Информационно-аналитическая система для конструирования новых неорганических соединений

© Н. Н. Киселева © В. А. Дударев © А. В. Столяренко
Федеральное государственное бюджетное учреждение науки Институт металлургии и материаловедения им. А. А. Байкова Российской академии наук (ИМЕТ РАН), Москва
kis@imet.ac.ru vic@imet.ac.ru stol-drew@yandex.ru

Аннотация

Рассмотрены принципы разработки систем поиска закономерностей в виртуально интегрированных распределенных базах данных. Разработана методология интеграции программ анализа данных, основанных на различных алгоритмах. Предложенные методы использованы для создания информационно-аналитической системы для автоматизации процесса компьютерного конструирования новых неорганических соединений, основанной на использовании программ распознавания образов для поиска закономерностей в информации баз данных по свойствам неорганических веществ и материалов. Приведены примеры использования разработанной системы для конструирования новых неорганических соединений.

Авторы благодарят В. В. Рязанова, О. В. Сенько, А. А. Докукина за помощь в создании ИАС.

Работа выполнена при частичной финансовой поддержке РФФИ, проекты 12-07-00142, 14-07-00819, 14-07-31032 и 15-07-00980.

1 Введение

Методы обнаружения сложных закономерностей в данных широко применяются в различных областях науки и техники [1, 2]. Этим термином сегодня обозначают процесс получения из “сырых” первичных данных новой, потенциально полезной информации о взаимосвязях между объектами и явлениями в конкретной предметной области. Процедура обнаружения закономерностей включает несколько этапов. Сюда относится накопление сырых данных, оценка, преобразование, подготовка информации к вводу в базу данных, поиск закономерностей в данных, анализ найденных закономерностей и их использование для принятия решения. Это направление информатики в настоящее время выдвигается на передний план, как по развитию инструментальных средств, так и по приложениям. Причиной этому служит создание интегрированных систем баз данных (БД) и развитие методов интеллектуального анализа данных, которые в англоязычной литературе обозначаются термином data mining. Программы data mining наряду с базами данных являются ядром систем обнаружения сложных закономерностей. В настоящее время data mining объединяет в себе алгоритмы методов прикладной статистики, распознавания образов, искусственного интеллекта и пр.

В настоящей статье рассмотрены принципы организации и результаты разработки информационно-аналитической системы (ИАС), предназначенной для поиска закономерностей в больших объемах данных о свойствах неорганических веществ и использования найденных закономерностей для конструирования еще не полученных неорганических соединений и оценки их свойств. В этой системе интегрированы БД по свойствам неорганических веществ и материалов, разработанные в разных организациях и странах, а подсистема поиска закономерностей объединяет программы, основанные на различных алгоритмах распознавания образов.

2 Интегрированная система баз данных как информационная основа ИАС

Интегрированная система БД по свойствам неорганических веществ и материалов [3, 20], являющаяся информационным ядром разработанной нами ИАС, в настоящее время объединяет шесть баз данных: БД по свойствам неорганических соединений “Фазы”, БД по фазовым диаграммам систем с полупроводниковыми фазами “Диаграмма”, БД по веществам с особыми акустооптическими, электрооптическими и нелинейно-оптическими свойствами “Кристалл”, БД по ширине запрещенной зоны неорганических соединений

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

“BandGap” и БД по свойствам химических элементов «Элементы» [18, 22, 22, 25], разработанные ИМЕТ РАН совместно с другими организациями России, и БД “AtomWork” по свойствам неорганических соединений, созданную в National Institute for Materials Science (Япония) [12]. Интегрированная система баз данных доступна зарегистрированным пользователям из сети Интернет (http://imet-db.ru/).

