры реализуются в качестве Web-сервисов, имеющих стандартизированное в рамках общей схемы WSDL-описание (Web Services Description Language – язык описания Web-сервисов). Результатом работы адаптеров являются стандартизированные XML-документы, что обеспечивает унифицированную работу предметного посредника со всеми интегрируемыми источниками данных.

Разработан предметный посредник, являющийся точкой входа в интегрированную ИС. Он отвечает на запросы пользователей интегрированной ИС, осуществляя взаимодействие со всеми программными адаптерами, поиск и агрегацию данных. Фрагмент ответа приведен на рис. 8. В настоящее время к

предметному посреднику выполнено подключение источников данных для ИС “BandGap” и “Кристалл”.

Отмечается, что оригинальная методика интегрирования Web-интерфейсов ИС и их информационных источников позволила объединить ИС по свойствам веществ и технологиям их получения. Таким образом, реализована интегрированная ИС, которая может использоваться как конечными пользователями для поиска и сбора информации, так и программными средами в качестве информационного источника по свойствам веществ и технологиям их получения.

```
<?xml version="1.0" encoding="windows-1251" ?>
- <root>
- <row PropertyID="27" Name="Акустооптические свойства"
  DBPropID="crystal.AcOpTabl" Rating="1.0">
- <PropertyXML>
- <row SystemID="1" SubstanceID="86" ModificationID="0" WaveLeng="1.15"
  Nzv="[100]" Uzv="[010]" E="произв." M1="155" M2="46.3" M3="49.2"
  Reference="<B>Ярив А.,Юх П.</B> <I>Оптические волны в кристаллах,
  М.: Мир, 1987, 616 с.</I> //" Rating="1.0">
- <SystemXML>
- <ChemicalSystem>
  <Element>As</Element>
  <Element>Ga</Element>
</ChemicalSystem>
</SystemXML>
- <SubstanceXML>
- <ChemicalSubstanceComposition>
  <Item Element="As" value="1" />
  <Item Element="Ga" value="1" />
</ChemicalSubstanceComposition>
</SubstanceXML>
<ModificationName />
</row>
</PropertyXML>
</row>
</root>
```

Рис. 8. XML-документ с фрагментом ответа предметного посредника интегрируемой ИС.

В четвертой главе отмечено, что интеграция ИС является первым шагом к разработке интеллектуальных ИС. Интегрированная ИС решает проблему поиска затребованных пользователем сведений об определенных веществах и их свойствах в различных интегрируемых источниках данных. Дальнейшая интеллектуализация разработанной интегрированной ИС связана с ее оснащением программными комплексами анализа огромных массивов химической информации и с поиском взаимосвязей в этих данных. Найденные взаимосвязи позво-

лят прогнозировать свойства еще экспериментально неизученных веществ, оценивать их параметры и принимать решение о путях поиска новых веществ с заданными свойствами. Это расширит возможности интегрированной ИС, превращая ее из компьютерного справочника в интеллектуальную ИС.

В рамках диссертационной работы были проведены исследования по использованию данных из интегрированной ИС для поиска сложных взаимосвязей в химической информации и их применению для конструирования новых неорганических соединений, перспективных для использования в электронной промышленности. Физико-химической основой разработки такой системы является Периодический закон, из которого следует, что существуют периодические зависимости между свойствами соединений и свойствами элементов, входящих в их состав. Поскольку все известные соединения, информация о которых хранится в ИС, должны подчиняться Периодическому закону, то, следовательно, возможен поиск закономерностей образования соединений определенных типов с использованием компьютерных методов анализа информации интегрированной ИС.

На основе системного анализа процесса компьютерного конструирования неорганических соединений разработана методика использования интегрированной ИС в качестве источника данных исследовательских систем, используемых для принятия решений при исследовании неорганических соединений (рис. 9).

Прогнозирующая ИС используется, если в интегрированной ИС нет данных по соединениям с нужными свойствами. При этом первый этап компьютерного конструирования – это экспертный анализ информации в специализированных БД с целью выбора соединений-прототипов, свойства которых близки к требуемым. Затем ведется отбор информации об известных аналогах по составу и/или типу кристаллической структуры в общих БД. Каждая система описывается в виде набора значений свойств химических элементов, входящих в её состав. На данном этапе эксперт отбирает свойства, по которым будет вестись анализ. Данные о свойствах химических элементов извлекаются из БД по свойствам химических элементов. Как правило, используется множество самых различных свойств элементов и/или их простых соединений.

