

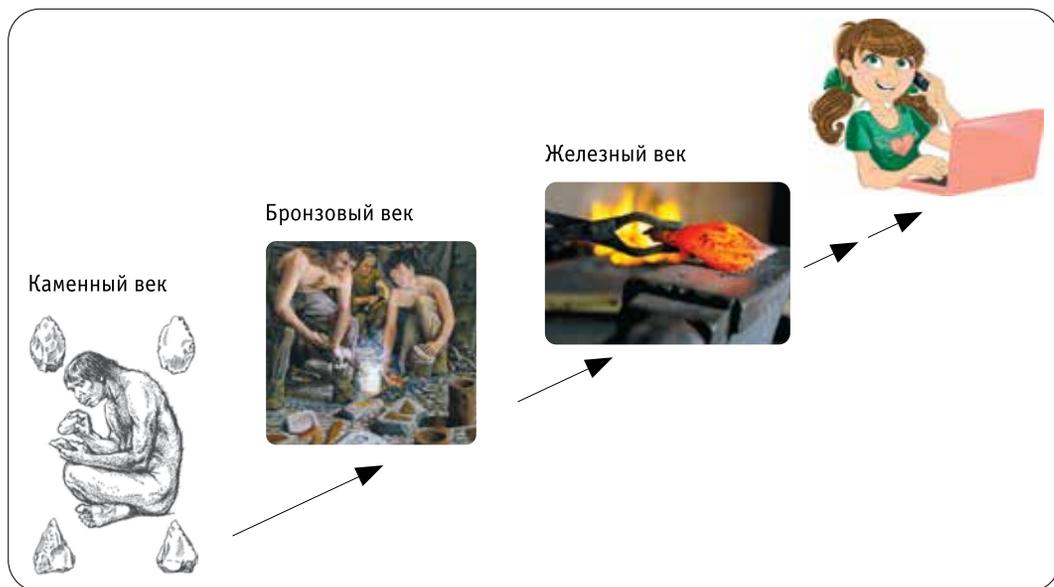
ИНФОРМАЦИОННАЯ ИНФРАСТРУКТУРА СОВРЕМЕННОГО МАТЕРИАЛОВЕДЕНИЯ – ПРОЕКТЫ И РЕЗУЛЬТАТЫ

Доктор химических наук КИСЕЛЕВА Н.Н.
(ИМЕТ им. А.А. Байкова РАН, Москва)

Конкурентоспособность продукции в значительной степени зависит от используемых в её производстве материалов. Во многих случаях открытие нового вещества с интересными для практических применений функциональными свойствами приводит к революционному перевороту в жизни людей. Например, применение полупроводни-

ков знаменовало новую эру нашей цивилизации. Даже обозначения многих этапов развития человечества производны от названий материалов: каменный век, бронзовый век, железный век и т.д. (рис. 1). Много лет назад мне попала статья американского авиаконструктора. Запомнилась одна фраза из этой статьи: инженер может выдумать самое перспективное устройство или машину, однако их создание упирается, в первую очередь, в наличие соответствующих материалов. Важность материалов для

Рис. 1.
Этапы развития
человеческой цивилизации.



развития промышленности и всей экономики стран в целом требует ускорения поиска, исследования и внедрения новых материалов с заданными функциональными свойствами. Однако в настоящее время, по мнению американских специалистов¹, между открытием нового материала и началом его практического использования проходит от 10 до 20 лет. Во многом это связано с тем, что очень часто потребители не имеют достаточной информации даже об очень перспективных материалах, работы по поиску, исследованию и созданию технологии получения и обработки материалов необоснованно дублируются, в практике используются не самые лучшие по потребительским и прочим параметрам вещества, что приводит к снижению качества продукции, росту затрат на её производство и, в конечном счете, к утрате рыночной привлекательности выпускаемого продукта.

Одним из путей ускорения поиска, разработки и внедрения новых материалов является создание развитой инфраструктуры информационного обеспечения специалистов, в первую очередь, распределённой виртуально интегрированной сети баз данных (БД), содержащих информацию о свойствах веществ и материалов и технологиях их получения и обработки, а также систем компьютерного конструирования и моделирования материалов. Вся эта инфраструктура должна быть доступна в Интернете специалистам самого разного профиля: научным работникам, инженерам, технологам, бизнесменам, госслужащим, студентам и т.д. Извечный спор между теоретиками и экспериментаторами на тему, всё ли можно рассчитать, должен родить истину: если можно что-то посчитать заранее или найти уже готовые данные в базах данных и тем самым уменьшить количество экспериментов,

¹ Сайт Materials Genome Initiative: <https://www.mgi.gov/>

то проводить расчёты или искать информацию в БД разумно. Я уже не говорю о компьютерных системах обработки информации в автоматизированных приборах и установках, которые экономят время и сокращают затраты на расшифровку самой разнообразной информации: от спектров до дифрактограмм. Для расчётов и информационного поиска нужны соответствующие инструменты и базы данных, т.е. информационная инфраструктура.

В последние годы в развитых странах были выдвинуты и поддержаны правительствами инициативы, направленные на организацию такой инфраструктуры доступа к экспериментальным и расчётным данным о материалах².



