

SOFIA INITIATIVE
„MINERAL DIVERSITY PRESERVATION”

X International Symposium
**MINERAL DIVERSITY
RESEARCH AND PRESERVATION**

СОФИЙСКАЯ ИНИЦИАТИВА
„СОХРАНЕНИЕ МИНЕРАЛЬНОГО РАЗНООБРАЗИЯ”

X Международнѳ симпозиум
**МИНЕРАЛЬНОЕ РАЗНООБРАЗИЕ
ИССЛЕДОВАНИЕ И СОХРАНЕНИЕ**

EARTH AND MAN NATIONAL MUSEUM
4, Cherny vruh Blvd., 1421 Sofia, Bulgaria
14-16 OCTOBER 2019

НАЦИОНАЛЬНЫЙ МУЗЕЙ „ЗЕМЛЯ И ЛЮДИ”
бул. „Черни връх” 4, София 1421, България
14-16 ОКТЯБРЬ 2019

БАЗА ДАННЫХ «WWW-MINCRYST» ПО КРИСТАЛЛОХИМИИ МИНЕРАЛОВ И СОПУТСТВУЮЩИЕ ИНФОРМАЦИОННО-РАСЧЕТНЫЕ ИНСТРУМЕНТЫ ДЛЯ МИНЕРАЛОГОВ И КРИСТАЛЛОГРАФОВ

ДМИТРИЙ ВАРЛАМОВ, ТАТЬЯНА ДОКИНА, НАТАЛЬЯ ДРОЖЖИНА,
ОЛЬГА САМОХВАЛОВА

*Институт экспериментальной минералогии имени академика Д.С. Коржинского
Российской Академии Наук, г. Черноголовка, РФ, dima@iem.ac.ru*

Аннотация. В статье рассматривается разработанная авторами общедоступная информационно-вычислительная система (ИВС) «WWW-Mincrust» по структурам и кристаллохимии минералов (более 10850 структур для 4470 минералов и их аналогов). ИВС предназначена в качестве эффективного инструмента минералогов и кристаллографов для работы с кристаллическими структурами минералов и их синтетических аналогов через Интернет ресурс. Описаны идеология и методы реализации системы, ее основные компоненты: база данных, комплексы интеллектуальных средств поиска и выбора информации, средства мультимедийного интерактивного представления информации, средства формирования перекрестных ссылок и использования вспомогательной информации.

Abstract. The article discusses a publicly available information-computing system (ICS) «WWW-Mincrust» developed by the authors on the structures and crystal chemistry of minerals (more than 10,850 structures for 4,470 minerals and their analogues). The system is intended as an effective tool for mineralogists and crystallographers to work with the crystal structures of minerals and their synthetic analogues through an Internet resource. The ideology and implementation methods of the system, its main components are described: a database, complexes of intelligent tools of searching and selecting information, tools of multimedia interactive presentation of information, means of forming cross-references and using auxiliary information.

ИВС «WWW-Mincrust», ее создание и развитие

В конце 1997 г. на основе возникших первых средств разработки интерактивных Интернет-ресурсов была создана база данных WWW-MINCRYST, ставшая одним из первых интерактивных мультимедийных Интернет-ресурсов в области наук о Земле в России и в мире (Chichagov et al., 2001; Чичагов и др., 2007). Инициатором работ был заведующий группой РСА ИЭМ РАН выдающийся кристаллограф А.В. Чичагов (увы, скончавшийся в 2010 году). Целью ресурса стало обеспечение интерактивного Интернет-доступа пользователям к большим массивам минералого-кристаллохимической информации. Стало ясно, что обеспечить доступ к большим, постоянно растущим объемам информации возможно только через стремительно формирующийся тогда Интернет, а чтобы снабдить исследователей максимумом возможности обработки информации – необходимо ввести в состав ресурса интерактивные инструменты по обработке информации. Фактологической основой WWW-MINCRYST послужили накопленные с 1985 года литературные данные по кристаллическим структурам минералов, сведенные в локальную базу данных (на основе литературных источников и авторских аналитических данных — всего около 3500 объектов на старте проекта). Далее, была создана ИВС WWW-MINCRYST (<http://mincryst.iem.ac.ru>) с различными инструментами по интерактивной обработке данных и их анализу в ее составе. В ходе проекта в состав ИВС вводились авторские программные разработки по обработке кристаллоструктурных и кристаллохимических данных, их анализу и мультимедийной визуализации. Основы идеологии и технологии ИВС описаны здесь (Чичагов и др., 2007; Варламов и др., 2013; Варламов и др., 2018а,б). ИВС WWW-MINCRYST изначально опиралась на использование Интернет-технологий и стала одним из пионерских интерактивных научных Интернет-ресурсов как в России, так и в мире (в области наук о Земле). Первый полностью работоспособный вариант WWW интерфейса был создан в рамках гранта РФФИ № 96-07-89162 и представлен пользователям в декабре 1997 года, т.е. ИВС функционирует, развивается и активно востребована пользователями почти 23 года.

