

Интегрированная информационная система по свойствам неорганических веществ и материалов

© В. А. Дударев, © Н. Н. Киселева

ФГБУН Институт металлургии и материаловедения им. А.А. Байкова РАН (ИМЕТ РАН),
Москва

vic@imet.ac.ru, kis@imet.ac.ru

Аннотация

Рассматривается созданная в ИМЕТ РАН интегрированная информационная система по свойствам неорганических веществ и материалов. Кратко описываются предпосылки для создания системы, приводится информация о разработке интегрированной системы в области неорганического материаловедения, направленной на консолидацию созданных информационных систем и их источников данных. В заключении обсуждаются перспективы развития системы.

Работа выполнена при частичной финансовой поддержке РФФИ, проекты 14-07-31032 и 15-07-00980.

1 Введение

Развитие многих отраслей промышленности обусловлено использованием новых материалов и связанных с ними технологий. Поэтому важной является проблема обмена информацией между разработчиками и потребителями материалов. Традиционная система публикации результатов научных разработок (статьи, монографии, справочники) в области материаловедения не обеспечивает высокие темпы развития таких высокотехнологичных отраслей, как, например, электроника.

Разработка и активное использование информационных систем (ИС) по свойствам неорганических веществ и материалов (СНВМ) – как основа эффективного информационного обмена между разработчиками и потребителями информации – ведется во всех развитых странах [23]. При этом наибольшего прогресса в этой разработке добились США и Япония, которые на базе NIST (National Institute of Standards and Technology – Национальный институт стандартов и технологий, США) [12] и NIMS (National Institute for Materials Science Technology – Национальный

институт материаловедения, Япония) [11] создали и активно развивают обширные комплексы материаловедческих информационных систем, основанных на базах данных. Россия также обладает неплохим потенциалом и все еще находится в лидирующей группе, однако, в последние годы можно отметить определенный застой в этой области, связанный с недостаточным финансированием. В нашей стране существует ряд специализированных баз данных (некоторые даже обладают всемирной известностью, например, ИВТАНТЕРМО, созданная в ОИВТ РАН [2]) разработанных различными организациями, и никак не связанных друг с другом.

Большинство разработанных систем являются фактографическими, хотя объем информации в области неорганических веществ и материалов, содержащейся в документальных ИС (Science-direct [13], Wiley [16], Springer [14], American Chemical Society [1], Chemical Abstract Service [3], STN [15], ВИНТИ [17], e-library [6] и т.д.), значительно превосходит суммарный объем данных фактографических систем [20].

Несмотря на увеличивающиеся объемы данных, содержащиеся в рамках материаловедческих БД, ни одна из них не содержит полного описания всех свойств веществ. Поэтому всестороннее изучение параметров того или иного материала требует сбора и анализа информации из целого ряда информационных систем. Такой анализ является необходимым, поскольку в современных многофункциональных устройствах только исчерпывающая характеристика материала позволяет материаловеда принять решение о целесообразности его использования. Одним из способов обеспечения доступа ко всему массиву материаловедческой информации является интеграция ИС СНВМ с целью предоставления полной совокупности данных о неорганических веществах и материалах.

2 Каталог ИС СНВМ IRIC

При создании интегрированных систем в любой предметной области самой первой встает задача систематизации самих исходных ИС СНВМ. Именно задача систематизации наиболее значимых БД по свойствам неорганических веществ решалась

Труды XVII Международной конференции DAMDID/RCDL'2015 «Аналитика и управление данными в областях с интенсивным использованием данных», Обнинск, 13-16 октября 2015

на базе ИМЕТ РАН при разработке БД IRIC по информационным ресурсам в области неорганической химии (IRIC – Information Resources on Inorganic Chemistry) [21]. Отметим, что в настоящее время в мире, как это ни странно, не существует подобного каталога информационных ресурсов, посвященных неорганическому материаловедению.

2.1 Логическая схема базы данных IRIC

Как известно, любая ИС состоит наполовину из данных, а наполовину – из программного кода. Схема данных является наиболее критичной частью для реализации любой ИС, поскольку основные функции ИС разрабатываются именно на уровне схемы данных. Таким образом, если схема данных поддерживает некоторую функциональность, то программный код способен реализовать ее. Если нет, то, как бы хороша не была программная реализация, конечная ИС не сможет качественно поддерживать функции, изначально не заложенные в схему БД. Поэтому важно было выделить основные сущности для справочной системы IRIC и отношения между ними, которые позже легли в основу проектируемой БД.