3 Подсистема поиска закономерностей

При разработке подсистемы анализа данных важнейшей задачей является отбор методов data mining, наиболее подходящих для поиска закономерностей для конкретной предметной области. Как правило, решение такой задачи выполняется методом «проб и ошибок». При отборе методов распознавания образов для анализа химической информации учитывался многолетний опыт применения этих методов для конструирования неорганических соединений [18]. В результате были выбраны следующие методы и программы:

- -набор алгоритмов системы РАСПОЗНАВАНИЕ, разработанной в ВЦ РАН [17]. Эта многофункциональная система распознавания образов, помимо широко известных методов линейной машины, линейного дискриминанта Фишера, k-ближайших соседей, опорных векторов, нейросетевых и генетических алгоритмов, включает алгоритмы, разработанные в ВЦ РАН: алгоритмы распознавания, основанные на вычисления оценок, алгоритмы голосования по тупиковым тестам, алгоритмы голосования по логическим закономерностям, алгоритмы статистического взвешенного голосования и т.д.;
- система обучения ЭВМ процессу формирования понятий ConFor, разработанная в Институте кибернетики НАН Украины [16]. Система основана на организации данных в памяти ЭВМ в виде растущих пирамидальных сетей.

Выбор вышеуказанных программ анализа данных обусловлен в первую очередь универсальностью относительно размерностей решаемых задач. Системы дают возможность решения как задач прогноза редких или уникальных событий, явлений или процессов, когда начальная (обучающая) информация мала (десятки прецедентов), так и задач больших размерностей (десятки тысяч прецедентов).

Как правило, заранее невозможно указать, какой алгоритм является наиболее эффективным при решении конкретной задачи. В связи с этим перспективным является использование методов распознавания коллективами алгоритмов. При синтезе коллективного решения во многих случаях удается компенсировать возможные ошибки распознавания отдельных алгоритмов правильными ответами других алгоритмов. Исходя из этого, в разработанную ИАС включены программы, реализующие разные стратегии принятия коллективных решений: основанные на: комитетных методах, методе Байеса, методе выпуклого стабилизатора, динамическом методе Вудса, методах, использующих области компетенции, шаблоны принятия решений, логическую коррекцию и т.д. [17].

Вышеуказанные алгоритмы распознавания образов и принятия коллективных решений основаны на различных принципах и используют различные формы представления искомых закономерностей (алгебраические функции, логические выражения, нейронные или растущие пирамидальные сети и т.д.). Для интеграции таких разнородных программ была использована сервисно-ориентированная архитектура (SOA), которая позволила учесть различия в данных и информационных структурах, используемых в интегрируемых программах, сложные механизмы их взаимодействия и обеспечила возможность достаточно простого добавления новых программ анализа данных в подсистему поиска закономерностей. При интеграции приложений вместо специализированных интерфейсов между отдельными программами применена связующая среда, которая играет роль универсального программного ядра, соединяющего все приложения [5, 7, 24]. Преимуществом используемой технологии на основе интегрирующей среды является, в первую очередь, простота поддержки и расширения разработанной на ее основе системы.

4 Информационно-аналитическая система для компьютерного конструирования неорганических соединений

Результатом практической реализации вышеуказанных принципов была разработка информационно-аналитической системы для компьютерного конструирования неорганических соединений, основанной на использовании методов обучения ЭВМ распознаванию образов для поиска закономерностей в информации баз данных по свойствам неорганических веществ и материалов [5, 7, 20, 24].

Помимо интегрированной системы БД [3, 18, 20-23, 25] и программ распознавания образов [16, 17] в состав ИАС (рис.1) входят подсистемы поиска классифицирующих признаков и визуализации полученных результатов, база знаний, база прогнозов для различных классов неорганических веществ и управляющая подсистема.

Для отбора классифицирующих свойств химических элементов в ИАС были включены программы [10, 13, 14]. Отбор свойств химических элементов, наиболее информативных для классификации веществ, имеет двоякое значение. С

одной стороны, удается резко сократить признаковое описание, которое для многокомпонентных веществ включает сотни свойств элементов. С другой стороны, выбор свойств элементов, наиболее важных для классификации химических веществ, дает возможность физической интерпретации найденных закономерностей, что повышает доверие к полученным прогнозам и позволяет найти существенные причинно-следственные связи между параметрами объектов и разработать физические и химические модели явлений.