Цель создания ИВС WWW-MINCRYST и входящих в ее состав интерактивных Интернет-инструментов – обеспечить возможность работы максимально широкого круга пользователей во всех областях науки, оперирующих с кристаллическим веществом

(геология, геохимия, кристаллография, физика твердого тела, физика поверхности, материаловедение и т.п.) с достоверной и актуальной информацией в области структурной кристаллохимии минералов, их синтетических аналогов и элементов, по возможности – развить методы обработки и визуализации кристаллоструктурных данных.

В ходе проекта ИВС WWW-MINCRYST нарастил свой информационный фонд почти в 3 раза (более чем на 7500 объектов) и по праву вошел в первые ряды рентгеноструктурных и кристаллохимических баз данных, связанных с изучением минерального вещества, как по оценкам пользователей, так и по мнению составителей отраслевых каталогов информационных ссылок.

Базовые компоненты ИВС и техническая реализация

Основным компонентом ИВС WWW-MINCRYST является собственно база данных или Информационный фонд (на сентябрь 2019 года — около 10850 информационных объектов для 4470 уникальных фаз, включая 4250 природных минералов и 200 синтетических фаз, не имеющих пока натуральных аналогов).

Информационный фонд содержит информацию о большинстве минеральных видов (около 4250 из 5562) официально признанных на январь 2020 года International Mineralogy Association ([http://cnmnc.main.jp/IMA_Master_List_\(2020-01\).pdf](http://cnmnc.main.jp/IMA_Master_List_(2020-01).pdf)), кристаллические структуры которых расшифрованы к настоящему времени. Помимо природных минеральных видов, в базе данных представлены синтетические минералы – их структурные аналоги, отличающиеся по составу (например, с заменой одного из катионов/анионов), и неорганические соединения (силикаты, фосфаты, бораты и др.), близкие по свойствам к природным веществам. Информационный фонд содержит данные структурных работ из более чем 140 иностранных и отечественных журналов за период от 1930-х годов по настоящее время. Ежегодный прирост новых или заново переопределенных/уточненных кристаллических структур минералов и их аналогов достаточно значителен, чтобы требовалась постоянная актуализация информационного фонда (в среднем, до 350-400 *новых* структур в год + 150-200 структур, подвергнутых ревизии и изменениям). В последние годы основной акцент был сделан на новые минералы, активно пополняемые в том числе российскими исследователями.

Пользовательский интерфейс ИВС WWW-МИНКРИСТ представляет собой многоуровневую информационную двуязычную (РУС/ENG языки интерфейса/компонентов и названия минералов) систему. В интерфейс входят: (а) комплексные поисковые интерфейсы, использующие как критерии поиска (в различных комбинациях) названия минералов, химический состав, кристаллоструктурные параметры, литературные ссылки и вспомогательную информацию, причем возможны комбинации поисковых признаков; (б) модули мультимедийного представления информации (интерактивные полиэдрические и шаровые структуры, линейчатые и непрерывные спектры — через Java-апплеты WWW-CrystPic и WWW-MixiPol; (в) классификационные схемы нескольких авторов [Годовиков, Бокий, Chirioti и т.д.]; (г) модули организации взаимосвязей с внешними информационными ресурсами через систему динамически формируемых ссылок; (д) WWW-ориентированный инструментальный разработчика (включая систему импорта, проверки и редакции данных, а также архивации и резервного копирования).

Технологически ИВС WWW-MINCRYST как web-сервис реализована на "классической" связке Linux-Apache-Mysql-PHP (LAMP) с использованием JavaScript и (для интерактивных апплетов) Java (на базе Sun/Oracle JDK). Размещена ИВС на сервере баз данных ИЭМ РАН, доступ не лимитирован.

