

Представление кристаллохимической информации структур слоистых гибридных галогенидов свинца в машиночитаемом виде

Научный руководитель – Марченко Екатерина Игоревна

Бучинский Владимир Витальевич

Студент (магистр)

Московский государственный университет имени М.В.Ломоносова, Геологический факультет, Кафедра кристаллографии и кристаллохимии, Москва, Россия

E-mail: buchinskiy.vova@gmail.com

На сегодняшний день полупроводниковые материалы на основе слоистых гибридных галогенидов свинца являются перспективными в области оптоэлектроники. Кристаллические структуры этого обширного класса соединений, насчитывающего более тысячи структур [1], являются производными от структурного типа перовскита, то есть содержат слои из октаэдров PbX_6 ($X=I, Cl, Br$), соединенных по вершинам. Органические катионы, локализованные между слоями октаэдров, диктуют искажение неорганической подструктуры, что усложняет рациональный кристаллохимический дизайн таких материалов и влияет на физические свойства этих соединений [2]. Для успешного построения моделей по предсказанию корреляций состав-структура-свойства в исследуемом классе соединений с применением алгоритмов искусственного интеллекта необходимо создать рациональный алгоритм по автоматической обработке кристаллохимической информации об исследуемых соединениях в машиночитаемый вид, чему и посвящена настоящая работа.

В работе приведен обзор методов представления кристаллографической и кристаллохимической информации в машиночитаемых видах, реализованных на алгоритмическом языке Python. По результатам литературного обзора выбраны наиболее подходящие методы представления, и с использованием стандартных библиотек языка Python написан собственный алгоритм по автоматическому представлению стандартных CIF (crystallographic information file) файлов структур, содержащих кристаллографическую информацию. Для обработки кристаллографической информации использовалась база данных слоистых гибридных галогенидов свинца NMSE [3].

Программа селективно выбирает органические катионы из CIF-файлов и записывает их в машиночитаемом формате SMILES. Неорганическая подструктура в результате работы алгоритма записывается в виде уникального топологического индекса подрешетки. Такое представление CIF-файлов можно записывать в виде многомерного вектора, подходящего для применения в качестве входной информации об исследуемых кристаллических структурах для современных алгоритмов с применением искусственного интеллекта

Источники и литература

- 1) Marchenko E.I., Fateev S.A., Petrov A.A. et al. // Chem. Mater. 2020. V. 32. № 17. P. 7383. <https://doi.org/10.1021/acs.chemmater.0c02290>
- 2) Marchenko E.I., Korolev V. V., Fateev S.A. et al. // Chem. Mater. 2021. V. 33. № 18. P. 7518. <https://doi.org/10.1021/acs.chemmater.1c02467>
- 3) NMSE 2D Perovskite Database: <http://pdb.nmse-lab.ru/>