



ArDI – система обработки и анализа колебательных спектров минералов

Шендрик Р.Ю.¹, Плечов П.Ю.², Смирнов С.З.³

¹Институт геохимии им. А.П. Виноградова СО РАН, Иркутск

²Минералогический музей им. А.Е. Ферсмана РАН, Москва, pplechov@gmail.com

³Институт геологии и минералогии им. В.С. Соболева СО РАН, Новосибирск

ArDI (Advanced spectRa Deconvolution Instrument) – это веб-приложение для обработки и анализа колебательных спектров минералов (<https://ardi.fmm.ru/>). Приложение предназначено для быстрой и надежной идентификации минералов в геологических образцах. ArDI позволяет обрабатывать спектры, проводить поиск похожих спектров в базах данных и загружать эталонные спектры минералов в базу данных для ее расширения. ArDI имеет перспективы развития в нескольких направлениях, включая улучшение эргономики интерфейса и алгоритмов автоматической обработки спектров комбинационного рассеяния света (КРС), наполнение базы эталонных спектров и сопряжение базы данных эталонных спектров с информационной системой Минералогического музея им. А.Е. Ферсмана РАН и другими информационными системами, имеющими дело с минералами. Инструментарий ArDI может быть использован для быстрой диагностики минералов и интерпретации отдельных полос колебаний на спектрах. Он может быть полезен в минералогии, экспертизе сырья и ограненных драгоценных камней, а также в медицине, фармацевтике и криминалистике.

Ключевые слова: ArDI, система обработки спектров, рамановская спектроскопия, спектроскопия комбинационного рассеяния света (КРС), Минералогический музей им. А.Е. Ферсмана РАН, эталонные коллекции, новые минералы, минералогия.

Введение

Одной из важнейших характеристик минерала является его кристаллическая структура, отражением которой могут быть данные рентгеновской дифракции и/или данные колебательной спектроскопии. Среди методов колебательной спектроскопии наиболее популярными являются инфракрасная спектроскопия (ИКС) и спектроскопия комбинационного рассеяния света (КРС, она же рамановская спектроскопия). КРС очень активно развивается в последние годы благодаря появлению дешевых, но достаточно мощных светодиодных лазеров, а также чувствительных ССD-матриц. Это определило появление компактных приборов и позволило перевести КРС-спектроскопию в

ряд рутинных методов, который имеет ряд преимуществ перед другими распространенными методами исследования вещества (ИК-спектроскопия, рентгеновская дифрактометрия и т.д.). Одним из явных преимуществ КРС является то, что можно снять спектр в микроскопических выделениях, а многие редкие минералы встречены и описаны только в таком виде. Кроме этого, КРС-спектроскопия не требует специальной пробоподготовки и может применяться даже для анализа включений в минералах. Компактность современных источников лазерного излучения и детекторов позволяет создавать приборы, легко интегрируемые в настольные микроскопы и спектрометры, приборы оперативной

диагностики, беспилотные аппараты дистанционного зондирования и т.п. Несмотря на ряд преимуществ КРС-спектроскопии перед ИК (локальность, простота съемки, высокая чувствительность к некоторым внекаркасным комплексам и т.п.), КРС и ИК-спектроскопия являются взаимодополняющими методами. Часто моды, активные в ИК, неактивны в КРС, и наоборот. Более того, в настоящее время приборы для локального ИК-анализа с использованием специализированных микроскопов более доступны, чем КРС-микроскопы для схожих задач. Новые перспективы для использования локальной ИК-спектроскопии открывают и синхротроны нового поколения. В силу этого представляется эффективным разрабатывать универсальные программные продукты для обработки колебательных спектров без привязки их к конкретному методу.

Мы предполагаем, что в ближайшей перспективе методы колебательной спектроскопии станут основой стратегии геохимических и технологических исследований и разработок в самых разных отраслях деятельности человека как на Земле, так и при исследовании объектов Солнечной системы.

Для описания минерального разнообразия в XIX и XX веках создавались многочисленные справочники и картотеки. На сегодняшний день известно шесть тысяч минеральных видов, и каждый год добавляется около сотни новых. Бумажная основа не очень удобна для быстрого обновления, и, кроме того, она плохо подходит для спектральной информации. Переход с бумажных на электронные носители позволяет решить много проблем, в том числе способствует сохранению лесов на нашей планете, поэтому с конца XX века сообщество минералогов занимается созданием баз данных, которые с развитием информационных технологий становятся более комплексными. Целями таких баз данных и информационных систем является накопление знаний о минералах и помощь в их быстрой диагностике.