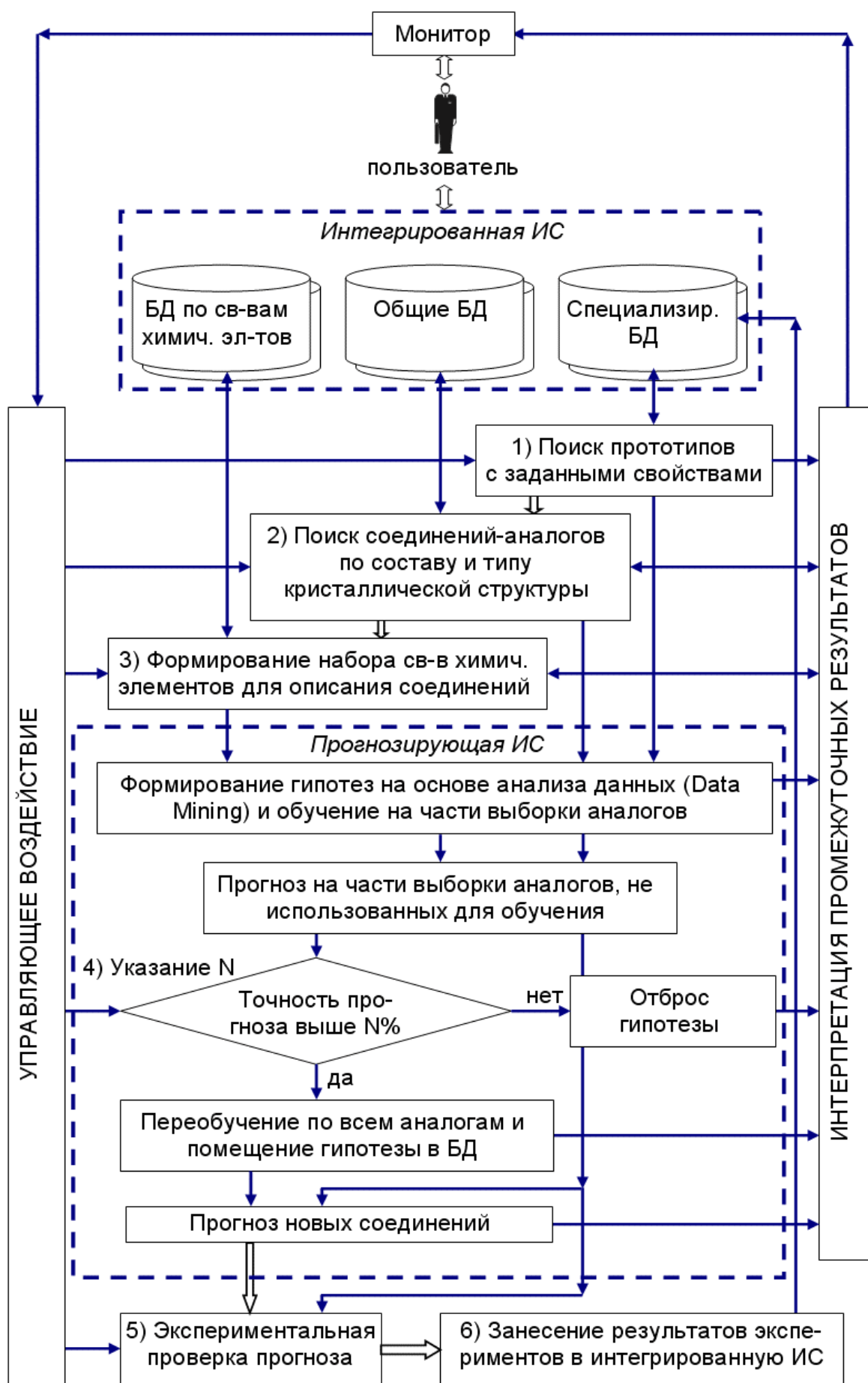


Рис. 9. Методика использования интегрированной ИС в СПДР.