Стратегическая Инициатива Геном Материалов (Materials Genome Initiative (MGI))

В 2011 г. в США были начаты работы по проекту, названному “Инициативой Геном Материалов” (Materials Genome Initiative (MGI)). Такое странное название было связано с успехом международного

² Kiselyova N.N., Dudarev V.A. *Inorganic Chemistry and Materials Science Data Infrastructure for Specialists // Selected Papers of the XVIII International Conference on Data Analytics and Management in Data Intensive Domains (DAMDID/RCDL 2016). Ershovo, Moscow Region, Russia, October 11–14, 2016. CEUR Workshop Proceedings. 2016. V. 1752. P. 121–128. <http://ceur-ws.org/Vol-1752/paper21.pdf>*

проекта “Геном человека” (The Human Genome Project (HGP))³, начатого в 1990 г. под эгидой Национального института здоровья США (National Institutes of Health (NIH)), на выполнение которого было выделено несколько млрд долл. Уже первые успехи в расшифровке человеческого генома позволили осуществить прорывы в биотехнологии и медицине и показали, что немалые средства, вложенные в проект HGP, были потрачены не зря.

Однако вернёмся к геному материалов. Цели MGI – ускоренное создание новых материалов, обладающих заданными свойствами, что критично для достижения высокого уровня конкурентоспособности промышленности США и будет способствовать поддержке их лидирующей роли во многих секторах современного материаловедения и промышленности: от энергетики до электроники, от обороны до здравоохранения. Особое внимание в MGI уделяется поддержке прорывных исследований в теории, моделировании свойств материалов и data mining⁴ как средств достижения существенного прогресса в материаловедении, что приведёт к снижению затрат на разработку, исследование и получение новых материалов. Задачи MGI – обеспечение разработки и внедрения новых материалов, в том числе и за счёт координации исследований и предоставления доступа к расчётным моделям и инструментарию для оценки свойств и поведения материалов, а также использования перспективных методов моделирования и анализа данных. Особое внимание уделяется обучению студентов современным методам исследования и расчёта материалов, а также созданию БД с уже готовыми результатами расчётов. Реализация

проекта MGI позволит создать механизмы, способствующие обмену данными и знаниями о материалах не только между исследователями, но и между академической наукой и промышленностью. Основой MGI является “Инфраструктура инноваций в материаловедении” (Materials Innovation Infrastructure), которая обеспечивает интеграцию методов и средств современного моделирования и экспериментальных исследований. Инфраструктура включает также комплекс взаимосвязанных обслуживающих структур и объектов (в том числе и установок megascience), составляющих и/или обеспечивающих основу функционирования материаловедения как науки и прикладной области. На первом этапе на реализацию программы MGI выделено несколько сотен млн долл. Следует особенно отметить, что в Подкомитет MGI Национального научно-технического Совета США (The National Science and Technology Council (NSTC)) входят представители ведущих ведомств США: Министерства обороны, Министерства энергетики, National Institute of Standards and Technology (NIST), National Science Foundation (NSF), National Aeronautics and Space Administration (NASA), National Institutes of Health (NIH), United States Geological Survey (USGS), Defense Advanced Research Projects Agency (DARPA) и т.д.

Среди успешных результатов выполнения проектов этой инициативы можно выделить совместные исследования NIST с Национальными лабораториями и университетами США, направленные на разработку баз данных и расчётных систем, например, виртуальной лаборатории, предназначенной для организации доступа и обработки материаловедческой информации, в том числе, и с использованием методов искусственного интеллекта⁵, Центра конструирования иерархи-

³ Сайт The Human Genome Project: http://web.ornl.gov/sci/techresources/Human_Genome/index.shtml

⁴ Data mining – интеллектуальный анализ данных, процесс выявления скрытых закономерностей в экспериментальных данных.

⁵ Сайт NIST: <https://mgi.nist.gov/high-throughput-experimental-materials-science-virtual-laboratory>

ческих материалов (Center for Hierarchical Materials Design (CHiMaD))⁶, задачами которого является использование баз данных, расчётных и экспериментальных средств для конструирования, моделирования и получения новых материалов для электроники, медицины, авиационной и космической техники, а также проект NASA и NIST по созданию репозитория⁷ с информацией о производстве и поведении материалов в условиях микрогравитации на орбитальной станции⁸. Но это теория и базы данных. Можно привести и практические результаты использования теоретического инструментария, например, открытие нового вида высокопрочного и износостойкого стекла⁹ или материала, сочетающего металлические и полярные (диэлектрические) свойства¹⁰. Эти открытия были сделаны путём широкого использования расчётов.