Базы данных собираются как отдельными исследователями и исследовательскими группами, так и компаниями, выпускающими спектроскопические приборы. Первые обычно доступны для членов этих групп, а вторые предоставляются с ПО, поставляемым с коммерческими моделями приборов. Проблемами этих двух категорий баз данных является их «локальность», ограниченность источников спектральной информации (обычно это одна лаборатория). Гораздо более выгодной формой является публичная база данных, доступ к которой на тех или иных условиях имеет широкий круг специалистов и заинтересованных в функционале

этого комплекса пользователей. Кроме того, большое количество спектров в эталонных спектральных библиотеках получено на старом оборудовании и требует серьезной ревизии. В силу этого определение минералов с помощью КРС-спектроскопии пока затруднено – требует очень высокой квалификации или часто приводит к ошибкам.

Одной из первых минералогических баз данных такого типа была база Калифорнийского технологического университета (<http://minerals.gps.caltech.edu/>), созданная и поддерживаемая стараниями профессора Дж. Россмана. На сегодняшний день в этой базе представлены данные оптической, инфракрасной, КРС- и мёссбауэровской спектроскопии минералов и разнообразные ссылки, связанные с этой тематикой.

Позднее данные по КРС-спектроскопии были трансформированы в базу данных RRUFF (<https://rruff.info/>) [Lafuente et al., 2015] – совместного проекта Калифорнийского технологического института и Государственного университета штата Аризона, где они были объединены с данными рентгеновской дифракции. База данных RRUFF приобрела популярность в минералогическом сообществе благодаря большому количеству данных и наличию инструмента для поиска необходимой информации по названию минералов, их химическому составу и библиографическим данным. Еще одной важной особенностью, стимулирующей популярность этого ресурса, было наличие специального приложения Crystal Sleuth, которое позволяло проводить расчеты по пользовательским данным рентгеновской дифракции и проводить сопоставление этих данных и данных КРС-спектроскопии с тем, что загружено в базу данных проекта. Кроме этого, приложение позволяло применять базовые операции обработки КРС-спектров: автоматическое вычитание базовой линии, удаление выбросов на спектрах, вырезание или выделение части спектра и поиск по нему. База данных RRUFF пополняется новыми данными и в настоящее время содержит спектральные данные о 2351 минерале. При этом обновление Crystal Sleuth было прекращено после 2009 года.

В 2017 году появилась еще одна публичная база данных, подобно проекту RRUFF, основанная на совмещении возможностей КРС-спектроскопии и рентгеновской дифракции для идентификации различных материалов, включая минеральные виды [ElMendili et al., 2019]. Она носит название Raman Open Database (ROD) (<https://solsa.crystallography.net/rod/>). База поддерживается командами из нескольких академических институтов и университетов Франции, Италии и Литвы и интегрирована с базой кристаллографических

данных COD (<http://www.crystallography.net>). Одним из направлений применения ROD является разработка системы, совмещающей ультразвуковое бурение и автоматизированный анализ минерального и химического состава, включающий в себя в том числе и КРС-спектроскопию керна. Такие системы особенно востребованы при изучении планет Солнечной системы с помощью беспилотных устройств. Сейчас эта база содержит 1133 записи. В ROD также включен поисковый инструмент, позволяющий производить поиск информации как по базе данных, так и по связанном с ней библиографическом ссылкой. Однако никаких способов автоматизированного сопоставления пользовательских данных с уже имеющимися загруженными данными нет, что не дает использовать ROD как удобное средство диагностики минералов. Несмотря на это, участие в проекте SOLSA (<https://solsa-dem-up.eu/en>) предполагает, что собираемая база данных должна входить в состав автоматизированного комплекса, который должен идентифицировать и минералы.

Обзоры-справочники по колебательной спектроскопии в виде серии книг периодически издаются Н.В. Чукановым с коллегами. В книгах [Chukanov, 2013; Chukanov, Chervonnyi, 2016; Chukanov, Vigasina, 2020] собрана максимально полная библиотека инфракрасных спектров минералов. Многие из этих спектров были получены на эталонах исследований новых минералов из коллекции Н.В. Чуканова и представляют собой огромную ценность. В книге [Chukanov, Vigasina, 2020], кроме новых данных по ИК-спектрам, приведены характеристики спектров комбинационного рассеяния (положение основных полос спектров) более 2000 минералов, существенно дополняющие базу данных RRUFF. К сожалению, открытых удобных и полных электронных баз данных инфракрасных спектров минералов и инструментов работы с ними в настоящее время практически не существует.

Существуют и другие проекты, которые способствуют накоплению огромных массивов эмпирических данных, имеющих большое значение не только для идентификации минеральных видов, но и для решения более сложных задач, связанных с определением кристаллических структур, особенностей состава минералов и ориентированных не только на геологические проблемы, но важных для области материаловедения, в том числе для создания современных высокотехнологичных устройств. Необходимо отметить проект WURM (<https://www.wurm.info/>), в рамках которого колебательные спектры для 325 минералов рассчитаны из первых принципов [Carcas, Bobocioiu, 2011].