Основным источником информации для ИВС служат журнальные статьи. Извлеченная кристаллоструктурная информация помещается (в соответствии со специальным форматом записи) в ASCII текстовый файл с последующей программной экспертизой (проводимой авторами-экспертами на локальных ПК) по результатам расчета межатомных расстояний и других кристаллоструктурных характеристик и, в случае положительного решения,

импортируется в ИВС специальными средствами через web-интерфейс как в виде единичных записей, так и пакетов записей (до нескольких сот) с входным контролем данных, что позволяет проводить постоянную актуализацию информационного фонда. Также реализована возможность редактирования записей в on-line режиме, их удаления, замены служебных файлов записей, архивации и восстановления базы данных через web-интерфейс.

В теоретической основе ИВС WWW-MINCRYST был положен ряд принципиальных на тот момент "идеологических" моментов. Синтез информации о кристаллической фазе, рассматриваемой как "монокристалл" и/или "поликристалл" с заменой экспериментальных поликристалл-стандартов расчетными – базовая идея WWW-MINCRYST, существенно расширяющая информационную базу сравнительно с чисто аналитической. На базе введенной информации программный пакет эксперта-разработчика позволяет автоматически сформировать вторую, производную от первой (аналитической), базу *расчетных* поликристалл-стандартов, проводя синтез двух типов информации о кристаллической фазе. Связка "Кристаллическая структура фазы и ее *расчетная* поликристалл-рентгенограмма" является информационной основой ИВС WWW-MINCRYST и служит важнейшим инструментарием в руках пользователя. Стержневая идея WWW-MINCRYST о синтезе двух типов кристаллоструктурной информации о кристаллической фазе реализует принципиально новый подход к формированию всей кристаллоструктурной информации о веществе и организации доступа к ней.

Информационная запись для индивидуального кристаллического вещества содержит информацию о названии (в соответствии с классификацией International Mineralogy Association или рекомендациями по наименованию неорганических веществ IUPAC), химическом составе, симметрии, параметрах элементарной ячейки, координатах атомных позиций с изотропными температурными факторами и заселенностями, информацию о межплоскостных расстояниях, HKL-индексах и интенсивностях сильнейших рефлексов рентгенодифракционной картины поликристалл-фазы, а также ссылки на соответствующие публикации по расшифровке или уточнению кристаллической структуры. Запись может быть специфицирована по полезным свойствам, особенностям химического состава и структуры, Р-Т условиям синтеза, принадлежности к условным минеральным группам. Каждая запись содержит "монокристалльные" и "поликристалльные" характеристики кристаллической фазы. Минералы классифицированы в соответствии с таксонами структурно-химических классификаций А.А. Годовикова, М. Чириотти, Г.Б. Бокия. Для 2500 фаз сделаны экспресс-оценки потенциальной энергии кристаллической решетки. В настоящее время создается система расчета содержаний элементов в соответствии с идеальной и реальной формулами минерала. В последние годы в записи вводились также географические привязки (при их наличии в статье) и прочие дополнительные сведения. Эти дополнительные сведения будут реализованы во вновь создаваемом варианте представления пользователю найденной информации.

В ИВС реализована система поиска фаз по отдельным наборам или комплексу критериев: название минерала (полное или частичное, по спецификации, в составе общеупотребительных групп – типа цеолитов, возможны алфавитные списки), химический или элементный состав в различных комбинациях (присутствие/отсутствие элементов, их комбинации, устойчивые химические группы катионов/анионов и силикатных групп, химические классы веществ), кристаллоструктурные характеристики (симметрия, пространственные группы, параметры элементарной ячейки, межплоскостные расстояния $d(hkl)$), что в сочетании с химическим (элементным) составом дает возможность прямого интерактивного качественного рентгенофазового анализа по результатам измерений. В систему поиска добавлен поиск по классификационным параметрам нескольких кристаллохимических и структурных классификаций (Годовиков, Чириотти, Бокий). Система поиска по этим параметрам обеспечивает поиск (и группировку) минералов по указанным нижним и промежуточным таксонам классификаций. Возможен поиск по литературным источникам (авторы, журналы, названия статей). По ряду параметров поиска ИВС WWW-

MINCRYST до сих пор не имеет аналогов в мире среди минералого-кристаллографических баз данных.