Перечислим основные из сущностей, представленных на схеме данных: страны (CountriesInfo), организации-разработчики (OrganisationsInfo), БД (Databases), ключевые слова (KeywordsInfo), литературные публикации (LitReferences) и их авторы (AuthorsInfo), условия доступа к ИС (LicenseType). Перечисленные сущности, а также связи между ними представлены на логической схеме БД (рис. 1).

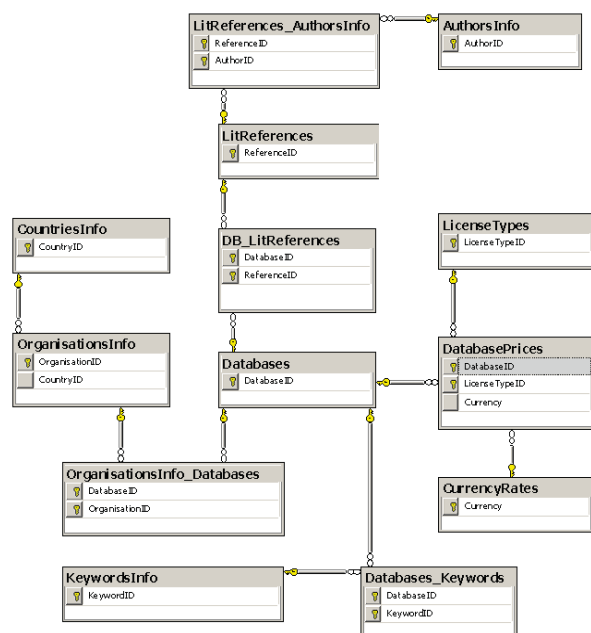


Рис. 1. Сущности БД IRIC и связи между ними.

Разработанная структура данных позволяет обработать все возможные запросы пользователей, упрощает взаимодействие с БД и гарантирует непротиворечивость хранящихся данных.

Ссылочная целостность обеспечивается использованием первичных и внешних ключей в СУБД Microsoft SQL Server 2008.

2.2 Разработка ИС IRIC

Следует отметить, что при разработке схемы БД учитывалась возможность представления всей информации на двух языках: русском и английском. Это впоследствии открыло возможность для написания русскоязычного и англоязычного интерфейса к БД IRIC, что позволило широкому кругу материаловедов не только в нашей стране, но и за рубежом использовать данную ИС.

Интернет является средой, обеспечивающей быстрый доступ к информации из любой точки мира. Именно поэтому для доступа к информации ИС IRIC было разработано Web-приложение, позволяющее конечным пользователям получить мгновенный доступ к информации через любую программу-браузер. Web-приложение реализовано по классической трехзвенной архитектуре (на основе концепции "клиент-сервер"), включающей в себя три звена: браузер (клиент); Web-сервер (сервер приложений); сервер БД.

Информация предоставляется в открытом доступе на двух языках (русскоязычный интерфейс доступен по адресу <http://iric.imet-db.ru>, полный англоязычный аналог – по адресу <http://en.iric.imet-db.ru>), что открывает возможность использования БД широким кругом специалистов не только в нашей стране, но и за рубежом [21, 22].

ИС IRIC поддерживает ряд поисковых запросов, которые отличаются количеством и типом критериев поиска: общий запрос – просмотр всей информации по выбранному разделу меню; простой запрос – поиск данных по одному критерию; сложный запрос – поиск данных по нескольким критериям.

Список БД по ключевым словам

Фильтр по ключевым словам: энтальпия, энтропия

| № | Название (аббревиатура) | Контакты | Комментарий |
|---|---|--|---|
| 1 | БД термодинамических свойств индивидуальных веществ (ИВТАНТЕРМО) [подробнее...] | Сайт: http://www.chem.msu.ru/rus/handbook/iwtan Тел: +74954851000 Факс: +74954851000 e-Mail: lorish@ihed.ras.ru | Термодинамические (рекомендованные) данные для неорганических веществ. |
| 2 | БД по термодинамическим и транспортным свойствам чистых газов и жидкостей (NISTFLUIDS) [подробнее...] | Сайт: http://www.nist.gov/srd/nist23.cfm Тел: +13019752208 Факс: +13019260416 e-Mail: Joan.Sauerwein@nist.gov | Термодинамические и транспортные свойства чистых неорганических и органических жидкостей и газов. |
| 3 | БД по термодинамическим свойствам органических и неорганических веществ с одним или двумя атомами углерода (JANAF) [подробнее...] | Сайт: http://kinetics.nist.gov/janaf/ Тел: Факс: e-Mail: | Термодинамические и химические свойства неорганических веществ и органических веществ с одним-двумя атомами углерода. |

Всего найдено: 3. 1-3

Рис. 2. Пример результатов запроса к ИС IRIC по ключевым словам.