Рис. 1 Схема ИАС для поиска закономерностей в химической информациии прогнозе новых соединений.

Подсистема визуализации строит проекции расположения объектов в двумерных пространствах свойств компонентов химических веществ, включающих не только исходные параметры компонентов, но и указанные пользователем алгебраические функции от этих параметров.

База знаний содержит полученные классифицирующие закономерности. При ее программной реализации возникла проблема, связанная с тем, что форма представления знаний в используемых методах распознавания образов существенно различается. В связи с этим было предложено новое программное решение для хранения полученных закономерностей, а также сопутствующей информации о параметрах программ и исследуемых объектов [24]. Хранение этой информации реализовано средствами SQL-сервера и файловых структур на дисках сервера. На сервере хранятся полученные закономерности в специальном внутреннем формате программ анализа данных, а в таблицах БД на SQL-сервере - служебная информация об этих закономерностях, а именно: уникальный идентификатор закономерности, обозначение прогнозируемой характеристики, формульный состав химических соединений, обозначения свойств химических элементов, используемых для описания веществ, пути к файлам на дисках, фамилия специалиста, проводившего оценку исходных данных и поиск закономерностей, дата формирования закономерности и т.д.

В базе прогнозов содержатся результаты предыдущих компьютерных экспериментов, а также ссылки на служебную информацию, хранящуюся в базе знаний. Использование базы прогнозов позволило повысить функциональность баз данных по свойствам неорганических веществ и материалов ИМЕТ РАН за счет предоставления пользователю не только известных сведений об уже изученных веществах, но и прогнозов еще не полученных неорганических соединений и оценок их свойств. В настоящее время идет заполнение этой базы.

Управляющая подсистема организует вычислительный процесс и осуществляет взаимодействие между функциональными подсистемами ИАС, а также обеспечивает доступ к системе из сети Интернет. Помимо этого, управляющая подсистема предоставляет пользователю программные средства для подготовки данных для анализа, выдачи отчетов и реализации других сервисных функций. В частности, для извлечения из БД информации, которая после оценки экспертом используется для компьютерного анализа, разработана специальная подсистема. Она предоставляет эксперту возможности редактирования найденной информации и формирования признакового описания. В последнем случае эксперт только отмечает выбранные свойства химических элементов в специальной таблице-меню, и подсистема подготовки выборки для анализа извлекает выбранные значения свойств из БД «Элементы», если нужно, то формирует сложные признаки в виде алгебраических функций от исходных, и «склеивает» признаковое описание в форме Excel-таблицы, которая затем поступает на вход системы анализа данных. Подсистема выдачи результатов предназначена для предоставления прогнозов в привычной для химиков и материаловедов табличной форме.

Важной особенностью программной реализации ИАС является то, что клиентская часть полностью построена на базе Web-интерфейса. То есть пользователи работают с ИАС посредством Web-браузера.

Процессы обучения и распознавания в ИАС реализованы с помощью специального

асинхронного Web-сервиса, что позволяет решать длительные по времени задачи обучения и прогнозирования в среде Интернет, в которой возможны сбои. Асинхронный Web-сервис позволяет пользователям инициировать длительное выполнение ресурсоемких операций, контролировать степень их выполнения в асинхронном режиме, получать оповещение о готовых результатах расчетов, прерывать выполнение задач с сохранением промежуточных результатов.

Разработанная ИАС доступна зарегистрированным пользователям из сети Интернет (http://ias.imet-db.ru).