Средства организации данных о материалах (The Materials Data Facility (MDF))

Учитывая важность материалов для достижения высокого уровня конкурентоспособности промышленности США, в июне 2014 г. Национальный консорциум сервисов данных (National Data Service (NDS)) объявил о пилотном проекте разработки средств для органи-

зации доступа к данным о материалах – The Materials Data Facility (MDF)¹¹, поддерживаемом крупнейшим разработчиком и распространителем баз данных – NIST. Этот проект является ответом на инициативу MGI Белого дома по ускорению разработки современных материалов. MDF обеспечит материаловедов масштабируемым репозиторием для хранения экспериментальных и расчётных данных, в том числе и до их публикации, снабжённых ссылками на соответствующие библиографические источники. MDF становится рычагом для создания американской национальной инфраструктуры коллективного использования информации, включая разработанные в мире БД по свойствам материалов и информационные системы для их расчёта и моделирования, а также будет способствовать организации обмена данными о материалах, в том числе (и это очень важно) и ещё не опубликованными. Доступность данных и средств расчёта обеспечивается современной информационной и телекоммуникационной инфраструктурой, которая позволяет предоставить данные исследователям материалов для многоцелевого использования, дополнительного анализа и проверки. Помимо NIST, среди исполнителей MDF необходимо выделить University of Chicago, Argonne National Laboratory, The University of Illinois, Northwestern University, Center for Hierarchical Materials Design и т.д. Репозиторий MDF сейчас включает, помимо многочисленных БД NIST¹², информационные системы с результатами квантовомеханических расчётов: AFLOW¹³,

⁶ Сайт NIST: <https://www.nist.gov/coe/advanced-materials-center-excellence>

⁷ Репозиторий – место, где хранятся и поддерживаются какие-либо данные.

⁸ Сайт NASA: <https://www.nasa.gov/marshall/news/news/releases/2015/15-070.html>

⁹ Сайт University of Chicago: http://www.uchicago.edu/features/microscopic_animals_inspire_innovative_glass_research/

¹⁰ Сайт University of Wisconsin-Madison: <http://news.wisc.edu/new-material-combines-useful-typically-incompatible-properties/>

¹¹ Сайт National Data Service: The Materials Data Facility. <http://www.nationaldataservice.org/mdf/>

¹² Сайт NIST Data Gateway: <http://srdata.nist.gov/gateway/gateway?dblist=0>

¹³ Curtarolo S., Setyawan W., Wang S., et al. AFLOW-LIB.ORG: A distributed materials properties repository from high-throughput ab initio calculations // *Comp. Mat. Sci.* 2012. V. 58. P. 227–235; Taylor R.H., Rose F., Toher C., et al. RESTful API for exchanging materials data in the AFLOWLIB.org consortium // *Comp. Mat. Sci.* 2014. V.93. P. 178–192.

The Open Quantum Materials Database (OQMD)¹⁴ и т.д.



Программа Поиска Новых Материалов (Novel Materials Discovery Laboratory (NoMaD))

Эта программа была ответом Евросоюза на американскую стратегическую инициативу MGI. Проект NoMaD¹⁵ направлен на создание Европейских центров превосходства (European Centres of Excellence) и предполагает разработку сети БД (Materials Encyclopedia) по свойствам веществ и материалов, а также средств анализа этих данных и расчёта веществ. Цель – ускорение разработки и использования материалов с заданными функциональными свойствами. Программа стартовала в ноябре 2015 г. в рамках проекта ЕС HORIZON2020 (объём финансирования около 5 млн евро). Существенным недостатком NoMaD является ориентация на информационные ресурсы США (главным образом, БД NIST по свойствам веществ и материалов) и информационные системы с расчётными данными. В настоящее время репозиторий NoMaD¹⁶ содержит,

в основном, результаты квантовомеханических расчётов уже полученных соединений. Программа NoMaD во многом коррелирует с проектом Евросоюза Materials design at the eXascale (MaX)¹⁷, включающим создание инфраструктуры для проведения квантовомеханических расчётов с использованием высокопроизводительных компьютерных систем (объём финансирования – свыше 4 млн евро). Среди исполнителей NoMaD следует отметить ведущие организации Европы, такие как Humboldt University, Fritz-Haber-Institute of the Max Planck Society, King's College London, University of Barcelona, Aalto University, Max Planck Institute for the Structure of Dynamics of Matter, Technical University of Denmark, Max Planck Computing and Data Facility, Barcelona Supercomputing Centre и т.д.



Инициатива исследования материалов путём интеграции информации ("Materials research by Information Integration Initiative" (M²I))

Естественно, что Япония не могла остаться в стороне в таком важном вопросе как ускоренное развитие материаловедения. Поэтому в 2015 г. японское правительство создало на базе Национального института материаловедения (National Institute for Materials Science (NIMS)) – крупнейшего разработчика и распространителя БД по свойствам материалов – Center for Materials Research

¹⁷ Сайт Materials design at the eXascale: http://cordis.europa.eu/project/rcn/198340_en.html

¹⁴ Saal, J.E., Kirklín, S., Aykol, M., et al. Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD) // JOM. 2013. V.65. P. 1501–1509.

¹⁵ Сайт The Novel Materials Discovery (NOMAD) Laboratory: <http://nomad-lab.eu/>; Сайт The Novel Materials Discovery (NOMAD) Laboratory: http://cordis.europa.eu/project/rcn/198339_en.html

¹⁶ Сайт The NoMaD Repository: <http://nomad-repository.eu/cms/>

by Information Integration¹⁸, и объявило свою ещё более заумно названную, чем MGI, "Инициативу исследования материалов путём интеграции информации" (Materials research by Information Integration Initiative (MI²I)). В отличие от европейских программ созданный центр ставит своей задачей не только широкое использование квантовомеханических расчётов, но и поддержку и развитие имеющихся в Японии БД по свойствам веществ и материалов¹⁹, их интеграцию с зарубежными информационными системами, а также применение методов искусственного интеллекта для прогноза новых веществ²⁰.