Примером сложного запроса является поиск материаловедческих БД, разработанных в определенной стране и организации. На рис. 2 представлены результаты поиска для двух ключевых слов “энтальпия” и “энтропия”. На

сегодняшний день в ИС IRIC содержится информация о трех БД, удовлетворяющих критерию поиска: отечественной разработке ИВТАНТЕРМО (ОИВТ РАН) и двух системах, принадлежащих NIST (рис. 2)..

2.3 Анализ тенденций в разработке ИС СНВМ

Основной задачей создания и эксплуатации ИС на основе БД по свойствам неорганических веществ является информационное обслуживание специалистов-химиков. Однако БД предоставляют значительно более широкие возможности манипулирования данными. Одной из таких возможностей является статистический анализ и выявление основных трендов развития информационных ресурсов в области неорганической химии и материаловедения [22].

Анализ тематики ключевых слов позволяет выделить наиболее распространенные параметры неорганических веществ и материалов, информация о которых содержится в БД СНВМ. Традиционное первое место по числу БД занимают фактографические информационные системы, аккумулирующие информацию по термодинамическим и теплофизическим свойствам. Интенсивно разрабатываются и широко используются в фундаментальных и прикладных исследованиях и промышленности БД с кристаллографической и кристаллохимической информацией. В последние годы увеличилось число БД, содержащих сведения о механических свойствах (прочности, усталости, ползучести и т.д.) и электрических свойствах неорганических веществ и материалов.

Число компьютерных информационных ресурсов и их объем могут служить показателем научного и промышленного потенциала страны. Анализ информации БД IRIC позволил выявить страны-лидеры в области разработки материаловедческих информационных ресурсов. Мировой лидер – США – занимает первое место в списке разработчиков БД СНВМ, главным образом за счет развития информационных систем NIST [12]. В последнее десятилетие Япония неуклонно наращивает объем информационных ресурсов во многом благодаря NIMS – основному провайдеру в области материаловедческих БД [11]. В тройку лидеров входит пока и Россия, главным образом за счет БД СНВМ, разрабатываемых в академических институтах (ИМЕТ РАН, ОИВТ РАН и др.).

Анализ информации, содержащейся в ИС IRIC, в том числе и предоставляемых ею полных текстов статей, позволил выявить и основные тенденции в области разработки и эксплуатации современных ИС СНВМ:

- доступ к информации из сети Интернет, который позволяет "доставить" необходимую и самую "свежую" информацию непосредственно на рабочее место химика или материаловеда;

- экспертная оценка хранящейся информации, для которой привлекаются высококвалифицированные специалисты, что дает в руки пользователя не просто "сырую" информацию, а рекомендуемые значения;
- оснащение БД СНВМ средствами анализа информации – от традиционных термодинамических расчетов и статистических процедур до современных средств поиска закономерностей в данных, позволяющих прогнозировать поведение объектов и обеспечивающих принятие решений [20, 26];
- интеграция БД по веществам и материалам в целях предоставления пользователю наиболее полной информации о свойствах конкретного вещества, а также для последующего анализа совокупной информации о веществах и материалах.

Разработанная ИС IRIC позволяет материаловедам не только получать информацию о существующих в мире ИС СНВМ на русском и английском языках, но и проводить поиск таких ИС по многокритериальным запросам. В последние годы для обеспечения специалистов полной информацией наблюдается тенденция к интеграции уже созданных БД, как на национальном, так и на международном уровнях, в том числе в рамках CODATA и ЮНЕСКО. При этом наиболее многообещающим подходом является виртуальная интеграция ИС СНВМ, разработанных в рамках разных организаций [8] в противоположность использованию технологии хранилищ данных, более уместных для получения консолидированных данных в одной организации [7, 25].

Создание ИС IRIC позволяет систематизировать имеющуюся информацию в материаловедческих БД на самом верхнем уровне и указать варианты наиболее разумной интеграции созданных информационных систем с целью минимизации времени, затрачиваемого специалистами на поиск требуемой информации. А это, в свою очередь, является важным шагом на пути к созданию единой интегрированной материаловедческой ИС [24].

3 Иерархия понятий предметной области

Ключевой проблемой при интеграции является стандартизация понятий предметной области. Описание химических сущностей и их свойств в разных БД СНВМ происходит с разной степенью детализации. Значения свойств, хранимые в разных информационных источниках, определяются, в первую очередь, составом неорганических веществ (набором образующих их химических элементов и соотношением их друг с другом). В свою очередь, физические свойства веществ во многом зависят от кристаллической структуры.

