

Неструктурированные теплофизические данные и ИИ

Еркимбаев А.О.^{1,®}, Зицерман В.Ю.¹ и Кобзев Г.А.¹

¹ Объединенный институт высоких температур Российской академии наук, Ижорская ул., 13, корп. 2, Москва, 125412, Россия

® adilbek@jiht.ru

К настоящему времени по теплофизическим свойствам веществ и материалов накоплены обширные массивы плохо структурированных или неструктурированных данных [1, 2]. Как правило, они представляют собой различные текстовые и графические объекты. Все эти объекты обычно хранятся в "исходном" формате в стандартных базах данных (БД) как дополнительные архивные ресурсы с возможностью их выгрузки и отдельного изучения [2]. Однако, работа с ними (поиск, систематизация и т. п.) типовыми инструментами БД, практически нереализуема. Здесь предложено использовать для организации полноценной работы с неструктурированными данными методы Искусственного Интеллекта (ИИ) с применением больших языковых моделей (LLM) [3, 4] для выполнения запросов к БД на естественном языке (ЕЯ). Эти технологии снимают рутину программирования запросов на языках реляционных и NO SQL БД, активным привлечением возможностей ИИ. На основе появившихся новых возможностей был разработан и реализован проект организации запросов на ЕЯ с локальным использованием LLM технологий для поиска и анализа в массивах неструктурированных данных, дополняющих БД ТЕРМАЛЬ и другие теплофизические БД ОИВТ РАН.

- [1] Erkimbaev A O, Zitserman V Y and Kobzev G A 2023 *Automatic Documentation and Mathematical Linguistics* **57**(3) 178–192
- [2] Belov G V, Erkimbaev A O, Zitserman V Y, Kobzev G A and Morozov I V 2020 *High Temperature* **58**(4) 566–582
- [3] Erkimbaev A O, Zitserman V Y and Kobzev G A 2026 *Automatic Documentation and Mathematical Linguistics* **60**(1) S26–S36
- [4] Manning C D 2022 *Daedalus* **151**(2) 127–138