Китайская Инициатива Геном Материалов (Materials Genome Initiative (MGI–China))

В Китае при поддержке Министерства науки и технологий (Ministry of Science and Technology (MOST)) стартовал свой пятилетний проект MGI (2016–2020 г.г.)²¹. Ранее в 2014–2015 г.г. уже были образованы несколько центров MGI: Shanghai Institute of Materials Genome (2014), Beijing Key Laboratory for Materials Genome (2015) и International Institute of MGI в Ningbo (2015). Цели проекта аналогичны американским.

¹⁸ Сайт Center for Materials Research by Information Integration: <http://www.nims.go.jp/eng/research/MII-1/index.html>

¹⁹ Сайт NIMS Materials Database (MatNavi): http://mits.nims.go.jp/index_en.html

²⁰ Lee J., Seko A., Shitara K., Tanaka I. Prediction model of band-gap for AX binary compounds by combination of density functional theory calculations and machine learning techniques // *Phys. Rev.* 2016. V. B93. N. 11. P. 115104; Toyoura K., Hirano D., Seko A., et al. Machine-learning-based selective sampling procedure for identifying the low-energy region in a potential energy surface: A case study on proton conduction in oxides // *Phys. Rev.* 2016. V. B93. N. 5. P. 054112.

²¹ Lu X.-G. Remarks on the recent progress of Materials Genome Initiative // *Sci. Bull.* 2015. V. 60. N. 22. P. 1966–1968.

Анализ реализуемых в мире крупных инфраструктурных проектов информационного обеспечения в области материаловедения

Следует отметить общие тенденции в разработке систем информационного обеспечения в материаловедческих областях:

- создание виртуально интегрированной сети БД по свойствам веществ и материалов, доступных в Интернете;
- разработка и широкое применение расчётных методов;
- создание БД с расчётной информацией.

Анализ целей и предлагаемых в вышеуказанных инициативах методов и технологий их достижения показывает, что наиболее перспективны проекты США и, быть может, Китая. Именно они позволят создать полноценную инфраструктуру информационной поддержки инновационной деятельности в разработке и внедрении новых материалов, обеспечив науку и промышленность достоверными и полными данными о свойствах веществ и материалов и разнообразным инструментарием (пакеты квантовомеханических и термодинамических расчётов, data mining и т.д.) для расчётов параметров веществ и оптимизации технологии их получения и обработки. Японская инициатива более ограничена. Она основана на использовании системы БД по свойствам веществ и материалов, в основном NIMS, а также использует имеющийся у исполнителей задел по применению уже известных расчётных методов (например, широко используемого пакета квантовомеханических расчётов VASP)²². Начаты работы по применению перспективных методов искусственного интеллекта. К тому же прагматичные японские специалисты ограничили сферу деятельности материалами для электроники

²² Сайт VASP: <https://www.vasp.at/>

(источники питания, магнитные, термоэлектрические и спинтронные материалы). Проекты ЕС на их начальном этапе выглядят наименее перспективными. Ориентация на американские БД по свойствам веществ и пока только расчёты значительно снижают потенциал и возможности этих инфраструктурных проектов. Тем не менее, следует отметить, что объективной предпосылкой для успешной реализации предложенных в США, ЕС, Китае и Японии инициатив являются, с одной стороны, успехи в разработке и применении методов расчёта свойств веществ, и, с другой стороны, наличие множества баз данных по свойствам веществ и материалов, разработанных в последние годы в разных странах.²³ Несмотря на то, что на создание и поддержку таких информационных систем затрачены сотни млн долл., их использование экономически выгодно, т.к. они позволяют значительно сократить затраты на разработку новых материалов за счёт уменьшения дублирования исследований и оперативного предоставления пользователям самой свежей и достоверной информации о свойствах веществ. В свою очередь, расчётные методы дают возможность ещё до экспериментов оценить параметры веществ, указать перспективные для применений составы и разработать технологию получения и обработки материалов, а также автоматизировать процесс обработки данных.

Россия пока даже не “запрягает лошадей”, чтобы присоединиться к всемирной гонке, целью которой будет прорыв в материаловедении и химии. Упование наших чиновников на то, что российским специалистам откроют доступ к создаваемой информационной инфраструк-

туре, вряд ли обоснованны. К сожалению, из-за введённых санкций против России, предполагаемой высокой стоимости доступа к разрабатываемым информационным системам, наличия в них информации о материалах и технологиях двойного назначения и закрытых данных о секретах получения и обработки материалов, доступ российских специалистов к этим информационным ресурсам будет крайне ограничен. Если нам что-либо и перепадёт, то по принципу “на тебе боже, что нам негоже”. Всё это может привести к серьёзному отставанию в темпах разработки и внедрения новых материалов, что будет иметь следствием резкое уменьшение конкурентоспособности российской продукции, особенно в наукоёмких отраслях. Единственный путь решения этой проблемы – это создание собственной инфраструктуры, обеспечивающей науку, образование, промышленность, бизнес, административные органы данными о материалах, технологиях их получения и обработки, сферах применения, производителях и потребителях материалов, а также средствами обработки и анализа накопленной информации, компьютерного моделирования и конструирования новых веществ и материалов, позволяющими принимать решения о выборе материалов для конкретных применений, о перспективности разработки и использования конкретного вещества, о технологических особенностях производства, использования, утилизации и т.д. материалов. Средства, вложенные в такую импортозамещающую программу, достаточно быстро окупятся за счёт сокращения затрат на разработку и исследование новых материалов и прибыли от реализации наукоёмкой продукции, конкурентной на мировом рынке. Имея такую инфраструктуру, можно наладить и взаимовыгодный обмен информацией и вычислительным инструментарием с владельцами крупнейших информационных систем за рубежом.