В данной статье мы представляем новый

проект ArDI (Advanced spectRa Deconvolution Instrument), осуществляемый силами ИГХ СО РАН (Иркутск), Минералогическим музеем им. А.Е. Ферсмана РАН (Москва) и ИГМ СО РАН с Центральным Сибирским геологическим музеем (Новосибирск). Ядром проекта является приложение, позволяющее проводить обработку спектра, поиск аналогов загруженному спектру по доступным базам данных и опирающееся на собственную публичную базу данных, создаваемую участниками проекта.

Одна из поставленных целей – сделать доступным использование базы данных на различных типах устройств, в том числе и портативных, которые управляются различными операционными системами. Это как персональные компьютеры под управлением нескольких поколений Microsoft Windows, Linux, различных поколений BSD-систем, macOS и т.п., так и мобильные устройства, размеры и разрешение экрана которых позволяют работать с научными графиками. Таким образом, от современного приложения для поиска и обработки спектров требуется значительная универсальность, то есть возможность использования его на устройствах с различными операционными системами. Это сложно обеспечить в классической парадигме desktop-приложения даже в рамках семейства Microsoft Windows систем. Одним из возможных выходов является использование web-приложений, которые работают непосредственно в web-браузере.

Web-приложения имеют несколько значительных преимуществ перед обычными desktop-приложениями при решении научных задач. Во-первых, web-приложения позволяют работать с данными и запускать вычисления прямо в браузере, что значительно упрощает доступность и доступ к программам. Вместо необходимости устанавливать и обновлять программное обеспечение на каждом компьютере пользователи могут просто открыть ссылку на web-приложение и начать работу. Во-вторых, web-приложения поддерживают множество платформ, что означает, что они могут быть запущены на различных операционных системах, таких как Windows, macOS, BSD или Linux. Это обеспечивает гибкость и удобство в использовании. В-третьих, web-приложения могут легко обмениваться данными и результатами через Интернет, что позволяет ученым использовать базы данных, размещенных на разных серверах; удаленно работать с коллегами и делиться результатами исследований. Кроме того, web-приложения обычно имеют более простой и понятный пользовательский интерфейс, что упрощает работу с ними даже для пользователей без специальных навыков программирования или систем-

ного администрирования. В целом web-приложения предоставляют удобное и гибкое решение для научных задач, делая их доступными и удобными для всех пользователей. Так, web-приложения уже успешно работают и используются для организации научных исследований [Shevchenko, 2022].

Как говорилось выше, фактором, способствующим широкому использованию минералогических баз данных, является наличие приложения, которое позволяет проводить быстрый поиск по базе, предоставлять результаты в удобной форме, а также производить со спектральной информацией необходимые манипуляции: построение и вычитание базовой линии, определение положения пиков, деконволюцию спектров и т.п. Важно, чтобы это приложение могло работать, оперируя всеми доступными базами данных, и при этом не приводило к значительным затратам ресурсов компьютера пользователя. Некоторыми из этих возможностей обладает приложение Crystal Sleuth проекта RRUFF. Однако набор операций со спектром там сильно ограничен и, как уже говорилось выше, приложение не поддерживается с 2009 года. Еще одной особенностью этого приложения является то, что оно оперирует с копией базы данных, загружаемой на компьютер пользователя, а не обращается к оперативно обновляемой базе, расположенной на сервере проекта.

В данной статье описывается текущее состояние проекта ArDI и обсуждаются перспективы его развития.

Используемые методы

В качестве среды разработки был выбран язык Python 3.11 (<https://www.python.org/downloads/release/python-3110/>), для ее базисной внутренней части (далее backend) – Flask (<https://flask.palletsprojects.com/en/3.0.x/>), а для пользовательской презентационной части (далее frontend) – Dash (<https://dash.plotly.com/>). Для обработки кривых используются функции, встроенные в библиотеки numpy (<https://numpy.org/>) и scipy (<https://scipy.org/>). Разложение спектра на отдельные пики (деконволюция) производится с помощью пакета lmfit (<https://lmfit.github.io/lmfit-py/>). В настоящее время frontend Dash набирает все большую популярность благодаря удобству в представлении данных и инструментам работы с графиками. Язык Python содержит достаточно большое количество библиотек для работы с данными, является кросс-платформенным и обладает достаточно низким порогом вхождения, что в дальнейшем позволит сравнительно легко увеличивать число разработчиков. Критические по времени выполнения функции были переписаны на языке

C++ и интегрированы в качестве статической библиотеки в приложение. В частности, для вычисления базовой линии используется метод Asymmetric least square smoothing (ALS), реализованный по алгоритму, предложенному в работе [Eilers, 2005], с небольшими модификациями. Для хранения спектров используются базы данных в формате HDF5 (<https://www.hdfgroup.org/solutions/hdf5/>).