5 Использование ИАС для конструирования неорганических соединений

Применение разработанной ИАС позволило сконструировать тысячи еще не полученных неорганических соединений в двойных, тройных и более сложных химических системах, а также оценить некоторые их свойства. Ниже приведены некоторые примеры полученных результатов. Для расчетов использовался сервер 2x Intel Xeon E5-2650v2 / 128Gb RAM 1866 / Adaptec RAID 6405 / 4x 4TB HotSwap SATA / sDVD±RW / 2x GLAN / IPMI+ / 2x 750W HotSwap. Время обработки данных варьировалось от нескольких секунд до нескольких часов в зависимости от алгоритма и объема анализируемых данных.

5.1 Прогноз возможности образования соединений состава AB_3X_3

Целью работы [19] было конструирование соединений состава AB_3X_3 (A и B – здесь и далее разные химические элементы; X = S, Se или Te), аналогичных пруститу ($AsAg_3S_3$) и пираргириту ($SbAg_3S_3$), применяемых в нелинейно-оптических и электрооптических устройствах. Для компьютерного анализа использовалась информация БД «Фазы», входящей в состав ИАС, о 117 примерах образования соединений состава AB_3X_3 и 58 примерах отсутствия соединений этого состава. Каждое вещество описывалось набором более 240 параметров компонентов A, B и X (химических элементов и простых халькогенидов). При прогнозировании были использованы только эти широко известные значения свойств компонентов.

За 6 лет, прошедшие после публикации наших прогнозов [19], были экспериментально проверены 25 составов (таблица 1). В таблице приняты следующие обозначения: 1 – прогноз образования соединения AB_3X_3 при обычных условиях (298 K и 1 атм); 2 - прогноз отсутствия соединения состава AB_3X_3 при обычных условиях; знаком # здесь и далее отмечены примеры, информация о которых использована для обучения ЭВМ; пустые клетки – здесь и далее неопределенный прогноз; здесь и далее серым отмечены результаты совпадения прогнозов с экспериментальными данными.

Следует отметить, что все наши прогнозы образования при обычных условиях соединений состава AB_3X_3 совпали с экспериментом.

Таблица 1. Прогноз возможности образования соединений состава AB_3X_3

X	**S**															
A B	**Fe**	**Ga**	**In**	**Sn**	**Sb**	**La**	**Ce**	**Pr**	**Nd**	**Pm**	**Sm**	**Eu**	**Gd**	**Tb**	**Dy**	**Bi**
K	1	#2	1	1	#1	1		1	1	1	1	1	1	1	1	#2
Rb	1		#1	1	1	1			1	1	1	1	1	1	1	#1
Tl	1	2	#1	#1	#1	#2	2	#2	2	2	2	2				
X	**Se**															
A B	**Fe**	**Ga**	**In**	**Sn**	**Sb**	**La**	**Ce**	**Pr**	**Nd**	**Pm**	**Sm**	**Eu**	**Gd**	**Tb**	**Dy**	**Bi**
K	#1	#1	1	1	#1		1			1	1	1	1	1	1	#1
Rb	1	1	1	1	1	1	1			1	1	1	1	1	1	#1
Ag	2	#2	2		#2	2	2	2	2	2	2	2		2	2	2
Cs	1	#1	1	1	1	1				1	1	1	1	1	1	#1
Tl	1	2	#2	#1	#1	2	2	2	2							#2
X	**Te**															
A B	**Fe**	**Ga**	**In**	**Sn**	**Sb**	**La**	**Ce**	**Pr**	**Nd**	**Pm**	**Sm**	**Eu**	**Gd**	**Tb**	**Dy**	**Bi**
Rb	1	1	1	1	1	1	1			1	1	1	1	1	1	1
Ag	2	#2	#2	2	#2	2	2	2	2	2	2	2	#2	2	2	2
Cs	1	1	1	1	1	1	1			1	1	1	1	1	1	1
Tl	1	2	2	#1	#2	#2	2	2	2	2						#2

5.2 Прогноз типа кристаллической структуры соединений состава ABX_2

Халькогенидные соединения состава ABX_2 (X = S, Se или Te) относятся к перспективному классу полупроводниковых и нелинейно-оптических веществ. Особый интерес для исследователей представляют соединения этого состава с кристаллической структурой типа халькопирита, которые используются в качестве материалов для солнечных батарей (например, $CuInS_2$, $CuInSe_2$ или $CuGaSe_2$) [9, 11], или нелинейно-оптических устройств [8] (например, $ZnGeP_2$, $AgGaSe_2$ или $AgGaTe_2$).

