

## АВТОМАТИЗАЦИЯ ОБРАБОТКИ РЕЗУЛЬТАТОВ РЕНТГЕНОВСКОГО ДИФРАКЦИОННОГО АНАЛИЗА МАТЕРИАЛОВ

Шиманский Н.А.

Белорусский государственный университет,  
г. Минск, Республика Беларусь

Научный руководитель: Хорошко Л.С. – канд. физ.-мат. наук, доцент, ведущий научный сотрудник НИЛ Энергоэффективных материалов и технологий физического факультета

**Аннотация.** Описано разрабатываемое приложение для обеспечения анализа данных рентгеновской дифракции в автоматическом режиме. Преимуществом разрабатываемого решения является возможность использования компьютерного моделирования для сокращения временных затрат на обработку дифрактограмм, а также интеграция методов машинного обучения для повышения точности идентификации состава и свойств материалов.

**Ключевые слова:** рентгеновский дифракционный анализ, облачные данные, машинное обучение, автоматизация

**Введение.** Большое количество прикладных исследовательских задач в настоящее время решаются с использованием информационно-компьютерных технологий. Помимо уже ставших традиционными пользовательских программ и алгоритмов, позволяющих упростить извлечение текстовых данных, проверить условие соответствия и т.д., такие средства как нейронные сети позволяют оперировать более сложными задачами, как, например, анализ и предсказание свойств материалов. Так исследователи из Массачусетского технологического института в еще в 2018 г. предложили для прогнозирования свойств наноматериалов и анализа атомного уровня использовать обобщённые сверточные нейронные сети на кристаллическом графе (*Crystal Graph Convolutional Neural Networks*), обеспечивающие точность результата, сопоставимую с использованием теории функционала плотности (*Density Functional Theory*) [1]. Тем не менее, создание программных продуктов для качественной и быстрой обработки экспериментальных данных является неизменно востребованной. В рамках исследовательского учреждения также актуальной является возможность совместной работы исследователей на удаленном ресурсе, а при наличии исследовательского оборудования – последовательное накопление обработанных и верифицированных данных, в том числе для разработки авторских методик исследования материалов традиционными методами. Оптимизация процессов обработки экспериментальных данных и формирование единой исследовательской базы востребовано для материаловедческих исследований в рамках стратегий развития белорусской науки и концепции импортозамещения [2, 3].

В данной статье описано разрабатываемое приложение для автоматизации обработки результатов рентгеновского дифракционного анализа (РДА).

**Описание решаемой задачи.** Скорость и точность определения позиционирования и особенностей дифракционных пиков на получаемых дифрактограммах и их дальнейшего анализа в большей степени определяет эффективность оценки характеристик материалов с применением РДА. В традиционном представлении поиск и оценка параметров пика, таких как определение точности позиционирования, интенсивности, *FWHM* (англ. «full width at half maximum» – полная ширина на половине амплитуды) для определения размера областей когерентного рассеяния и др. [4], производятся оператором или исследователем преимущественно в «ручном» режиме, что значительно увеличивает время на оценку и интерпретацию результатов. Также исследователю необходимо специальное программное обеспечение, зачастую платное, позволяющее нивелировать и скорректировать такие факторы появления ошибок, как уровень шума на графике, базовая линия шума, разрывы данных и т.д. Еще одним фактором, замедляющим получение научного результата, может стать затруднение в обобщении результата и оперативном обмене данными между исследователями, работаю-

щими в рамках одного проекта, но аффилированных к разным учреждениям. Это может привести к повторяющимся действиям по обработке одних и тех же результатов, что снижает продуктивность научного коллектива в целом. Также поиск данных для верификации результатов может занимать достаточно продолжительное время, а в случае данных по рентгеновской дифракции еще и требовать приобретения лицензированного доступа к базам данных, которые в обоих случаях содержат большое количество наименований, зачастую, отличных от исследуемых и требующих трудоемкой проверки результатов поиска [5, 6].

Подход к решению описанных проблем является комплексным и должен сочетать в себе возможности оптимизации и ускорения процесса анализа дифрактограмм, сохранения результатов обработки и предоставления доступа к ним пользователям. Разработка описываемого проекта проводится с начала 2023 г. на кафедре физики твердого тела и нанотехнологий физического факультета БГУ. Целью разработки является предложение облачное решение в виде удобного пользовательского интерфейса для загрузки и анализа дифрактограмм. Учитывается также перспектива расширения функционала для возможности анализа данных, полученных с помощью других методов (спектроскопии комбинационного рассеяния света, фотолуминесценции и др.).