Приложение ArDI упаковано в Docker-контейнер (<https://www.docker.com/>), что позволяет легко масштабировать проект и обеспечивать его связность. Доступ к контейнеру осуществляется через обратный прокси NGINX (<https://www.nginx.com/>). Взаимодействие с обратным прокси обеспечивается через сервер web-приложений WSGI (Web Server Gateway Interface). Доступно несколько вариантов: пакеты UWSGI (<https://uwsgi-docs.readthedocs.io/>) или Gunicorn (<https://gunicorn.org/>).

Описание интерфейса

Web-приложение проекта ArDI (<https://ardi.fmm.ru>) решает несколько задач.

- 1). Обработка спектров с целью улучшения результатов поиска: сглаживание, кадрирование, вычитание базовой линии, удаление выбросов и нежелательных пиков и т.п.
- 2). Поиск близких спектров по существующим базам данных.
- 3). Загрузка эталонных спектров минералов в базу данных для ее расширения.

Структура web-приложения ArDI представлена на рис. 1. Вычислительная часть ArDI состоит из трех модулей.

Первый модуль, обозначенный на рис. 1 как Deconvolution, отвечает за первичную обработку спектра и его деконволюцию – разложение спектра на составляющие пики и вычисление базовой линии. Доступны следующие возможности при первичной обработке спектров: 1) сглаживание и уменьшение шума; 2) удаление «космических» пиков (cosmic rays); 3) выделение необходимой области на спектре и кадрирование спектра; 4) сдвиг спектра по оси x; 5) масштабирование спектра по оси y; 6) преобразование спектра в спектр поглощения при обработке спектров пропускания, что в дальнейшем может быть использовано при обработке ИК-спектров. Результаты всех действий со спектрами обратимы. Обработанный спектр возможно сохранить на локальном компьютере в формате .csv и загрузить заново уже обработанный спектр.

Важной функцией этого модуля является разложение загруженного спектра на компоненты и вычисление базовой линии (рис. 3). Вычисление базовой линии проводится по методу ALS [Eilers,

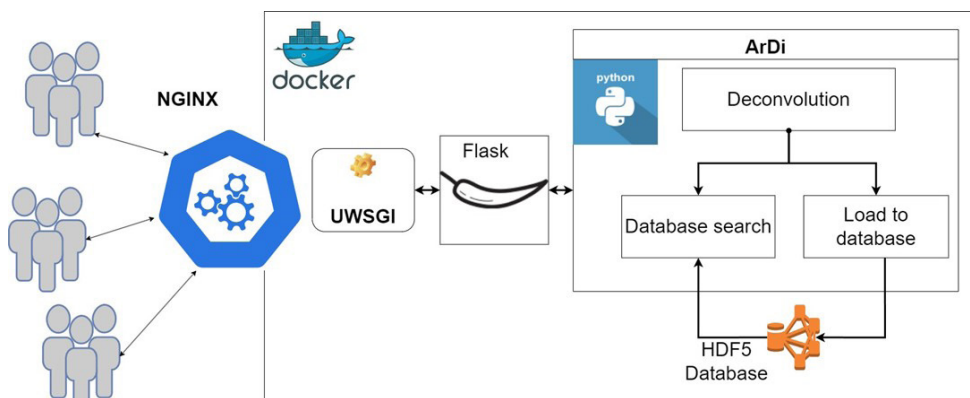


Рис. 1. Структура web-приложения ArDi.



Рис. 2. Автоматическое выделение пиков на КРС-спектре алмадина (обр. FMM_1_57724).

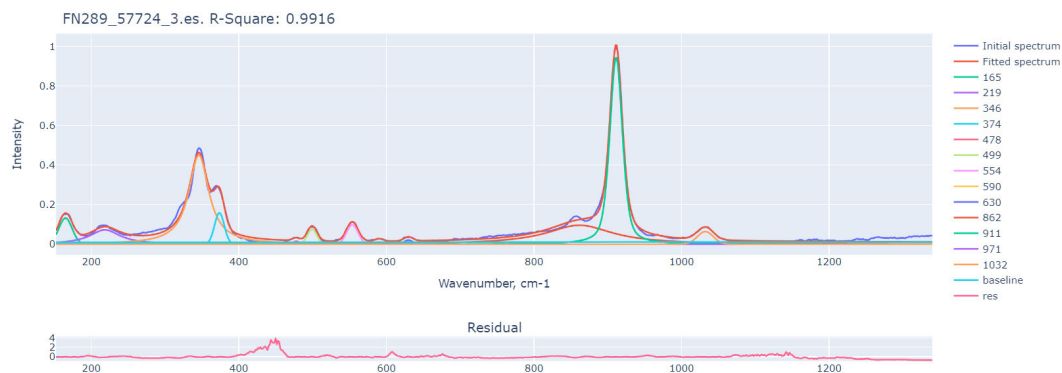


Рис. 3. Деконволюция КРС-спектра алмадина (обр. FMM_1_57724), выполненная в автоматическом режиме.