Нами было проведено прогнозирование новых соединений состава ABX_2 и предсказан тип их кристаллической структуры при обычных условиях [4]. Для компьютерного анализа была использована выборка из более тысячи известных веществ, каждое из которых было описано набором более 240 параметров компонентов A, B и X. В таблице 2 дан фрагмент полученных результатов. Приняты следующие обозначения: 1 – прогноз соединения ABX_2 с кристаллической структурой типа α-$NaFeO_2$ при обычных условиях; **2** – прогноз соединения ABX_2 со структурой типа NaCl; **3** – прогноз

соединения ABX_2 со структурой типа халькопирита; **4** – прогноз соединения ABX_2 со структурой типа TlSe; **5** – прогноз соединения ABX_2 с кристаллической структурой, отличной от приведенных выше; **6** – прогноз отсутствия соединения состава ABX_2 при обычных условиях; черным отмечены результаты несовпадения прогнозов с экспериментальными данными.

Таблица 2. Прогноз типа кристаллической структуры соединений состава AB_3X_3

X	S					Se				Te				
A / B	Cu	Rb	Ag	Cs	Tl	Li	Na	K	Tl	K	Rb	Ag	Cs	Tl
B	3	#5	3	#5	#5	1	1		4	4	1		1	4
Al	#3		#3		4	#5	#4	#5	#5	#4		#3	4	#4
Cr	#5	#1	#5		#5	2	#1	1		5	5	#5	5	#5
Fe	#3	#5	#3		#5		1	#5	#5	1	1	#3	1	5
Ga	#3	#5	#3	5	#5	5	4	#5	#5	#5		#3		#4
As	#5		#5	5	#5		#5	#5	5	4	5	#2	5	
Y	5	1	#5	5	#1	#1	#1	1	#1		1	#5	1	#1
In	#3	#5	#3	#5	#4	#5	#1	5	#4	#4		#3	4	#4
Sb	#5	#5	2	5	2	2	5	#5	5	4	#5	#2	#5	#1
La		#1		#1	#6	#5	#1	1	#6	1	1		1	6
Ce	#5	#1		#1	#6	#5	#1	1	#6	1	1		1	#6
Pr	#5	#1		#5		1	#1	1	#1	1	1		1	#1
Nd	#5	#1		#5			#1	1	#1	1	1		1	#1
Sm	#5	#1	#5	#5	#1		#1	1	1	1	1		1	#1
Gd	#5	#1	#5	#5	#1	#1	#1	1	#1	1	1	#5	1	#1
Tb	#5	#1	#5	#5	#1	#1	#1	1	#1		1	5	1	#1
Yb	#5	#1	#5	#5	#1		#1	#1	1	1	1		1	
Bi	#5	#1	#1		#1	#2	#2	#2	#2			1		#1

За 5 лет был проверен 31 наш прогноз. Обнаружено только 4 несовпадения с экспериментальными данными.