По результатам поиска пользователь получает ссылку на список фаз, отвечающих поисковым критериям, с указанием ID записи, названия фазы, формулы и пространственной группы с возможностью последующего перехода по гиперссылке. В случае единичного успешного ответа пользователь сразу автоматически переводится на найденный объект, если найдена группа объектов – предлагается список выбора.

В 2017-2019 годах в систему поиска по химическому составу сделаны важные дополнения для поиска по устойчивым сочетаниям элементов (типа анионов/катионов: SO_4 , CO_3 ; нейтральных групп – H_2O , NH_3 ; структурных силикатных мотивов – SiO_3 , Si_2O_7 , Si_3O_8 и т.п.), реализована возможность поиска по примесным изоморфным атомам, не отраженным в идеализированных формулах минералов (что важно для рассеянных элементов); улучшены схемы поиска по взаимным критериям «отсутствие-присутствие» элементов. Закончен переход на унифицированную кодировку utf-8 (для внутреннего представления текстовой информации в базе данных и выдачи пользователю), что упрощает работу с современными браузерами и позволяет использовать диакритические знаки в написании оригинальных имен минералов, что важно для европейских языков. В модуле WWW-Crystpic исправлен ряд неточностей и замечаний по работе с размножением сверхструктурных элементов, расширены возможности работы с большими (по количеству атомов, до 200, например, в сложных структурах цеолитах) структурами.

В «карте» найденного объекта представлены основные данные (как статические из БД, так и динамически рассчитываемые) – название, формула, параметры ячейки, различные кристаллографические данные (кол-во рефлексов, рентгеновские плотности, коэффициенты поглощения); CPDS карта – 20 максимальных рефлексов, их hkl позиции и интенсивности; базовые атомные позиции и их заселенности; полная информационно-расчетная карта (рентгеновские плотности, все рефлексы и т.п.); рассчитанные энергии решетки; принадлежность к различным классификационным схемам; источники данных; автоматически формируемые ссылки на внешние ресурсы и поисковые запросы.

Для всех записей через апплет WWW-Crystpic (пока – Java-3D с использованием библиотек OpenGL) доступны динамически создаваемые интерактивные изображения моделей кристаллических структур в шарах-сферах и в полиэдрических проекциях (до 138 позиций и до 1500(!) атомов на структуру). Программа позволяет делать всевозможные манипуляции с моделью структуры, включая масштабирование, непрерывное и/или автоматическое дискретное вращение вокруг "экранных" осей X,Y,Z; ориентацию по кристаллографическим осям, hkl-фрагментацию структуры (на hkl-ориентированные фрагменты толщиной $d(hkl)$), наращивание элементарных ячеек вдоль любых выбранных направлений для формирования "сверхструктур" и мотивов, а также прямой "ручной" и автоматизированный для малых полиэдров (тетраэдров и октаэдров) расчет любых межатомных расстояний и углов (плоских и телесных) в структуре. Программа изображает любые полиэдры, включая "дефектные" с необычно малыми ("плохими") межатомными расстояниями. Примеры моделей кристаллических структур (до 600 атомов) приведены на рис.1.

В настоящее время существуют проблемы с работой с внешними библиотеками OpenGL (на которых реализован апплет) и параноидальными настройками безопасности Java 8, которые решаются вмешательством пользователя. Однако из-за этого сейчас крайне необходим перенос апплета на новые программные платформы типа встроенных средств HTML5 или WebGL.

Модуль в виде апплета WWW-Mixipol предназначен для графического представления полных расчетных спектральных профилей поликристалл-рентгенограмм с возможностями манипулирования спектрами для разных источников излучения и разных типов спектральных шкал. Также модуль способен формировать рентгенограммы смесей фаз (до 6

фаз одновременно) при возможности варьирования относительными содержаниями компонентов смеси. Пример расчетных спектров смеси шести фаз приведен на рис.2.

Как для структур, так и для спектров минералов предусмотрены «упрощенные» варианты представления в виде интерактивных традиционных шаровых структур и линейчатых спектров.