2005] с незначительными модификациями. В данном методе параметры базовой линии задаются всего двумя величинами: p - асимметрия и λ - гладкость, что является безусловным преимуществом при автоматическом подборе параметров. В алгоритм ALS было внесено небольшое изменение, связанное с тем, что параметры базовой линии можно задать для двух участков спектра. Например, если для области валентных колебаний O–H требуются низкие значения асимметрии и высокие значения гладкости, то для низкочастотной области, напротив, требуется уменьшение гладкости и увеличение коэффициента асимметрии. Подбор параметров деконволюции ведется методом наименьших квадратов. Имеется возможность выбора различных распределений, описывающих наблюдаемые в спектре пики. Начальные параметры деконволюции вводятся в виде редактируемой и экспортируемой таблицы.

Второй модуль (рис. 4, 5) используется для работы с базами данных. В ArDI используются базы данных в формате HDF5. HDF5 – это векторная

база данных. Каждая запись представляет собой многомерный массив, или вектор, в котором хранится спектр образца, в качестве параметров записи выступает название образца, параметры съемки, его географическая привязка, номер, химическая формула и т.п. Доступ к записи осуществляется по ключу, который включает в себя название образца, его уникальный номер и номер записи в базе данных. Формат HDF5 удобен для хранения больших объемов векторных данных и обладает достаточно быстрым поиском и доступом к записям по их ключу [Gosink, 2006].

При работе с базами имеется возможность загрузки всех записей, содержащихся в них, сортировка записей по имени, номеру, химическим элементам, входящим в образец, его классификации по Nickel–Strunz [Strunz, 2001], которая в настоящий момент является базовой у Комиссии по новым минералам, номенклатуре и классификации (КНМНК) MMA и поддерживается проектами Mindat.org (<https://www.mindat.org/strunz.php>) и Webmineral.com (<https://webmineral.com/help/StrunzClass.shtml>). Найденные спек-

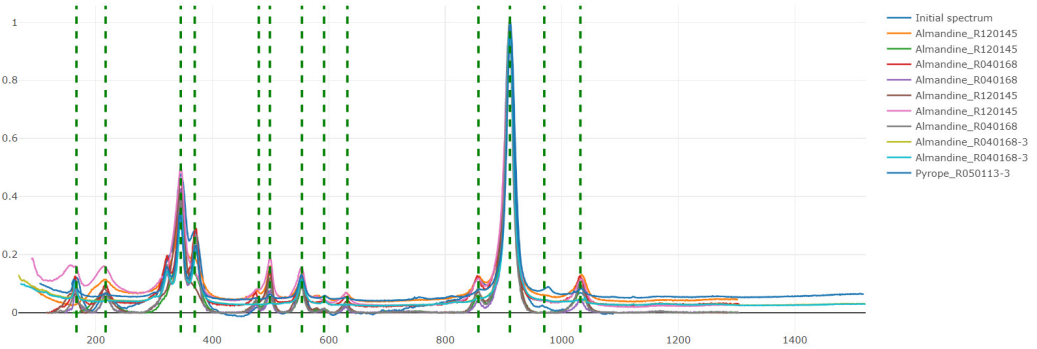


Рис. 4. Результат поиска похожих КРС-спектров для альмандина (обр. FMM_1_57724) среди спектров базы RRUFF.

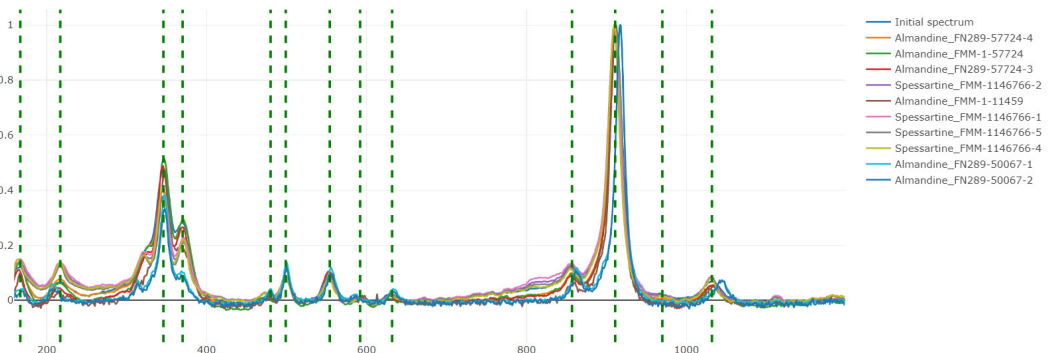


Рис. 5. Результат поиска похожих КРС-спектров для альмандина (обр. FMM_1_57724) среди спектров создаваемой базы ArDI.