²³ Киселева Н.Н., Дударев В.А., Земсков В.С. Компьютерные информационные ресурсы неорганической химии и материаловедения // Успехи химии. 2010. Т. 79. № 2. С. 162–188; Сайт БД IRIC (Information Resources on Inorganic Chemistry). <http://iric.imet-db.ru/>

Опыт разработки интегрированной информационной системы по свойствам неорганических веществ и материалов

Предпосылкой для успешного выполнения такого инфраструктурного проекта в России является наш опыт разработки и интеграции БД по свойствам неорганических веществ и материалов, доступных из сети Интернет, также методов и программных средств для компьютерного конструирования новых веществ и материалов, основанных на использовании технологий data mining, и в первую очередь, методов распознавания образов по прецедентам²⁴. Следует отметить, что интерес к применению методов data mining в неорганическом материаловедении связан с объективными трудностями, возникающими при квантовомеханических расчётах ещё не полученных многокомпонентных неорганических веществ, особенно в твёрдой фазе. Например, для того, чтобы рассчитать электронную структуру неорганического соединения с использованием пакета VASP, необходимо знать его кристаллическую структуру, т.е. нужно получить и исследовать это соединение. С помощью же методов распознавания образов, проанализировав имеющуюся информацию об уже известных веществах, хранящуюся в БД, можно прогнозировать ещё не синтезированные соединения и оценивать некоторые их свойства, зная только хорошо известные параметры компонентов (химических элементов или более простых соединений). Для решения этой задачи в ИМЕТ РАН разработана специальная информационно-аналитическая система (ИАС) (рис. 2),

²⁴ Киселева Н.Н., Дударев В.А., Земсков В.С. Компьютерные информационные ресурсы неорганической химии и материаловедения // Успехи химии. 2010. Т. 79. № 2. С. 162–188; Киселева Н.Н., Дударев В.А., Столяренко А.В. Интегрированная система баз данных по свойствам неорганических веществ и материалов // Теплофизика высоких температур. 2016. Т. 54. № 2. С. 228–236.

включающая виртуально интегрированную систему БД по свойствам неорганических веществ и материалов, подсистемы поиска закономерностей в данных, прогнозирования новых соединений и оценки их свойств, базу знаний, базу прогнозов и другие подсистемы.

Интегрированная система баз данных по свойствам неорганических веществ и материалов в настоящее время объединяет информационные системы, разработанные в ИМЕТ РАН в сотрудничестве с ведущими научными организациями России: по фазовым диаграммам полупроводниковых систем (“Диаграмма”), по свойствам акустооптических, электрооптических и нелинейнооптических веществ (“Кристалл”), по ширине запрещённой зоны неорганических веществ (“Bandgap”), по свойствам неорганических соединений

Рис. 2.
Схема информационно-аналитической системы для конструирования неорганических соединений.



("Фазы") и по свойствам химических элементов ("Elements"), а также БД "AtomWork" по свойствам неорганических веществ, разработанную в National Institute for Materials Science (Япония), и БД ТКВ по термическим константам веществ, разработанную в ОИВТ РАН и МГУ.

Система компьютерного конструирования неорганических соединений. Разработка методов прогнозирования возможности образования и свойств неорганических соединений на основе знания только свойств компонентов является одной из важнейших задач химии и материаловедения. Её решение позволяет резко сократить затраты на поиск новых веществ с заданными свойствами. В настоящее время существует **два основных подхода к прогнозированию** ещё не полученных соединений и расчёту их свойств. **Первый** – это разнообразные квантовомеханические методы расчёта свойств веществ "из первых принципов" (т.е. на основе информации об электронном строении химических элементов, входящих в состав соединений). Несмотря на успехи квантовомеханических методов, связанные, в первую очередь, с появлением в последние годы супер-ЭВМ, эти методы ещё не позволяют рассчитать сложные неорганические соединения, особенно в твёрдом состоянии²⁵. В связи с этим химики и материаловеды часто используют **второй подход** – эмпирические методы прогнозирования неорганических соединений и оценки их свойств. **Суть этих методов состоит в поиске закономерностей в экспериментальной информации.** Очень часто речь идёт о нахождении так называемых областей устойчивости неорганических фаз – областей в пространстве свойств компонентов, которые соответствуют соединениям с определёнными свойствами, например, с заданными типами