Таким образом, может быть построена иерархия понятий, которая используется при описании свойств химических сущностей. Обозначив сущности второго уровня общим термином "вещество" получаем трехуровневую иерархию

химических объектов: система, вещество и кристаллическая модификация (рис. 3) [18].

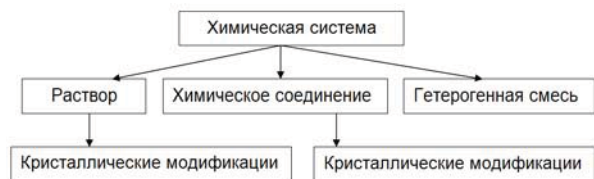


Рис. 3. Иерархия понятий, используемых в ИС СНВМ.

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 [8]. Множество химических систем, образуемых элементами, обозначается S , множество химических веществ – C , а множество кристаллических модификаций – M . Химическая система обозначается s (где $s \in S$), химическое вещество – c (где $c \in C$), а химическую модификацию – m (где $m \in M$).

Химическая система s представляется множеством элементов e_i : $s = \{e_1, e_2, \dots, e_n\}$, т.е. определяется качественный состав вещества. Неорганическое вещество c определяется не только множеством атомов, но и количественным вхождением последних в состав вещества, раствора или смеси. Поэтому вещество c представлено кортежем (s, f) , где $s \in S$, а f является отображением множества элементов, которые образуют вещество, на множество пар $R^+ \times R^+$, задающих соответственно минимальное и максимальное вхождение заданного химического элемента в вещество, раствор или смесь c . То есть $f: e_i \rightarrow (R_{\min}^+, R_{\max}^+)$, где R^+ – множество неотрицательных действительных чисел. R_{\min}^+ и R_{\max}^+ , соответственно, минимальная и максимальная концентрация химического элемента e_i в веществе c . В случае, когда концентрация конкретного химического элемента e_i в веществе c фиксирована, то $R_{\min}^+ = R_{\max}^+$. Кристаллическая модификация m представляется кортежем (s, f, mod) , где $s \in S$, $f: e_i \rightarrow (R_{\min}^+, R_{\max}^+)$, а mod – строковое обозначение кристаллической модификации вещества, принятое в интегрированной ИС СНВМ. Для разрешения конфликтов наименований модификаций используются тезаурусы.

Далее множества C и M , расширяются пустым элементом $null$, т.е. $null \in C, null \in M$. Любой

химический объект (система, вещество и модификация) может быть описан тройкой (s, c, m) , где $s \in S, c \in C, m \in M$. Таким образом, получаем шаблоны для записи химических объектов следующего вида: $(s, null, null)$ – для систем, где определен качественный состав вещества; $(s, c, null)$ – для веществ, где определен количественный состав; (s, c, m) – для кристаллических модификаций. Предложенная иерархия химических понятий используется в контексте построения интегрированной ИС СНВМ.

4 Методы интеграции ИС

В настоящее время выделяют три метода интеграции. Это интеграция корпоративных приложений (Enterprise Application Integration, EAI) [9], интеграция корпоративной информации (Enterprise Information Integration, EII) [10] и программное обеспечение для извлечения, преобразования и загрузки данных (Extract, Transform, Load – ETL) [7].

Принципы интеграции, заложенные в этих методах, используются для решения широкого круга задач: от интеграции в режиме реального времени до пакетной интеграции, и от интеграции данных до интеграции приложений. На рис. 4 показано положение названных методов по отношению к этим двум типам задач.

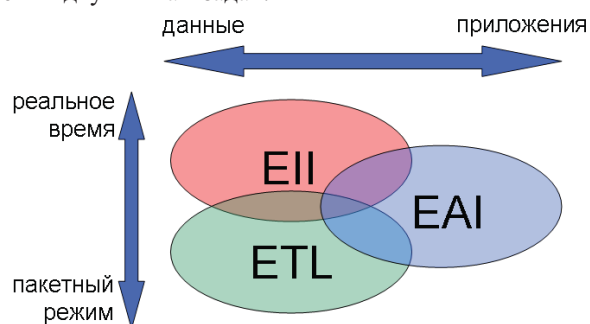


Рис. 4. Современные методы интеграции ИС.