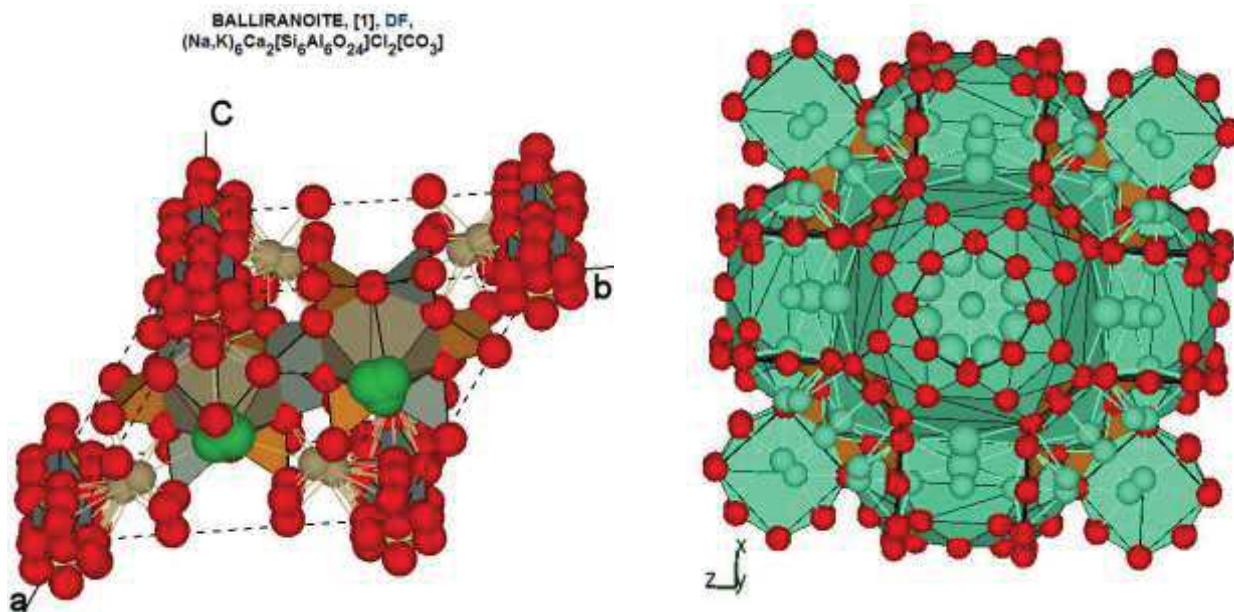


Рис. 1. Примеры интерактивных моделей кристаллической структуры (смешанное представление в шарах и полиэдрах)

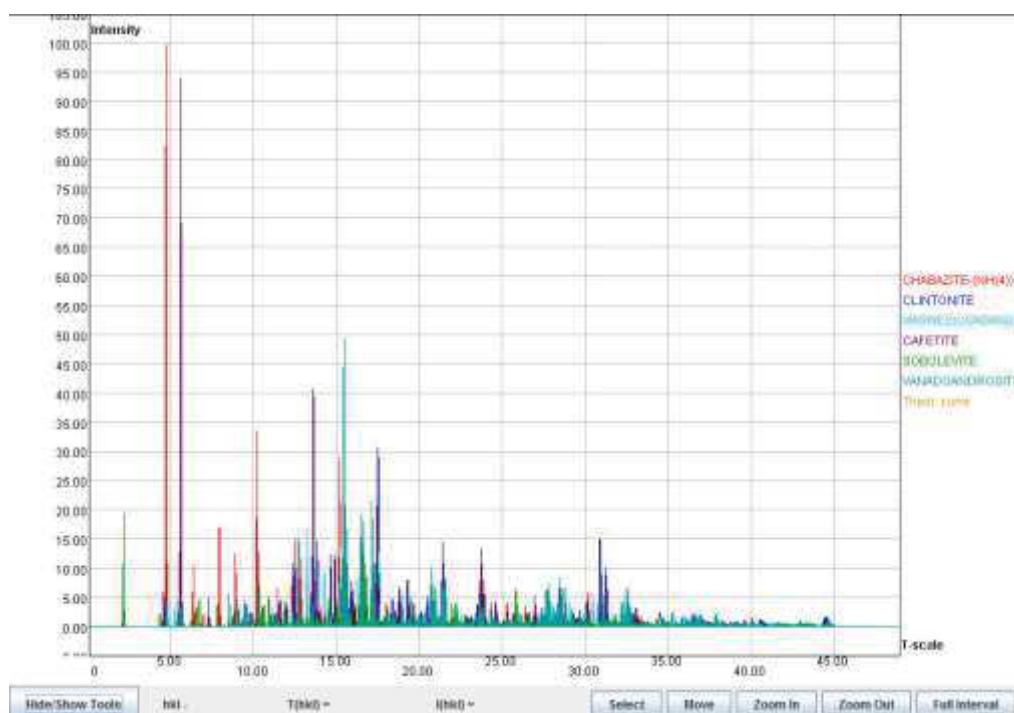


Рис. 2. Пример комбинации расчетных спектральных профилей поликристалл-рентгенограмм (смесь 6 фаз)