5.3 Оценка величины ширины запрещенной зоны для халькопиритов состава ABX_2

Развитие оптоэлектроники и других прикладных областей, в частности, солнечной энергетики, устройств детектирования фотонов и заряженных частиц вызвало интерес к поиску новых широкозонных полупроводников. Широкозонными принято считать полупроводники, у которых энергия межзонных электронных переходов превосходит значение, близкое к 2 эВ [15]. В последнее время внимание исследователей привлекли широкозонные полупроводники с кристаллической структурой халькопирита [11]. В связи с этим для части прогнозированных выше соединений со структурой халькопирита нами была оценена величина ширины запрещенной зоны [6]. Использовалась выборка из 41 соединения (таблица 3). Исходя из физико-химических представлений, в критерий были включены следующие параметры элементов A, B и X (всего 15 параметров):

- функция $\Delta\chi = |2\chi_X - \chi_A - \chi_B|$, где χ_i – электроотрицательности по Мартынову-Бацанову;
- валентности Z_A, Z_B, Z_X (для переходных металлов – номер группы в Периодической системе);
- среднее число валентных электронов: $n = (n_A + n_B + 2n_X)/4$;
- электроотрицательность по шкале Петтифора;
- функция $(I_z/Z)_A - \{6 + 0.1(I_z/Z)_C\}$, где I_z – последний потенциал ионизации.

Таблица 3. Сравнение прогнозов ширины запрещенной зоны (E_g) с экспериментальными данными для изученных халькопиритов

Соединение	$E_{g\ эксп.}$, эВ	Прогноз E_g
$CuAlS_2$	3.5	>2 эВ
$CuGaS_2$	2.44	>2 эВ
$CuInS_2$	1.5	<2 эВ
$CuAlSe_2$	2.67	>2 эВ
$CuGaSe_2$	1.63	<2 эВ
$CuInSe_2$	0.95	<2 эВ
$CuAlTe_2$	2.06	>2 эВ
$CuGaTe_2$	1.18	<2 эВ
$CuInTe_2$	0.88	<2 эВ
$AgAlS_2$	3.13	>2 эВ
$AgGaS_2$	2.75	>2 эВ
$AgAlSe_2$	2.55	>2 эВ
$AgGaSe_2$	1.65	<2 эВ
$AgInSe_2$	1.24	<2 эВ
$AgAlTe_2$	1.8	<2 эВ
$AgGaTe_2$	1.1	<2 эВ
$AgInTe_2$	0.96	<2 эВ
$ZnSiP_2$	2.07	>2 эВ
$ZnSiAs_2$	2.1	>2 эВ
$ZnGeN_2$	2.9	>2 эВ
$ZnGeP_2$	2.1	>2 эВ
$ZnGeAs_2$	1.16	<2 эВ
$ZnSnP_2$	1.45	<2 эВ
$ZnSnAs_2$	0.74	<2 эВ
$ZnSnSb_2$	0.4	<2 эВ
$CdSiP_2$	2.2	>2 эВ
$CdGeP_2$	1.8	<2 эВ
$CdGeAs_2$	0.53	<2 эВ
$CdSnP_2$	1.16	<2 эВ
$CdSnAs_2$	0.3	<2 эВ
$AgInS_2$	1.9	<2 эВ
$CdSiAs_2$	1.51	<2 эВ
$CuFeS_2$	0.53	<2 эВ
$CuFeSe_2$	0.16	<2 эВ
$CuFeTe_2$	0.1	<2 эВ
$LiGaTe_2$	2.31	>2 эВ
$LiInTe_2$	1.46	<2 эВ
$AgFeSe_2$	0.23	<2 эВ
$MgSiP_2$	2.35	>2 эВ
$MnGeP_2$	0.24	<2 эВ
$MnGeAs_2$	0.6	<2 эВ

В таблице 3 приведены результаты оценки ширины запрещенной зоны для известных халькопиритов, для которых в БД “BandGap” хранились сведения о ширине запрещенной зоны. Было отмечено согласие экспериментальных и прогнозируемых значений ширины запрещенной зоны (таблица 3), что подтверждало обоснованность применяемых компьютерных методов. С помощью найденных критериев была оценена ширина запрещенной зоны халькопиритов (таблица 4), для которых в БД “Bandgap” [21] не было соответствующей информации. Согласно нашим прогнозам, халькопириты $ZnAlS_2$ и $ZnAlSe_2$ относятся к широкозонным полупроводникам, перспективным для оптоэлектронных приложений.