тры могут быть нанесены на график сравнения (рис. 5), при этом существует возможность фильтрации и выбора спектров. Важной функцией модуля работы с данными является возможность поиска в базах данных спектров, похожих на загруженный пользователем. Поиск может осуществляться как по спектру целиком, так и по какой-то из его частей. В качестве меры схожести спектров используется евклидова метрика. Для этого происходит приведение спектров к одной размерности, далее вычисляется косинус угла между многомерными векторами, которые представляют собой значения по оси ординат у спектров. Похожесть спектров определяется как косинус угла между их значениями ординаты:

$$\text{Similarity} \equiv \cos \alpha = \frac{\sum_{i=1}^N y_i^{\text{exp}} y_i^{\text{base}}}{\sqrt{\sum_{i=1}^N (y_i^{\text{exp}})^2} \sqrt{\sum_{i=1}^N (y_i^{\text{base}})^2}} \quad (1),$$

где y^{exp} – спектр, который ищется, y^{base} – спектр из базы данных, N – число точек в спектре. Метрика схожести спектров выводится в результатах поиска, она лежит в интервале от 0 до 1, где 1 означает полностью идентичные спектры, а 0 – совершенно отличные друг от друга. В целом эта метрика схожа с коэффициентом корреляции Пирсона, однако использование евклидовой метрики более корректно с математической точки зрения [Tan, 2005].

При работе со спектрами КРС значительную сложность представляет появление артефактов. Методы интерпретации и обработки таких спектров достаточно подробно освещены в литературе [Chukanov, Chervonnyi, 2016]. Наиболее часто

встречающийся из них – это люминесценция образцов, интенсивность которой на несколько порядков выше интенсивности комбинационного рассеяния света. Возможности ArDI позволяют уменьшить вклад полос люминесценции и выделить непосредственно спектр КРС. В частности, ионы переходных металлов в основном обладают сравнительно широкой полосой в области 2000–5000 cm^{-1} при возбуждении в зеленой области спектра. Такие полосы эффективно удаляются при помощи процедуры удаления базовой линии. Ионы трехвалентных лантаноидов обладают сравнительно узкими полосами в широком спектральном диапазоне. Однако их положение известно и слабо зависит от матрицы. Таким образом, эти полосы можно удалить после проведения деконволюции, используя инструмент удаления «лишних» пиков.

Еще одной важной проблемой является исследование анизотропных образцов, когда спектры КРС или ИК-отражения анизотропного образца могут существенно отличаться друг от друга при изменении его ориентации. В таком случае метрики корреляции спектров Пирсона или Евклида могут давать малую степень схожести для одного и того же образца при изменении его пространственной ориентации. Возможным решением этой проблемы является использование кластеризации спектров с использованием методов машинного обучения. Мы планируем интегрировать такие возможности в ArDI.

Третий модуль (рис. 6) отвечает за загрузку спектров в базу данных. Помимо исходного спектра образца, спектра с вычтенной базовой линией, в отдельную базу заносится результат разложения спектра на пики. Также в файле json сохраняется информация о географическом

Рис. 6. Интерфейс системы ArDI для загрузки эталонного спектра в базу данных.

адресе образца, владельце, дате загрузки, параметрах съемки и произвольный комментарий. Изображение образца и его химический состав могут быть загружены отдельными файлами. В дальнейшем эти данные будут находиться на статической странице с описанием образца по типу существующих на сайте RRUFF. Название минерального вида выбирается как из списка минералов, утвержденных КНМНК ММА (перечень утвержденных минеральных видов регулярно обновляется), так и может быть введено вручную. Для каждого образца перед загрузкой необходимо указать его уникальный номер. Также имеется возможность указать специфические условия съемки и т.п. В дальнейшем для каждого образца планируется создание карточки с его составом и условиями съемки. В настоящее время такая информация хранится в базе данных, но не выводится на экран.

Обсуждение и перспективы развития

В настоящее время приложение ArDI является удобным инструментом для быстрой обработки, поиска похожих спектров и загрузки спектров в базу данных. Оно может быть использовано как для рутинной диагностики минералов, так и для интерпретации отдельных колебаний на спектрах известных и новых минералов. Планируется развитие проекта по нескольким направлениям:

1). Улучшение эргономики интерфейса и алгоритмов автоматической обработки КРС-спектров. В текущем виде ArDI за считанные секунды позволяет загрузить спектр, выделить на нем колебания и вычистить базовую линию на основе проведенной деконволюции. Иногда возникают проблемы с деконволюцией «шумных» спектров, а также со спектрами, в которых есть полосы как в низкочастотной, так и в высокочастотной областях. Улучшение алгоритмов автоматической обработки спектров уже позволило решить часть проблем, однако работа над дальнейшим их совершенствованием продолжается.