Для интеграции данных в режиме реального времени лучше всего подходит подход EII [10]. Для пакетной интеграции данных – ETL [7]. А для интеграции приложений, в режиме реального времени или пакетном, наиболее подходящим инструментом является применение метода EAI [9]. Следует отметить, что ни один из существующих на сегодняшний день методов интеграции не способен решить все проблемы, возникающие при объединении ИС [24].

Для консолидации на уровне данных применимы два метода интеграции: ETL и EII. Первый подразумевает слияние всех данных в единую информационную систему – Data Warehouse. Именно такой подход является основой для методологии интеграции ETL (Extract, Transform, Load), при которой данные из разных БД после унификации и очистки от неточностей загружаются в общее хранилище данных – Data Warehouse [7].

Однако, при всей заманчивости такой мегабазы данных, процедура унификации информации баз данных, созданных в разных организациях и/или разных странах, с использованием различных языков, аппаратных и программных средств, отличающихся по точности и методам измерения данных, является крайне сложной технической и организационной задачей, а ее создание и эксплуатация требует огромных финансовых вложений.

Другой путь – это виртуальная интеграция БД СНВМ и создание неоднородной распределенной информационной системы. Именно этот путь позволяет обеспечить независимость развития отдельных баз данных и организовать доступ ко всему массиву данных о конкретном веществе или материале для конкретного пользователя, что является основной целью интеграции [5].

При виртуальной интеграции ИС СНВМ возможно использовать два основных метода:

- интеграция корпоративной информации (Enterprise Information Integration, EII) [10];
- интеграция корпоративных приложений (Enterprise Application Integration, EAI) [9].

В первом случае разрабатывается программный интерфейс доступа (API – Application Programming Interface) к информационным источникам, с помощью которого можно извлекать необходимые данные из разных БД. То есть, строится некая центральная информационная система, которая взаимодействует с распределенными источниками данных, извлекает и предоставляет пользователю агрегированную информацию о запрашиваемом веществе из разных БД СНВМ.

Использование второго подхода (EAI) наиболее целесообразно, когда ИС СНВМ включают прикладные программы, которые невозможно использовать вне оригинального контекста. При реализации этого подхода объединяются не сами БД ИС, а только их пользовательские интерфейсы, которые осуществляют доступ к расчетным подсистемам. Такими интерфейсами, как правило, являются Web-приложения соответствующих ИС.

В связи с тем, что интегрируемые ИС СНВМ могут пересекаться по набору свойств веществ, а качество информации (достоверность и полнота) в каждой ИС отличается для разных свойств, необходима подсистема, поддерживающая экспертные оценки информации интегрируемых БД. Экспертиза должна проводиться высококвалифицированными специалистами, которые выставляют оценки, характеризующие качество данных в разнородных интегрируемых ИС. Таким образом, при наличии информации по какому-либо свойству вещества в нескольких интегрируемых БД, интегрированная ИС должна выдавать не только сами данные, но и степень их достоверности, рассчитанную на основе экспертных оценок. Разработка такой подсистемы экспертизы информации интегрируемых БД – одна из самых

сложных организационных задач интеграции ИС СНВМ.

Необходимо также учесть, что квалифицированные пользователи не всегда доверяют экспертным оценкам коллег. В связи с этим ИС СНВМ должны содержать гиперссылки на полные тексты исходных публикаций, из которых извлечена информация. Таким образом, квалифицированные пользователи смогут получить доступ не только к значениям конкретных свойств, но также и к информации о том, каким образом, когда и кем были получены соответствующие результаты.

5 Интегрированная ИС СНВМ ИМЕТ РАН

Опыт интеграции информационных ресурсов в области неорганического материаловедения, показал, что ни один из существующих подходов не способен решить все проблемы объединения информационных источников и программных приложений ИС СНВМ. Поэтому была предложена методология интеграции, сочетающая в себе интеграцию на уровне данных (ETL+EII) и пользовательских интерфейсов (EAI) [4]. В рамках этой методологии предоставляется как доступ к текущим пользовательским интерфейсам ИС и свободное перемещение пользователей между ними (EAI), так и богатые возможности по сбору и агрегации информации, полученной из разнородных распределенных источников данных по свойствам веществ, согласно общей разработанной информационной схеме (ETL для организации, EII для консолидации информационных ресурсов разных организаций). Использование разработанной методологии для интеграции российских ИС СНВМ, разработанных в ИМЕТ РАН, показало, что она предоставляет мощные инструментальные средства для консолидации гетерогенных информационных ресурсов, созданных с использованием разных компьютерных платформ, операционных систем, СУБД, отличающихся по языку, достоверности информации.