кристаллических структур. Как правило, в 2D-пространстве параметров компонентов (в большинстве случаев это свойства химических элементов) не удаётся отделить области устойчивости различных соединений. Поэтому **основная задача разработчиков эмпирических критериев** состоит в поиске алгебраических функций от исходных параметров химических элементов, в пространстве которых можно разделить области устойчивости соединений различных типов. Наиболее распространённые критерии – это, по сути, 2D-проекция расположения точек, соответствующих различным типам соединений, на плоскость, координатами которой являются найденные алгебраические функции от исходных параметров химических элементов – компонентов этих соединений. Процесс поиска "разделяющих" функций весьма трудоёмок (недаром очень часто полученные эмпирические критерии названы именами разработчиков, например, правило Юм-Розери, критерий Маттиаса, правило толерантности Гольдшмидта и т.д.). **Разработанные критерии являются результатом кропотливого анализа экспериментальных данных, а не следствием каких-либо расчётов.** Более того, в большинстве случаев теоретическая физика не может даже объяснить причину успешного выполнения таких правил. Существенным недостатком таких критериев является **необходимость их пересмотра при появлении новых соединений, не подчиняющихся этим правилам.** В этом случае нужно искать или перedefинировать имеющиеся критерии, чтобы учесть особенности новых объектов. К тому же простейшие эмпирические критерии не могут учитывать всю совокупность свойств химических элементов (или более простых соединений), определяющих принадлежность веществ к определённому классу, что часто ведёт к значительному пересечению разных классов веществ в пространстве выбранных свойств элементов. Применение компьютеров и специальных

²⁵ Грибов Л.А. О некоторых базовых положениях при постановке квантовых задач теории строения молекул и молекулярных превращений // Ж. структур. химии. 2010. Т. 51. № 4. С. 631–643.

программ поиска многомерных классифицирующих закономерностей в больших объёмах экспериментальной информации, моделирующих процесс нахождения таких закономерностей человеком, позволяет резко сократить время разработки новых критериев и реконструкции старых в связи с появлением новых данных.

В 60-е годы в ИМЕТ были начаты работы по применению методов обучения ЭВМ распознаванию образов (эти методы часто относят к инструментам искусственного интеллекта) для автоматизации поиска многомерных эмпирических критериев образования неорганических соединений с разными свойствами²⁶ с целью их использования для конструирования новых веществ²⁷. Уже первые эксперименты по применению методов распознавания образов в неорганической химии дали хорошие результаты – специалистам ИМЕТ с 90%-ой точностью удалось найти многомерные критерии, которые позволили разделить двойные химические системы с образованием соединений разного состава. Это был успех, следствием которого было появ-

ление многочисленных работ по применению методов распознавания образов в неорганической химии и материаловедении, выполненных в разных странах. За полвека в ИМЕТ были получены прогнозы возможности образования тысяч новых неорганических соединений в двойных, тройных, четверных и более сложных химических системах²⁸. Использование полученных нами прогнозов позволяет ускорить поиск новых магнитных, полупроводниковых, сверхпроводящих, электрооптических, акустооптических, нелинейнооптических и других материалов. Сотни предсказанных соединений были синтезированы, и сравнение прогнозов с экспериментальными данными, полученными позже публикации наших прогнозов, показали, что средняя достоверность прогнозирования неорганических соединений выше 80%. Применение методов распознавания образов по прецедентам для теоретического конструирования новых неорганических соединений позволяет значительно сократить затраты на поиск новых веществ с заданными свойствами за счёт рационального использования огромной информации, накопленной химией и сконцентрированной в БД по свойствам веществ и материалов, в том числе, и разработанных в ИМЕТ РАН.

²⁶ Савицкий Е.М., Девингталь Ю.В., Грибуля В.Б. Прогноз металлических соединений типа АзВ с помощью электронно-вычислительной машины // Докл. АН СССР. 1968. Т. 183. № 5. С. 1110–1112.

²⁷ В связи с этим очень странно выглядит фраза из Википедии (раздел "Materials Genome") о большом успехе проекта MGI в 2016 г. – "впервые использованных методах обучения ЭВМ для конструирования материалов" (https://en.wikipedia.org/wiki/Materials_Genome). По-видимому, либо репозитории, созданные в рамках MGI, ещё не заполнены достаточным образом и поэтому не содержат информации о многочисленных наших статьях (в том числе и напечатанных в зарубежных журналах) в области применения методов обучения ЭВМ распознаванию образов в химии, либо авторы этого раздела Википедии столь не любопытны и некомпетентны, что поленились провести элементарный поиск соответствующих публикаций в Интернет. Однако этот казус может быть и примером того, когда из-за недостатка информации "изобретается велосипед" – заново тратятся время и средства на открытие давно известных и широко используемых методов.

²⁸ Савицкий Е.М., Грибуля В.Б. Опыт прогнозирования состава и свойств соединений с помощью ЭВМ // Докл. АН СССР. 1970. Т. 190. № 5. С. 1147–1150; Киселева Н.Н. Компьютерное конструирование неорганических соединений. Использование баз данных и методов искусственного интеллекта. М.: Наука, 2005. 288 с.; Бурханов Г.С., Киселева Н.Н. Прогнозирование интерметаллических соединений // Успехи химии. 2009. Т. 78. № 6. С. 615–634; Kiselyova N., Stolyarenko A., Ryazanov V., et al. Application of Machine Training Methods to Design of New Inorganic Compounds // In: "Diagnostic Test Approaches to Machine Learning and Commonsense Reasoning Systems". Ed. By X.A. Naidenova & D.I. Ignatov. Hershey: IGI Global. 2012. P. 197–220; Савицкий Е.М., Грибуля В.Б., Киселева Н.Н. и др. Прогнозирование в материаловедении с применением ЭВМ. М.: Наука, 1990. 86 с.