2). Наполнение базы эталонных спектров. Мы планируем провести ревизию спектров, накопленных проектами RRUFF и ROD. Эта ревизия будет заключаться в сохранении высококачественных спектров и удалении некачественных спектров. Таким образом, мы рассчитываем получить спектры примерно двух тысяч минералов. Коллекция Минералогического музея им. А.Е. Ферсмана РАН содержит образцы около четырех тысяч минеральных видов. В первую очередь мы будем снимать спектры тех минералов, которые отсутствуют в базах RRUFF и ROD, а также спектры минералов, качество которых в имеющихся базах вызывает критику. Таким образом, мы надеемся в ближайший год создать

наиболее полную базу высококачественных КРС-спектров для минералов. Наполнение базы спектров также будет проводиться на основе коллекции (около 1300 минеральных видов) Центрального Сибирского геологического музея (Новосибирск) и рабочих коллекций сотрудников ИГМ СО РАН и других научных организаций.

3). Мы ожидаем возникновения ряда методических проблем, связанных с получением спектров минералов, изменяющихся под воздействием лазерного пучка, сильно люминесцирующих минералов и минералов с высоким показателем отражения. Мы надеемся, что системный подход поможет решению части этих проблем. К примеру, нам уже удалось подобрать параметры съемки для некоторых сульфосолей, для которых ранее не удавалось получить КРС-спектры. Также существуют минералы, у которых нет активных колебаний в рутинно используемом диапазоне частот. Такие минералы, к сожалению, останутся за рамками текущей версии системы.

4). Сопряжение базы данных эталонных спектров с информационной системой Минералогического музея им. А.Е. Ферсмана РАН и другими информационными системами по минералам. Информационная система Музея содержит данные о 6006 минеральных видах и 157 210 образцах для 4202 минералов. На сегодняшний день это самая крупная информационная система по минералогии в русскоязычном сегменте Интернета. Эта система построена с учетом накопления разнородной информации о минералах и музейных образцах [Plechov et al., 2019], что позволяет сделать такое сопряжение достаточно просто. В дальнейшем возможна кооперация с другими минералогическими проектами мирового уровня (RRUFF, Mindat.org). Возможна кооперация с основными производителями оборудования для имплементации разработок в стандартное программное обеспечение, используемое при получении КРС-спектров.

5). В данный момент проект ArDI инициативно поддерживается тремя основными участниками. Это Минералогический музей им. А.Е. Ферсмана РАН, предоставивший вычислительные мощности, поддержку программного обеспечения и коллекцию минералов для пополнения эталонной базы спектров, ИГХ СО РАН (в лице Р.Ю. Шендрика), выполняющий основную работу по разработке и развитию интерфейса, а также ИГМ СО РАН, также предоставляющий коллекцию Центрального Сибирского геологического музея для наполнения эталонной базы спектров и участвующий в отладке эргономики и функционала интерфейсов. Таким образом, уже сложился

консорциум, где каждый участник вносит свой ощутимый вклад и пользуется результатами совместных усилий. Мы хотели бы расширить этот консорциум, пригласив в него все ведущие институты и лаборатории РФ, которые смогли бы принять деятельное участие в том или ином виде. В данный момент идет активное обсуждение условий вхождения в консорциум. Часть опций мы планируем открыть для широкого круга пользователей, но какая-то часть опций будет доступна только для участников консорциума или по подписке. Возможное коммерческое использование ArDI будет регулироваться участниками консорциума. Наши представления о будущей архитектуре консорциума предполагают какую-то количественную оценку вклада каждого участника. Современные технологии позволяют автоматизировать такую оценку, например, на базе смарт-контрактов [Zheng et al., 2020]. Однако использование для этого технологии блокчейна представляется нам избыточным для данной задачи. Мы проводим консультации с IT-специалистами, чтобы найти решение, которое будет отвечать задачам консорциума и соответствовать современному уровню развития информационных технологий.

6). Разработанные алгоритмы обработки могут применяться не только к КРС-спектрам, но и к многим видам спектральной информации. Мы проводим предварительные работы по обработке ИК-спектров, измеренных при различных режимах, дифрактограмм, оптических спектров поглощения и т.д. При расширении консорциума возможна работа по созданию эталонных баз различных типов спектров.