Таблица 4. Прогноз E_g для халькопиритов, информация о которых не использована при компьютерном анализе

Соединение	**Прогноз E_g**
$ZnAlS_2$	
$ZnAlSe_2$	
$ZnAlTe_2$	
$AgFeS_2$	
$AgFeTe_2$	
$ZnGaTe_2$	
$CdGaTe_2$	
$HgGaTe_2$	<2 эВ

6 Заключение

Информационно-аналитическая система позволяет решить две важные задачи. Во-первых, она дает возможность частично автоматизировать анализ огромной экспериментальной информации, накопленной химией, для поиска закономерностей в данных и последующего конструирования новых соединений с заданными свойствами. Во-вторых, она расширяет возможности традиционных БД по свойствам веществ и материалов, предоставляя пользователю не только информацию об уже исследованных веществах, но и прогнозы для еще неизученных веществ и их свойств.

Разработанная ИАС широко используется для конструирования новых неорганических соединений, перспективных для создания материалов - активных компонентов полупроводниковых, электрооптических, акустооптических, нелинейно-оптических, лазерных, магнитных и прочих устройств современной электроники. С ее помощью удалось осуществить прогноз возможности образования тысяч новых соединений и оценить некоторые их свойства (тип кристаллической структуры, температуру перехода в сверхпроводящее состояние, температуру плавления и т.д.) [4-7, 18, 19]. Следует отметить, что при прогнозировании неисследованных соединений используется только информация о хорошо известных свойствах компонентов (химических элементов или более простых соединений). Как показывает сравнение прогнозов с экспериментальными данными [18], средняя достоверность прогнозирования неорганических соединений превышает 80 %.

Литература

[1] U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, R. Uthurusamy. Advances in Knowledge Discovery and Data Mining. Cambridge: AAAI Press/The MIT Press. 1996.

[2] D. T. Larose. Discovering Knowledge in Data: An Introduction to Data Mining. UK: John Wiley & Sons. 2004.

[3] N. Kiselyova, S. Iwata, V. Dudarev, I. Prokoshev, V. Khorbenko, V. Zemskov. Integration principles of Russian and Japanese databases on inorganic materials. Int. J. “Information Technologies and Knowledge”, 2(4), p. 366-372, 2008.

[4] N. N. Kiselyova, V. V. Podbel’skii, V. V. Ryazanov, A .V. Stolyarenko. Computer-aided design of new inorganic compounds with composition ABX_2 (X = S, Se or Te). Inorganic Materials: Applied Research, 1(1), p. 9-16, 2010.

[5] N. Kiselyova, A. Stolyarenko, V. Ryazanov, V. Podbel’skii. Information-analytical system for design of new inorganic compounds. Int.J. "Information Theories & Applications", 15(4), p. 345-350, 2008.

[6] N. N. Kiselyova, A. V. Stolyarenko, T. Gu, W. Lu. Computer-aided design of new wide bandgap semiconductors with chalcopyrite structure. Перспективные материалы. Спецвыпуск, с. 351-355, 2007.

[7] N. N. Kiselyova, A. V. Stolyarenko, V. V. Ryazanov, O. V. Senko, A. A. Dokukin, V. V. Podbel’skii. A system for computer-assisted design of inorganic compounds based on computer training. Pattern Recognition and Image Analysis, 21(1), p. 88-94, 2011.

[8] M. C. Ohmer, J. T. Goldstein, D. E. Zelmon, A. W. Saxler, S. M. Hegde, J. D. Wolf, P. G. Schunemann, T. M. Pollak. Infrared properties of $AgGaTe_2$, a nonlinear optical chalcopyrite semiconductor, J. Appl. Phys. 86(1), p.94-99, 1999.