Вернёмся к рассмотрению разработанной нами специальной системы компьютерного конструирования – ИАС. Её основу составляют алгоритмы и программы распознавания образов по прецедентам, входящие в многофункциональную систему “РАСПОЗНАВАНИЕ”, разработанную в вычислительном центре (ВЦ) РАН²⁹ и объединяющую, помимо широко известных методов линейной машины, линейного дискриминанта Фишера, *k*-ближайших соседей, опорных векторов, нейросетевых и генетических алгоритмов, также уникальные алгоритмы, разработанные в ВЦ РАН: алгоритмы распознавания, основанные на вычислениях оценок, алгоритмы голосования по тупиковым тестам, алгоритмы голосования по логическим закономерностям, алгоритмы статистического взвешенного голосования и т.д. В систему также интегрирована программа обучения ЭВМ процессу формирования понятий ConFor, разработанная в Институте кибернетики НАН Украины³⁰, в основу которой положена оригинальная организация данных в памяти ЭВМ в виде растущих пирамидальных сетей. Для отбора свойств компонентов химических соединений и алгебраических функций от этих свойств, наиболее важных для прогнозирования, в ИАС были включены программы, основанные на алгоритмах³¹. Важным компонентом ИАС явля-

ется подсистема визуализации областей устойчивости разных неорганических фаз в 2D-пространстве наиболее важных для прогнозирования свойств компонентов и/или алгебраических функций от этих свойств, отобранных компьютерными программами. **Использование искусственного интеллекта позволяет в течение часов или даже минут проанализировать хранящуюся в базах данных информацию о неорганических соединениях и найти многомерные эмпирические критерии образования тех или иных соединений** (например, критерии образования соединений с определённым типом кристаллической структуры, широкозонных полупроводников и т.д.), а также визуализировать 2D-проекции точек, соответствующих различным классам соединений, в пространствах наиболее удачных сочетаний параметров компонентов, и осуществить прогнозирование ещё не полученных соединений и оценку некоторых их свойств. Следует отметить, что на этапе прогнозирования используются только данные о свойствах компонентов соединений (химических элементов или более простых соединений).

Проект инфраструктуры обеспечения данными российских специалистов в области неорганической химии и материаловедения

Разработанная нами ИАС является, своего рода, одним из пилотных проектов создания информационной инфраструктуры для неорганического материаловедения. В ней виртуально интегрированы наиболее известные российские БД в этой области, а также начата их интеграция с зарубежными информационными системами. Большинство российских БД, включённых в ИАС, содержат ссылки на полные тексты публикаций, из которых извлечена информация, хранящаяся в БД. Подсистема компьютерного конструирования соединений позволяет

²⁹ Журавлев Ю.И., Рязанов В.В., Сенько О.В. “РАСПОЗНАВАНИЕ”. Математические методы. Программная система. Практические применения. М.: ФАЗИС, 2006. 176 с.

³⁰ Гладун В.П. Процессы формирования новых знаний. София: СД “Педагог-6”, 1995. 192 с.

³¹ Senko O.V. An Optimal Ensemble of Predictors in Convex Correcting Procedures // Pattern Recognition and Image Analysis. 2009. V. 19. N. 3. P. 465–468; Yuan G.-X., Ho C.-H., Lin C.-J. An Improved GLMNET for L1-regularized Logistic Regression // J. Machine Learning Research. 2012. V. 13. P. 1999–2030; Yang Y., Zou H.A. Coordinate Majorization Descent Algorithm for L1 Penalized Learning // J. Statistical Computation & Simulation. 2014. V. 84. N. 1. P. 1–12.

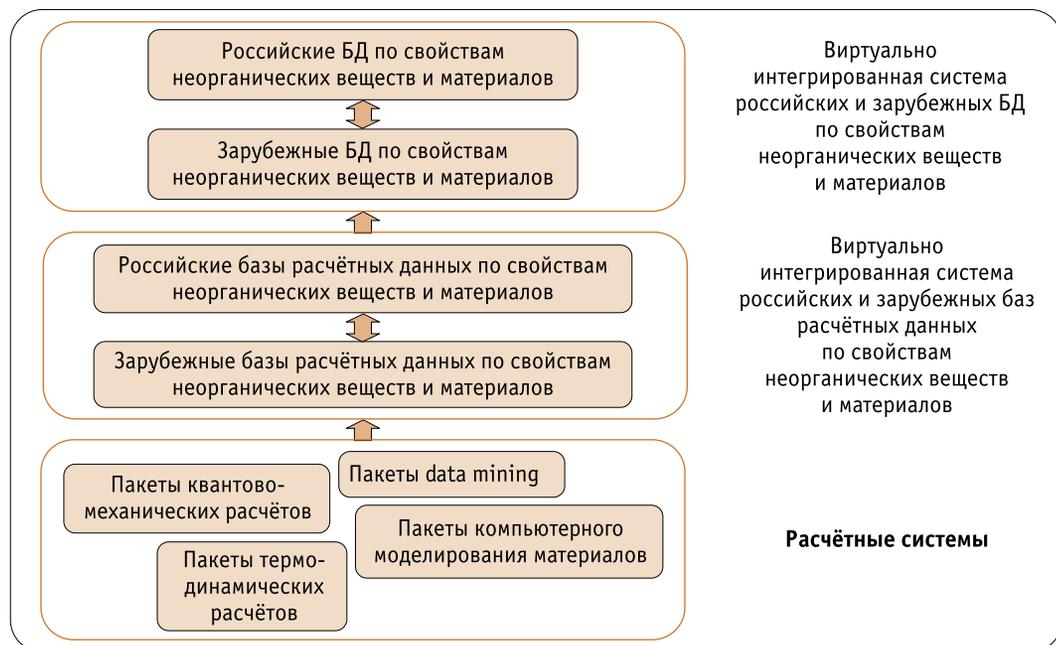
найти закономерности в информации БД и использовать их для прогнозирования ещё не полученных соединений и оценки их свойств. Полученные прогнозы хранятся в специальной базе прогнозов, что расширяет функциональные возможности традиционных баз данных (пользователь получает не только известные экспериментальные данные, но и прогнозы ещё не синтезированных соединений и оценки некоторых их свойств).