7). По мере накопления верифицированной информации в базе спектров планируется использовать методы машинного обучения для поиска и обработки. В частности, это откроет новые подходы к поиску минеральных видов и интерпретации данных колебательной спектроскопии. Ожидается, что такой подход позволит успешно работать со спектрами, содержащими пики от нескольких минеральных

видов, и перейти к фазовому анализу образцов на основе спектроскопических данных, подобно рентгенофазовому анализу. Также мы надеемся, что методы машинного обучения смогут помочь в работе со спектрами анизотропных минералов и в сравнении спектров, полученных на различном оборудовании и с различными условиями съемки.

Основное применение получившегося продукта – быстрое и надежное определение минералов в геологических образцах. Это крайне востребовано в самой минералогии (проверка найденных минеральных видов и поиск новых, пока неизвестных науке минералов). Кроме того, возможно широкое применение в экспертизе сырья и ограненных драгоценных камней, а также изделий из них. Эти объекты обладают особой материальной ценностью и требуют анализа неразрушающими методами (КРС-спектроскопия в большинстве случаев не разрушает образец). Применение ArDI позволит увеличить оперативность и эффективность экспертных работ.

Важной перспективой, на наш взгляд, является возможность интеграции ArDI в ПО современных приборов, использующихся при проведении научных исследований и разработке новых технологий. На основе метода возможно конструирование компактных устройств, полезных минералагам в полевых условиях и для минералогического картирования на объектах горнодобывающих предприятий. Кроме минералогии и геммологии, разработанный подход может быть широко использован в медицине, фармацевтике и криминалистике, так как КРС-спектроскопия уже максимально широко используется в материаловедении и для определения органических веществ.

Благодарности

Авторы благодарят Н.В. Чуканова и Е.А. Панкрушину за внимательное прочтение рукописи на стадии рецензирования и ценные конструктивные замечания, позволившие улучшить статью.

Список литературы:

Caracas R., Bobocioiu E. The WURM project – a freely available web-based repository of computed physical data for minerals // *American Mineralogist*. 2011. V. 96. P. 437–443.

Chukanov N.V. *Infrared spectra of mineral species: extended library*. Springer Science & Business Media, 2013. 1726 p.

Chukanov N.V., Chervonnyi A.D. *Infrared spectroscopy of minerals and related compounds*. Springer, 2016. 1109 p.

Chukanov N.V., Vigasina M.F. *Vibrational (Infrared and Raman) Spectra of Minerals and Related Compounds*. 2020. Springer, 1376 p.

Gosink L., Shalf J., Stockinger K., Wu K., Bethel W. HDF5-FastQuery: Accelerating Complex Queries on HDF Datasets using Fast Bitmap Indices, 18th International Conference on Scientific and Statistical Database Management (SSDBM'06), 2006, Vienna, Austria. P. 149-158. DOI: 10.1109/

SSDBM.2006.27.

Eilers P.H.C., Boelens H.F.M. Baseline correction with asymmetric least squares smoothing // Leiden University Medical Centre report. 2005. 1(1). P. 1–24.

El Mendili Y., Vaitkus A., Merkys A., Gražulis S., Chateigner D., Mathevet F., Gascoin S., Petit S., Bardeau J.-F., Zanatta M., Secchi M., Mariotto G., Kumar A., Cassetta M., Lutterotti L., Borovin E., Orberger B., Simon P., Hehlen B., & Le Guen M. Raman Open Database: first interconnected Raman–X-ray diffraction open-access resource for material identification // *Journal of Applied Crystallography*. 2019. V. 52(3). P. 618–625. DOI: 10.1107/s1600576719004229.

Lafuente B., Downs R.T., Yang H., Stone N. The power of databases: the RRUFF project. In: *Highlights in Mineralogical Crystallography*, T Armbruster and R M Danisi, eds. Berlin, Germany, W. De Gruyter, 2015. P. 1–30.

Plechov P.Y., Trousov S.V., Bychkov K.A., Konovalova K.A. Multilayered mineralogical information in spectroscopy

of minerals. In *XIX INTERNATIONAL MEETING ON CRYSTAL CHEMISTRY, X-RAY DIFFRACTION AND SPECTROSCOPY OF MINERALS*. 2019. P. 43.

Shevchenko Y. Open Lab: A web application for running and sharing online experiments // *Behavior Research Methods*. 2022. V. 54. P. 3118–3125. DOI: 10.3758/s13428-021-01776-2

Strunz H. and Nickel E.H. *Strunz Mineralogical Tables*, 9th Edition. Berlin and Stuttgart: E. Schweizerbart'sche Verlagsbuchhandlung, 2001. 870 p.

Tan P.-N., Steinbach M., Kumar V. *Introduction to Data Mining*, Addison-Wesley, ISBN 0-321-32136-7. 2005. Chapter 8. p. 500.

Zheng Z., Xie S., Dai H.N., Chen W., Chen X., Weng J., Imran M. An overview on smart contracts: Challenges, advances and platforms // *Future Generation Computer Systems*. 2020. 105. P. 475–491.