[9] M. Purwins, A. Weber, P. Berwian, G. Műller, F. Hergert, S. Jost, R. Hock. Kinetics of the reactive crystallization of $CuInSe_2$ and $CuGaSe_2$ chalcopyrite films for solar cell applications. J. Crystal Growth, 287(2), p. 408-413, 2006.

[10] O. V. Senko. An Optimal Ensemble of Predictors in Convex Correcting Procedures. Pattern Recognition and Image Analysis, 19(3), p. 465-468, 2009.

[11] S. Siebentritt, U. Rau. Wide gap chalcopyrites. Heidelberg-Berlin: Springer. 2006.

[12] Y. Xu, M. Yamazaki, P. Villars. Inorganic Materials Database for Exploring the Nature of

Material. Jap. J. Appl. Phys., 50(11), p. 11RH02-1-11RH02-5, 2011.

[13] Y. Yang, H. Zou. A Coordinate Majorization Descent Algorithm for L_1 Penalized Learning. J. Statistical Computation & Simulation, 84(1), p. 84-95, 2014.

[14] G.-X. Yuan, C.-H. Ho, C.-J. Lin. An Improved GLMNET for L_1-regularized Logistic Regression. J. Machine Learning Research, 13(6), p. 1999-2030, 2012.

[15] В. С. Вавилов. Особенности физики широкозонных полупроводников и их практических применений. Успехи физ. наук, 164(3), с. 287–296, 1994.

[16] В. П. Гладун. Процессы формирования новых знаний. София: СД "Педагог 6". 1995.

[17] Ю. И. Журавлев, В. В. Рязанов, О. В. Сенько. «РАСПОЗНАВАНИЕ». Математические методы. Программная система. Практические применения. М.: ФАЗИС. 2006.

[18] Н. Н. Киселева. Компьютерное конструирование неорганических соединений. Использование баз данных и методов искусственного интеллекта. М.: Наука. 2005.

[19] Н. Н. Киселева. Прогнозирование существования AB_3X_3. Неорганические материалы, 45(10), с. 1157–1160, 2009.

[20] Н. Н. Киселева, В. А. Дударев, В. С. Земсков. Компьютерные информационные ресурсы неорганической химии и материаловедения. Успехи химии, 79(2), с. 162-188, 2010.

[21] Н. Н. Киселева, В. А. Дударев, М. А. Коржуев. База данных по ширине запрещенной зоны неорганических веществ и материалов. Материаловедение, 7, 2015.

[22] Н. Киселева, Д. Мурат, А. Столяренко, В. Дударев, В. Подбельский, В. Земсков. База данных по свойствам по свойствам тройных неорганических соединений «Фазы» в сети Интернет. Информационные ресурсы России, 4, с. 21-23, 2006.

[23] 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, 42(3), с. 380-384, 2004.

[24] А. В. Столяренко, Н. Н. Киселева, В. В. Подбельский. Система компьютерного конструирования неорганических соединений. Автоматизация и современные технологии, 9, с. 23-28, 2008.

[25] Ю. И. Христофоров, В. В. Хорбенко, Н. Н. Киселева, В. В. Подбельский, И. Н. Белокурова, В. С. Земсков. База данных по фазовым диаграммам полупроводниковых систем с доступом из Интернет. Изв. ВУЗов. Материалы электронной техники, 4, с. 50-55, 2001.

Information-Analytical System for Design of New Inorganic Compounds

N. N. Kiselyova, V. A. Dudarev, A. V. Stolyarenko

The principles of development of systems of knowledge discovery in virtually integrated distributed databases are considered. The methodology of integration of data mining programs based on different algorithms is developed. The proposed methods are applied to development of the information-analytical system for automation of process of new inorganic compounds computer-aided design based on use of pattern recognition programs for discovery of regularities in information of the databases on inorganic substances and materials properties. The examples of application of the developed system to design of new inorganic compounds are given.