Далее используются программы распознавания образов (были использованы системы распознавания образов “Распознавание”, разработанная в ВЦ РАН, и “Confor”, разработанная в Институте кибернетики Национальной АН Украины), анализирующие интервалы изменения значений свойств, которые соответствуют различным классам систем. В качестве обучающей используется часть выборки соединений-аналогов. Результатом этого этапа является гипотеза о взаимосвязях в наборе свойств химических элементов, разделяющая системы на разные классы. Затем ИС осуществляет распознавание классов веществ, согласно полученной гипотезе. Прогнозирование осуществляется на экзаменационной выборке, являющейся частью выборки соединений-аналогов. Если точность прогноза превышает указываемый пользователем-экспертом порог (обычно 80%), то гипотеза считается верной, а иначе отбрасывается.

На заключительном этапе в гипотезу подставляются наборы значений свойств элементов – компонентов еще не исследованных систем, и пользователь-эксперт получает прогноз, будет ли образовываться в данной системе соединение заданного состава. Аналогично можно получить прогноз соединений с определенным типом кристаллической структуры или с параметрами, значения которых находятся в определенном интервале.

После экспериментальной проверки результатов прогнозирования информация помещается в разработанную интегрированную ИС, пополняя соответствующие БД. При этом, если эти данные не совпадают с результатом прогноза, то пользователь может инициировать переобучение системы с учетом новых экспериментальных сведений. За счет использования большей выборки, очень часто, обучающей системе удается найти больше взаимосвязей, построить лучшие гипотезы и тем самым повысить точность будущих прогнозов.

На основе предложенной методики проведены эксперименты по использованию интегрированной ИС для анализа больших массивов хранящейся информации с целью поиска взаимосвязей в данных и использования их для конструирования новых веществ с заданными свойствами. Получены следующие результаты:

- Осуществлено конструирование еще не полученных перспективных полупроводниковых соединений состава ABX_2 ($X = S, Se, Te$). Сравнение результатов расчетов с новыми экспериментальными данными показало, что из 61 проверенного прогноза только 3 оказались неправильными, т.е. ошибка прогнозирования составила менее 5 %.

- Для соединений состава ABX_2 ($X = S, Se, Te$) с кристаллической структурой халькопирита решена задача прогноза ширины запрещенной зоны, при этом ошибка прогнозирования оказалась порядка 20 %, а при применении метода “голосования” результатов прогнозов с использованием различных алгоритмов удалось добиться правильных прогнозов.
- Осуществлен прогноз более сложных по составу соединений $A_2B_2(XO_4)_3$ с кристаллической структурой лангбейнита, перспективных для поиска новых пьезоэлектрических, сегнетоэлектрических, нелинейнооптических, электрооптических и люминесцентных веществ. Из 17 экспериментально проверенных прогнозов 12 оказались верными, т.е. ошибка прогнозирования составила менее 30 %.

На основе анализа полученных результатов сделан вывод о перспективности использования разработанной интегрированной ИС как информационной основы для программного комплекса компьютерного конструирования неорганических соединений.

ЗАКЛЮЧЕНИЕ

В диссертационной работе получены следующие результаты:

- на основе анализа современных технологий интеграции гетерогенных информационных систем, выработаны критерии и осуществлен выбор программной платформы для реализации информационной системы;
- на основе проведенного анализа архитектуры информационных систем по свойствам веществ и технологиям их получения разработан комплексный подход к интеграции информационных систем, как на уровне пользовательских интерфейсов, так и на уровне информационных источников;
- на основе теории множеств формализовано понятие релевантной информации, содержащейся в интегрируемых информационных системах рассматриваемой предметной области;
- разработана методика интеграции распределенных Web-приложений информационных систем с учетом требований информационной безопасности, оснащенная поисковым механизмом для обнаружения релевантной информации, которая содержится в интегрируемых информационных системах;
- разработана методика интеграции разнородных информационных источников на основе предложенных механизмов разрешения конфликтов гетерогенности и схемы предметной области;
- разработан, реализован и внедрен в ИМЕТ РАН программный комплекс, интегрирующий информационные системы по свойствам веществ и технологи-

ям их получения, который позволяет пользователям осуществлять доступ ко всей информации, хранящейся в интегрированных информационных системах;

- использование интегрированной ИС для анализа больших массивов хранящейся информации с целью поиска взаимосвязей в данных и использования их для прогнозирования свойств веществ показало, что средняя ошибка прогнозирования – менее 20 %.