В настоящее время интегрированная ИС СНВМ объединяет все разработанные в ИМЕТ РАН информационные системы (БД Фазы, Elements, Диаграмма, Кристалл и BandGap), а также ИС AtomWork, разработанную в NIMS (Япония) [4, 5]. Интегрированная система также включает в себя подсистему для компьютерного конструирования неорганических соединений и прогнозирования их свойств (ИАС – Информационно-Аналитическая Система) [20]. Особенностью разработанной интегрированной системы является то, что входящие в ее состав БД СНВМ созданы с использованием различных СУБД и функционируют на принципиально различных компьютерных платформах: Sun UltraSPARC (БД «Диаграмма») и Intel (прочие БД) под управлением разных операционных систем: Sun Solaris (БД Диаграмма),

Unix (AtomWork) и Microsoft Windows Server (прочие БД).

При интеграции ИС СНВМ необходимо предусмотреть возможность просмотра информации, содержащейся в других ИС, о выбранном пользователем веществе. Таким образом, нужен некоторый координирующий центр, который знает о том, в каких ИС и какая информация хранится. То есть, должен существовать некоторый центр, который описывает информацию, содержащуюся в интегрируемых ИС. Такую функцию выполняет предложенная нами метабаза – специальная база данных, содержащая справочные сведения о содержимом интегрируемых ИС СНВМ [18].

5.1 Формализация метабазы и поиска релевантной информации

Ограничившись для простоты изложения только самым верхним уровнем иерархии (уровень химических систем), получим следующую формализацию для описания метабазы. В метабазе 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 . Можно предложить 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 $(s_2, s_1) \in R$.

Заметим, что в первом правиле можно заменить отношение $s_1 \subseteq s_2$ более строгим $s_1 = s_2$, тогда получим в качестве релевантных только те химические объекты, которые состоят из единого набора химических элементов.

Отметим, что ни одно из определений не является подходящим для решения всех задач по определению релевантной информации в распределенных ИС, и на практике часто используются несколько разных отношений релевантности R , называемые классами релевантности. Более того, возможно более четкое определение релевантной информации при использовании отношений R вида: $R \subset (d_1, s_1, p_1) \times (d_2, s_2, p_2)$, где $d_1, d_2 \in D; s_1, s_2 \in S; p_1, p_2 \in P$.

Улучшение релевантности поиска можно добиться также за счет использования обозначений веществ c_i или кристаллических модификаций m_i вместо обозначений химических систем s_i в случаях, когда пользователь запрашивает релевантную информацию, находясь на уровне неорганических веществ или их модификаций в предложенной иерархии понятий.

При поиске на уровне веществ учитывается количественный состав соединения. Обозначим парой $(a_{i \min}, a_{i \max})$ количественное вхождение химического элемента $e_i \in S$ в состав, $a_{i \min}, a_{i \max} \in R^+$, $a_{i \min} \leq a_{i \max}$. Если $a_{i \min} = a_{i \max}$, то вещество имеет постоянный состав по элементу $e_i \in S$. Для каждого элемента химической системы $e_i \in S$ пользователь при поиске может задать пару $(r_{i \min}, r_{i \max})$, где $r_{i \min}, r_{i \max} \in R^+$, обозначающую допустимый интервал вхождения i -го элемента в состав вещества (R^+ – множество неотрицательных действительных чисел). Тогда релевантными будут все вещества, относящиеся к той же химической системе, у которых для каждой пары $(r_{i \min}, r_{i \max})$ выполняется $a_{i \min} \in [r_{i \min}, r_{i \max}]$ или $a_{i \max} \in [r_{i \min}, r_{i \max}]$. Другими словами, если логическая дизъюнкция $[r_{i \min} \leq a_{i \min} \ \& \ a_{i \min} \leq r_{i \max}] + [r_{i \min} \leq a_{i \max} \ \& \ a_{i \max} \leq r_{i \max}] = true$ для всех $e_i \in S$, то данные о веществе являются релевантными.

При поиске релевантной информации с учетом кристаллических модификаций m_i учитываются сингонии, т.к. часто информация о кристаллических структурах может указываться по-разному. Например, для ниобата лития (LiNbO_3) в разных информационных источниках ИС СНВМ указывается гексагональная или тригональная кристаллографическая система, что, по сути, соответствует одной кристаллической структуре.

5.2 Безопасный просмотр релевантной информации в интегрируемых ИС СНВМ

Для обеспечения информационной безопасности при переходах пользователей между Web-приложениями ИС предложено использовать систему шлюзов безопасности, санкционирующих переходы пользователей между ИС СНВМ и обеспечивающих отображение релевантной информации. Шлюз метабазы санкционирует переход между ИС, а шлюз ИС выполняет сопряжение централизованной системы безопасности с системой безопасности интегрируемой ИС СНВМ.