**Описание разрабатываемого приложения.** Разрабатываемое решение является веб-приложением (*SPA*), которое взаимодействует с платформой посредством *REST API*, что позволит обеспечить совместимость с большинством операционных систем. Облачный подход обеспечивается провайдером *AWS*, способным автоматически масштабировать нагрузку. Для хранения данных используется база данных *Elasticsearch (AWS OpenSearch)*, платформа реализуется в виде *serverless* функций (*AWS Lambda*). Работа приложения осуществляется с доступом к сети Интернет, для работы пользователи могут создавать учетные записи, загружать данные, создавать и публиковать графики. Интерфейс позволяет проводить автоматизированную обработку загруженных или находящихся в базе дифрактограмм: устранение шумов на графике, нахождение базовой линии шума, автоматический поиск позиционирования пиков, аппроксимация всех найденных или заданных пользователем пиков функциями Гаусса или Лоренца (и другими по выбору) с применением методов математического моделирования. По запросу пользователя приложение производит автоматический расчет параметров пиков, таких как амплитуда пика при скорректированной базовой линии, *FWHM*, смещение от эталона и др. Результаты обработки дифрактограммы отображаются на самом графике в виде комментариев (наложений). Также предусмотрена возможность добавления к дифрактограммам пользовательских пометок, комментариев, и их демонстрация другим пользователям. Обработанные дифрактограммы можно сохранить в облачной базе данных и поделиться с пользователями или избранными группами в режиме реального времени. Особенностью приложения является возможность сопровождения каждого идентифицированного пика такими метаданными, как химический состав, фаза вещества, концентрация, и т.д., что позволяет использовать для повышения точности и сокращения времени интерпретации машинное обучение. Обработанные дифрактограммы, имеющие полное и подтвержденное описание идентифицированных пиков и отмеченные как эталонные, используются в специальном модуле проекта (*Python*, фреймворк *Tensorflow AI*) для обучения нейронной сети, которая впоследствии сможет предсказывать состав исследуемого материала с большой достоверностью, исключая ручной поиск и использование сторонних приложений, на основании накопленных данных и проведенных циклов обучения нейронной сети. Достоверность прогнозов будет повышаться по мере работы с приложением и накопления эталонов в базе данных, что позволит, в частности, создать уникальные библиотеки по исследуемой номенклатуре химических соединений. Первичная оценка затраченного времени для исследования типовой дифрактограммы на примере выполнения практической работы в учебной группе показывает, что минимальное время, затрачиваемое для полного анализа дифрактограммы noticeably в ручном режиме, составляет около 2 ч, с использованием демоверсии лицензируемого платного пакета *Match!* – 22 мин, с использованием разрабатываемого приложения –

12–18 с, при этом в автоматическом режиме будет произведено устранение шумов, коррективировка базовой линии. Помимо очевидного выигрыша во временных затратах, использование обученной сети в разрабатываемом приложении также позволит освободить время высококвалифицированного персонала для решения иных задач, что может положительно сказаться на эффективности работы научных групп.

**Заключение.** Рассмотренная архитектура и методология позволяет разрабатываемому приложению обеспечить оптимизацию решения прикладных задач в области исследования свойств материалов. При этом применяемое для обработки результатов РДА приложение в последствии может быть адаптировано для обработки результатов, полученных с помощью ряда других методик, таких как комбинационное рассеяние света, люминесцентная спектроскопия и др. Хотя достоверность автоматизированного анализа с использованием нейросетей все еще является дискуссионным предметом в научном сообществе, тем не менее, можно отметить непрерывное повышение точности такой обработки по мере совершенствования технологий в области машинного обучения и больших данных. Необходимо также отметить, что используемые вычислительные методологии не налагают особенных ограничений на источник и методики анализа материалов, которые могут быть автоматически обработаны, а применяемый подход машинного обучения позволяет практически мгновенно получать предположения о составе исследуемого материала с помощью библиотеки накопленных пользовательских данных и обучаемой нейронной сети что подтверждает актуальность разрабатываемого продукта и перспективность проводимых исследований.

### Список литературы

1. Xie, T. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties / T. Xie, J. Grossman // *Physical Review Letters*. – 2018. – Vol. 120. – P. 145301-1–145301-6.
2. Стратегия «Наука и технологии: 2018–2040» [Электронный ресурс]. URL [https://nasb.gov.by/congress2/strategy\\_2018-2040.pdf](https://nasb.gov.by/congress2/strategy_2018-2040.pdf) (Дата обращения: 22.03.2023).
3. Перечень импортзамещающей продукции – Министерство экономики Республики Беларусь [Электронный ресурс]. URL: <https://econotny.gov.by/ru/importoz-ru/> (Дата обращения: 22.03.2023).
4. Гусев, А.И. Наноматериалы, наноструктуры, нанотехнологии / А.И. Гусев. – М. : ФИЗМАТЛИТ, 2005. – 416 с.
5. Crystallography Open Database [Электронный ресурс]. URL: <http://www.crystallography.net/cod/> (Дата обращения: 22.03.2023).
6. Cambridge Structural Database System [Электронный ресурс]. URL: <https://software.chemucla.edu/CSD/> (Дата обращения: 23.03.2023).

UDC 004.822-501

## AUTOMATION OF PROCESSING OF THE RESULTS OF X-RAY DIFFRACTION ANALYSIS OF MATERIALS

*Shymanski N.A.*

*Belarusian State University, Minsk, Republic of Belarus*

*Khoroshko L.S. – PhD, associate professor, lead researcher in R&D Lab of Faculty of Physics*

**Annotation.** The application being developed to provide analysis of X-ray diffraction data in automatic mode were described. The advantage of the developed solution is the possibility of using computer simulation to reduce the time spent on processing diffraction patterns, as well as the integration of machine learning methods to improve the accuracy of identifying the composition and properties of materials.

**Keywords:** X-ray diffraction analysis, cloud data, machine learning, automation