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи в журналах, рекомендованных ВАК для опубликования результатов диссертационных работ:

1. Н.Н. Киселева, И.В. Прокошев, В.А. Дударев, В.В. Хорбенко, И.Н. Белокурова, В.В. Подбельский, В.С. Земсков. Система баз данных по материалам для электроники в сети Интернет // Неорганические материалы, 2004, т.40, №3, стр. 380-384.
2. В.А. Дударев, Н.Н. Киселева, В.С. Земсков. Интегрированная система баз данных по свойствам материалов для электроники // Перспективные материалы, 2006, №5 – стр. 20-25.
3. Н.Н. Киселева, В.А. Дударев, А.В. Столяренко, В.С. Земсков. Компьютерное конструирование неорганических соединений, перспективных для поиска новых материалов для электроники // Изв.ВУЗов. Материалы электронной техники, 2006, №3 – стр. 61-68.

Статьи и тезисы докладов:

4. И.Н. Белокурова, В.А. Дударев, В.С. Земсков и др. Базы данных по материалам для электроники, доступные пользователям Интернета // Информационное общество, 2001, №5, с.24-27.
5. В.А. Дударев. Программа удаленного администрирования базы данных по физико-химическим свойствам веществ. XXVIII Гагаринские чтения. Тезисы докладов Международной молодежной научной конференции. М.: МАТИ, 2002 – стр. 18-19.
6. В.А. Дударев. Программа удаленного администрирования базы данных по свойствам кристаллов акустооптических, электрооптических и нелинейно-оптических веществ. “Новые информационные технологии”. Тезисы докладов. Том 2. М.: МГИЭМ, 2002 – стр. 359-360.
7. В.А. Дударев. Подходы к интеграции российских баз данных по материалам для электроники. “Новые информационные технологии”. Тезисы докладов. М.: МГИЭМ, 2004 – стр. 83-85.
8. N.N. Kiselyova, V.A. Dudarev, et al. The Distributed System of Databases on Properties of Inorganic Substances and Materials. “Information Research, Appli-

- cations and Education – i.Tech” – Proceedings of Third International Conference. Sofia: FOI-Commerce, 2005. pp. 22-27.
9. V.F. Kornyshko, V.A. Dudarev. Software Development for Distributed System of Russian Databases on Electronics Materials. “Information Research, Applications and Education – i.Tech” – Proceedings of Third International Conference. Sofia: FOI-Commerce, 2005. pp. 27-33.
 10. В.А. Дударев. Разработка программного обеспечения интегрированной системы баз данных ИМЕТ РАН по свойствам материалов для электроники. “Научный сервис в сети Интернет: технологии распределенных вычислений”. Труды Всероссийской научной конференции. М.: МГУ, 2005 – стр. 227-229.
 11. В.А. Дударев. Создание интегрированной системы баз данных по материаловедению. Труды молодых специалистов ИМЕТ им. А.А. Байкова РАН 2005: Сб. статей под ред. Ю.К. Ковнеристого и др. – Москва, Издательство Интерконтакт Наука, 2005 – стр. 148-150.
 12. В.А. Дударев, К.Ю. Колыбанов. Повышение экономической эффективности разработки полифункциональных материалов на основе интеграции баз данных. Ученые записки МИТХТ, выпуск 14: Сб. статей под ред. В.С. Тимофеева и др. – Москва, Изд.-полиграф. центр МИТХТ им. М.В. Ломоносова, 2005 – стр. 79-80.
 13. V. Kornyshko, V. Dudarev. Software Development for Distributed System of Russian Databases on Electronics Materials // Int. Journal “Information Theories & Applications”, vol. 13, number 2, 2006. pp. 121-126.
 14. N. Kiselyova, V. Dudarev et al. The Distributed System of Databases on Properties of Inorganic Substances and Materials. // Int. Journal “Information Theories & Applications”, vol. 12, number 3, 2005. pp. 219-224.

Подписано в печать 09.11.2006. Сдано в производство 13.11.2006.

Формат бумаги 60x90 1/16. Объем 1,25 п.л.

Тираж 100 экз. Заказ № 496

Отпечатано в ООО "Фирма БЛОК"

107140, г.Москва, ул. Краснопрудная, вл.13. т.264-3073

Изготовление брошюр, авторефератов, печать и переплет диссертаций