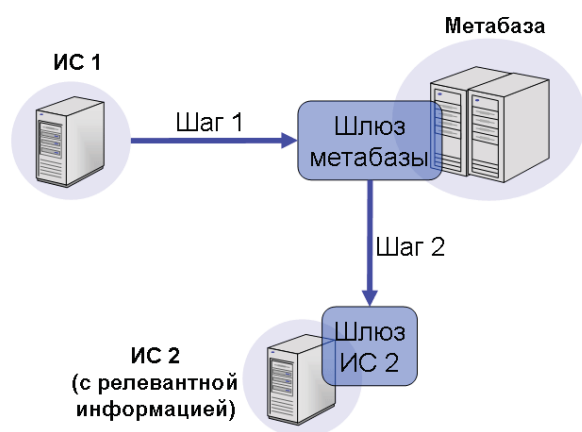


Рис. 5. Интеграция ИС с поиском релевантной информации в метабазе.

При работе с интегрированной системой зарегистрированный пользователь входит в одну из баз данных, размещенных по адресу в Internet: <http://www.imet-db.ru>. При работе, например, с ИС BandGap, пользователь, выбрав вещество из списка, получает на экране меню, содержащее ссылки на другие ИС с релевантной информацией. При «щелчке» по этой ссылке пользователь автоматически переходит в ИС Диаграмма на систему или в ИС Кристалл на вещество, содержащие тот же набор элементов, которые образуют выбранное соединение (рис. 5). Естественно, что такие переходы возможны, если ИС содержат информацию об аналогичных веществах.

5.3 Единая точка входа в ИС СНВМ

Очень часто при поиске данных по свойству того или иного вещества неискушенный пользователь не знает к какой ИС СНВМ стоит прибегнуть для

первичного сбора информации. Поэтому актуальным является создание специализированной ИС, позволяющей потребителю данных по свойствам неорганических веществ получить возможность просмотра связанной информации по свойствам заданной химической системы в разных ИС СНВМ из одного места, которое условно называется “единой точкой входа”.

Основная идея заключается в предоставлении пользователю возможности выбора химических элементов (из периодической таблицы Менделеева), образующих химическую систему. Имея набор выбранных пользователем элементов, ИС единой точки входа должна осуществить поиск ИС СНВМ, содержащих сведения о свойствах фаз выбранной химической системы, для чего используется метабаза, разработанная ранее при создании интегрированной ИС СНВМ.

Поскольку поиск релевантной информации выполняется в метабазе, единая точка входа предоставляется для всех ИС СНВМ, описанных в метабазе. В настоящее время интегрированная ИС СНВМ консолидирует все разработанные в ИМЕТ РАН ИС СНВМ и ИС AtomWork (разработанную в NIMS, Япония) [4].

Рассмотрим кратко особенности разработки Web-приложения единой точки входа, располагаемого по адресу <http://meta.imet-db.ru>. Web-приложение ASP.Net написано на языке C# (.Net Framework 3.5) с использованием ADO.Net для доступа к метабазе. Для построения запросов используются языковые средства Transact-SQL, являющегося диалектом языка SQL, который используется в СУБД Microsoft SQL Server 2008 [19].

Пользовательский интерфейс является интерактивным за счет использования библиотеки jQuery, облегчающей взаимодействие с HTML DOM и предоставляющей удобный API для работы с AJAX. Таким образом, после выбора химической системы из таблицы Менделеева в ИС единой точки входа пользователь в виде гиперссылок получает сведения о релевантной информации, содержащейся в распределенных БД интегрированной ИС ИМЕТ РАН. Затем, пройдя по гиперссылкам, пользователь получает возможность ознакомиться с информацией по свойствам выбранного вещества системы в разных информационных источниках. Другими словами, создана единая точка входа, используя которую пользователь может получить доступ ко всей информации, содержащейся в интегрированной ИС.

6 Заключение

В данной работе кратко рассмотрен подход к интеграции информационных систем на основе метабазы, успешно примененный в ИМЕТ РАН для обеспечения совместной работы ряда гетерогенных программных комплексов по свойствам неорганических веществ и материалов, созданных

как в нашей стране, так и за рубежом. В перспективе планируется подключение новых ИС СНВМ, прежде всего, справочника “Термические константы веществ”, разработанного в ОИВТ РАН и МГУ, и ИС CompES, разработанной NIMS [11] и содержащей результаты расчета электронной структуры неорганических веществ.