Одной из первых среди научных интернет-ориентированных баз данных для ИВС WWW-MINCRYST была разработана система динамически формируемых перекрестных веб-ссылок для связи объектов с записями для конкретных минералов в ведущих минералогических базах данных в Интернете (типа Mindat, Webmineral и т.п.). Система генерации

динамических гиперссылок на внешние информационные ресурсы (на минералогические базы данных и поисковые системы) позволяет "прозрачно" для пользователя подключать внешние массивы данных, используя метод "генеральных" запросов (Варламов и др., 2019). При этом пользователь сразу получает доступ к информации по интересующему его объекту, минуя стадии поиска или просмотра всей внешней базы. Кроме того, данный механизм реализует обратную связь, позволяя таким же образом ссылаться этим базам уже напрямую на наши информационные объекты, что резко повышает востребованность ИВС внешними пользователями.

Наличие в ИВС WWW-MINCRYST данных по 10850 кристаллическим структурам, встроенного расчетного комплекса и развитых средств визуализации позволили помимо возможностей, ориентированных на поиск и предоставление информации, использовать ИВС в разработке нетрадиционных научных подходов к интерпретации и представлению некоторых кристаллических структур. Обработка комплексными SQL запросами массива накопленных данных позволил получить весьма нетривиальные статистические выборки, например, по распределению в природе минералов среди групп симметрии кристаллов. WWW Xraypol позволил выявить в традиционных структурах возможность гибкого использования полиэдров и почти автоматически формировать различные варианты структурных моделей минералов. Благодаря гибкому использованию полиэдров, позволяющему формировать различные варианты структурных моделей минералов, для части минералов можно нестрого привязываться к традиционному катионно-анионному изображению, а формировать структуры на основе любых атомов, входящих в ее состав.

Востребованность WWW-MINCRYST хорошо подтверждается статистикой обращений (за 2019 год – 5 млн. успешных единичных запросов, более 217 Гб скачанной информации, почти 48000 уникальных сайтов-клиентов), а также большим количеством отзывов, описаний и внешних ссылок на WWW-MINCRYST (см. раздел "Ссылки" на сайте).

Заключение

Описанная ИВС «WWW-MINCRYST» является общедоступным, дружелюбным к пользователю интерфейсом к большому массиву кристаллохимической и минералогической информации с развитыми средствами поиска, представления и обработки и может служить мощным инструментарием для всех исследователей в минералогии, кристаллографии, физике твердого тела, материаловедении и прочих смежных областях науки.

Проведение работ по ИВС WWW-MINCRYST в течение 1997-2017 годов было поддержано 6-ю грантами РФФИ.

ЛИТЕРАТУРА

1. Chichagov A.V., Varlamov D.A., Dilanyan R.A., Dokina T.N., Drozhzhina N.A., Samokhvalova O.L., and Ushakovskaya T.V.. MINCRYST: a Crystallographic Database for Minerals, Local and Network (WWW) Versions // Crystallography Reports, 2001, v.46, 5, pp.876-879 DOI 10.1134/1.1405882
2. Варламов Д.А., Докина Т.Н., Дрожжина Н.А., Самохвалова О.Л. WWW-MINCRYST: Интернет-ориентированная информационно-вычислительная система по кристаллографии и кристаллохимии минералов // Вестник ЮУрГУ, Сер. «Вычислительная математика и информатика», 2013, т.2, вып.1, с.26-32 DOI:10.14529/cmse130103
3. Варламов Д.А., Докина Т.Н., Дрожжина Н.А., Самохвалова О.Л. Информационно-вычислительная система WWW-MINCRYST – общедоступный Web-инструментарий по кристаллическим структурам минералов и их аналогов // Научный сервис в сети Интернет: труды XX Всероссийской научной конференции (17-22 сентября 2018 г., г. Новороссийск). - М.: ИПМ им. М.В.Келдыша, 2018. с.111-121 DOI:10.20948/abrau-2018-49
4. Варламов Д.А., Подлесский К.К., Докина Т.Н., Дрожжина Н.А., Самохвалова О.Л. Мультимедийные Web-ориентированные базы данных в науках о Земле (кристаллохимия, минералогия, петрология) // Научный сервис в сети Интернет: труды XXI Всероссийской научной конференции (23-28 сентября 2019 г., г. Новороссийск). М.: ИПМ им. М.В.Келдыша, 2019. с. 198-209. DOI:10.20948/abrau-2019-89
5. Варламов Дмитрий, Докина Т., Дрожжина Н., Самохвалова О. Информационно-вычислительная система по кристаллохимии и минералогии «WWW-MINCRYST»: 20 лет развития // Сборник трудов IX Int.