При разработке российского проекта инфраструктуры информационного обеспечения специалистов в области неорганического материаловедения нужно учитывать всё многообразие возможных запросов пользователей. Вполне естественно, что запросы академических учёных могут кардинально отличаться от запросов инженеров-конструкторов

или производителей материалов. Однако общий проект информационной инфраструктуры должен включать в качестве необходимых элементов виртуально интегрированную систему российских и зарубежных баз данных по свойствам неорганических веществ и материалов, технологиям их получения и обработки, потребителям и производителям материалов и т.д., комплекс пакетов расчёта и моделирования материалов, пользователями которых в большинстве случаев являются академические учёные, и виртуально интегрированную систему баз данных уже рассчитанных значений (рис. 3). Следует подчеркнуть, что **технологии обработки, хранения, организации поиска необходимых сведений требуют разработки и применения самых современных программных средств и создания мощных центров обработки данных.**

Необходимо выделять средства на ежегодное продление лицензий для пользования зарубежными базами и организовать единый портал бесплатного для

Рис. 3.
Схема инфраструктуры обеспечения данными российских специалистов в области неорганической химии и материаловедения.



российских пользователей доступа к ним (сейчас такие базы доступны для ограниченного числа организаций). Необходимо всячески поддерживать перевод в электронную форму коллекций наиболее известных в мире российских журналов, что несомненно будет способствовать повышению их авторитета и цитируемости.

Оснащение исследовательских организаций системами расчёта необходимо начинать, в первую очередь, с обучения студентов и аспирантов пользованию наиболее известными пакетами квантовомеханических, термодинамических, статистических и т.д. расчётов. К сожалению, за последние четверть века отток специалистов в области теоретической химии сильно обескровил всемирно известные школы российских университетов и академических институтов, что значительно усложняет решение проблемы квалифицированного обучения молодых специалистов, а также использования расчётных методов. Нужно разрабатывать российские базы данных расчётных значений и интегрировать их с зарубежными информационными системами, которые сейчас ещё открыто доступны в Интернете, что позволит частично решить проблему квалифицированных расчётов веществ. Постановка экспериментов должна включать проведение или использование расчётов в качестве начального этапа исследований, что позволит сократить время и затраты на поиск и разработку новых материалов.

Важным компонентом разрабатываемого инфраструктурного проекта должна стать *подсистема анализа запросов пользователей*, особенно, специалистов в прикладных областях. Именно она позволит выявить группы материалов, на изучение которых нужно направить экспериментальные исследования. Статистика отказов в выдаче значений того или иного параметра вещества может стать стимулом к дополнительному экспериментальному исследованию запрашиваемого свойства.

В этой статье не рассматриваются вопросы, связанные с материаловедением органических веществ и материалов. Учитывая тот факт, что количество известных органических соединений на порядок больше, чем неорганических, необходимость создания информационной инфраструктуры в этой области крайне важна.

Подведём итог. Российским ответом на стратегические инициативы США, ЕС, Китая, Японии по созданию информационной инфраструктуры в области материаловедения может быть федеральная целевая программа, аналогичная MGI, направленная на организацию инфраструктуры доступа к экспериментальным и расчётным данным о материалах, а в дальнейшем создание федерального информационного центра, обеспечивающего специалистов информацией о свойствах веществ и материалах, технологиях их производства, а также расчётными данными, патентной информацией и т.д. Федеральная целевая программа и создание информационного центра, интегрирующего информационные ресурсы в области материаловедения, будет способствовать резкому ускорению поиска, разработки и внедрения новых материалов, в сочетании со значительным сокращением затрат за счёт уменьшения дублирования исследований и оперативного предоставления химикам и материаловедам достоверной экспериментальной и расчётной информации о веществах и материалах, а также позволит повысить конкурентоспособность российской продукции, особенно в высокотехнологических областях³².

³² Автор благодарит В.А. Дударева, А.В. Столяренко, В.В. Рязанова, О.В. Сенько, А.А. Докукина за помощь в разработке систем компьютерного конструирования неорганических соединений и баз данных по свойствам неорганических веществ и материалов.