Интеграция ИС СНВМ является основным направлением развития информационных ресурсов в материаловедческих областях. Именно она позволит создать информационную структуру 21-ого века в области химии и материаловедения, позволяющую обеспечить специалистов достоверными и полными данными о свойствах веществ и материалов и доставить эту совокупную информацию в любую точку мира по сети Internet.

Литература

- [1] American Chemical Society, <http://www.acs.org>.
- [2] G.V. Belov, V.S. Iorish, V.S. Yungman. IVTANTHERMO for Windows — database on thermodynamic properties and related software. CALPHAD. 23(2) p. 173-180, 1999.
- [3] Chemical Abstract Service, <http://www.cas.org>.
- [4] V.A. Dudarev. Databases on properties of inorganic substances and materials integration infrastructure // Proceedings of The 3rd Asian Materials Database Symposium (AMDS 2012), 2012, p. 71-76.
- [5] V.A. Dudarev, N.N. Kiselyova, Y. Xu, M. Yamazaki. Virtual integration of the Russian and Japanese databases on properties of inorganic substances and materials // Proc. MITS-2009. Symposium on Materials Database (NIMS), 2009, pp. 37-48.
- [6] e-library, <http://www.elibrary.ru>.
- [7] R. Kimball, J. Caserta. The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data // John Wiley & Sons, 2004. 416 pages
- [8] V.F. Kornyshko, V.A. Dudarev. Software Development for Distributed System of Russian Databases on Electronics Materials // Int. Journal "Information Theories & Applications", v. 13, № 2, 2006, pp. 121-126.
- [9] J.P. Morgenthal. B.L. Forge. Enterprise Applications Integration with XML and Java. Prentice Hall Ptr, 2001. 504 pages.
- [10] J.P. Morgenthal. Enterprise Information Integration: A Pragmatic Approach. LULU Press, Morrisville, 2005. 324 pages.
- [11] NIMS Materials Database, http://mits.nims.go.jp/db_top_eng.htm.
- [12] NIST Materials Database, <http://www.nist.gov/chemistry-portal.cfm>.
- [13] Science-direct, <http://www.sciencedirect.com>.
- [14] Springer, <http://www.springer.com>.
- [15] STN, <http://www.stn-international.de>.
- [16] Wiley, <http://www.wiley.com>.
- [17] ВИНТИ, <http://www2.viniti.ru>.
- [18] В.А. Дударев, О.А. Филоретова. Подход к интеграции баз данных по свойствам неорганических веществ на основе метабазы // Прикладная информатика, №4(46), 2013, с. 38–42.
- [19] В.А. Дударев, Е.Г. Шмакова. Web-интерфейс для доступа к гетерогенным информационным системам по свойствам неорганических веществ // Интеграл, №4(71), 2013, с. 55.
- [20] Н.Н. Киселева. Компьютерное конструирование неорганических соединений. Использование баз данных и методов искусственного интеллекта. Наука, Москва, 2005. 288 с.
- [21] Н.Н. Киселева, В.А. Дударев. База данных "Информационные ресурсы неорганической химии и материаловедения" // Информационные технологии, №12, 2010, с. 63-66.
- [22] Н.Н. Киселева, В.А. Дударев. Информационная система по ресурсам неорганической химии и материаловедения // Вестник Казанского технологического университета, Т. 17, №19, 2014, с. 356-358.
- [23] Н.Н. Киселева, В.С. Земсков, В.А. Дударев. Компьютерные информационные ресурсы неорганической химии и материаловедения // Успехи химии, № 2, 2010, с. 163-188.
- [24] В.В. Масютин, В.А. Дударев. На пути к единой информационной системе по свойствам неорганических веществ // Интеграл, №6, 2010, с. 30 – 31.
- [25] А.Е. Поляков, В.А. Дударев. Хранилище данных для интеграции информационных систем по свойствам неорганических веществ // Интеграл, №6, 2011, с. 18 – 19.
- [26] А.Е. Поляков, В.В. Масютин, В.А. Дударев. Компьютерное конструирование неорганических соединений на основе интегрированной информационной системы // Прикладная информатика, № 4(40), 2012, с. 38-43.

Integrated Information System on Inorganic Substances and Materials Properties

Victor A. Dudarev, Nadezhda N. Kiselyova

Integrated information system on inorganic substances and material properties created at IMET RAS is considered. Reasons for the system creation are described shortly and some information on integrated system development in the field of inorganic materials science are given. In conclusion, the integrated system development perspectives are discussed.