Symposium “Mineral Diversity – Research and Preservation”, 16-18 October, 2017, Sofia, Bulgaria. Sofia, National Museum «Earth and Man», 2018, (250 c.), ISSN - 1313-9231, с.52-58

6. Чичагов А.В., Варламов Д.А., Ершов Е.В., Докина Т.Н., Дроzhzhina Н.А., Самохвалова О.Л. Кристаллографическая и кристаллохимическая база данных для минералов и их структурных аналогов (WWW-MINCRYST) // Записки РМО, 2007, т.136, № 3, с.135-141

REFERENCES

1. A.V. Chichagov, D.A. Varlamov, R.A. Dilanyan, T.N. Dokina, N.A. Drozhzhina, O.L. Samokhvalova, and T.V. Ushakovskaya. MINCRYST: a Crystallographic Database for Minerals, Local and Network (WWW) Versions // Crystallography Reports, 2001, v.46, 5, pp.876-879 DOI 10.1134/1.1405882

2. D.A. Varlamov, T.N. Dokina, N.A. Drozhzhina, O.L. Samokhvalova. WWW-MINCRYST: Internet-orientirovannaya informatsionno-vychislitel'naya sistema po kristallografii i kristallokhimii mineralov // Herald of South-Urals State University, Ser. «Vychislitel'naya matematika i informatika», 2013, t.2, vyp.1, s.26-32 DOI:10.14529/cmse130103

3. D.A. Varlamov, T.N. Dokina, N.A. Drozhzhina, O.L. Samokhvalova Informatsionno-vychislitel'naya sistema WWW-MINCRYST – obshchedostupnyi Web-instrumentarii po kristallicheskim strukturam mineralov i ikh analogov // Nauchnyi servis v seti Internet: trudy XX Vserossiiskoi nauchnoi konferentsii (17-22 sentyabrya 2018 g., Novorossiisk). - M.: IPM im. M.V.Keldysha p.111-121 DOI:10.20948/abrau-2018-49

4. D.A. Varlamov, K.K. Podlesskii, T.N. Dokina, N.A. Drozhzhina, O.L. Samokhvalova Multimedia Web-based databases in Earth sciences (crystal chemistry, mineralogy, petrology) // Nauchnyi servis v seti Internet: trudy XXI Vserossiiskoi nauchnoi konferentsii (23-28 sentyabrya 2019 g., Novorossiisk). M.: IPM im. M.V.Keldysha, 2019. p. 198-209. DOI:10.20948/abrau-2019-89

5. Varlamov Dmitry, Dokina T., Drozhzhina Natalya, Samokhvalova O. Information-calculation system on crystal chemistry and mineralogy «WWW-MINCRYST»: 20 years of development // Materials of IX Int. Symposium “Mineral Diversity – Research and Preservation”, 16-18 October, 2017, Sofia, Bulgaria. Sofia, National Museum «Earth and Man», 2018, (250 p.), ISSN - 1313-9231, pp.52-58

6. A.V. Chichagov, D.A. Varlamov, E.V. Ershov, T.N. Dokina, N.A. Drozhzhina, O.L. Samokhvalova. Kristallograficheskaya i kristallokhimicheskaya baza dannykh dlia mineralov i ikh strukturnykh analogov (WWW-MINCRYST) // Zapiski RMO, 2007, t.136, № 3, s.135-141