**УДК 004.6**

**ПРОБЛЕМАТИКА ОПИСАНИЯ И АНАЛИЗА НАБОРОВ ДАННЫХ ДЛЯ ИНФОРМАЦИОННОГО ОБЕСПЕЧЕНИЯ РЕСУРСОЕМКИХ ПРИЛОЖЕНИЙ**

*Сеченых Полина Алексеевна, младший научный сотрудник1,*

[*p-sechenyh@mail.ru*](mailto:p-sechenyh@mail.ru)

*1 ФИЦ ИУ РАН, г. Москва*

**Аннотация:** В работе представлен подход к подготовке наборов данных, используемых для информационного обеспечения ресурсоемких приложений.

**Ключевые слова:** семантические требования, анализ данных, структуризация информационного контента, спецификация данных.

Результаты компьютерного моделирования в области материаловедения электронных компонентов существенно зависят от качества используемых данных, поэтому при их проведении требуется уделять внимание не только поиску и накоплению информации, но и ее детальному описанию и аналитическому исследованию по ключевым требованиям контроля полноты, достоверности, применимости и адекватности данных.

Можно сформулировать три ключевых фактора проблематики построения объектов данных, характерных для рассматриваемой предметной области:

* большой объем элементов данных и их сборок (агрегатов), обусловленный тем, что математическое и компьютерное моделирование базируется на молекулярных или атомарных расчетных уровнях с последующим масштабированием их в макро-структуры [1];
* большое количество прецизионных расчетов и чувствительных к входным параметрам алгоритмов;
* множество разнообразных источников данных, требующих проверки и согласования.

Необходимость хранения, анализа, сопоставления, поиска, агрегации и фильтрации значительных объемов разнообразных данных, требуемых для работы различных материаловедческих приложений, делает актуальным не только их сбор, но и подробное описание, включающее декомпозицию данных на множества по разнообразным критериям и требованиям, маркировку множеств по их назначению и степени достоверности, а также сопоставления им соответствующих ограничений и операций обработки.

Для проведения глубинной спецификации данных предлагается методология, включающая три базовые стадии. На первой формализуются количественные и семантические требования и определяются рабочие наборы данных, которые в зависимости от источника, могут быть классифицированы как данные из справочников, баз данных, инструментальных средств предметной области или иных экспериментально-расчетных источников информационного контента в их исходном виде. Они сохраняются для дальнейшего исследования в специализированной области. На второй стадии осуществляется анализ и выполняется иерархическая структуризация накопленных данных. Это делает возможным уточнение и сравнение данных, хранение различных наборов, и применение множества инструментов их обработки. На третьей стадии данные агрегируются и отображаются в объекты предметной области в соответствии со схемой базы данных системы информационной поддержки приложений. Перечисленные этапы предложенной методологии позволяют использовать специализированную область исследования информационного контента для построения объектных моделей сценариев приложений, структуризации схемы доменного представления данных и формирования событийно-информационных форм.

Апробация данной методологии на примере описания данных для задачи расчета метрических параметров кристаллических структур [2, 3] позволила уточнить требования, критерии и фильтры поиска информации, описания данных, каталогизацию параметров и их верификацию. Удалось также выявить взаимосвязь наборов подготовленных данных, применить фильтры и правила ограничений целостности, провести каталогизацию входных параметров и их проверку по критериям полноты и достоверности, реализовать первичное агрегирование и построение объектов доменной модели приложений.

Планируется программная реализация данного подхода, которая предполагает использование стека Web-ориентированных инструментальных средств, обеспечивающих локальный и удаленный доступ к контенту требований и каталогам решений с разделением полномочий для различных категорий исследователей и пользователей.

**Список использованных источников**

1. Абгарян К.К., Многомасштабное моделирование в задачах структурного материаловедения. – М.: МАКС Пресс, 2017. – 284 с.
2. Сеченых П.А. Математическое моделирование метрических параметров ГПУ металлов. Известия высших учебных заведений. Материалы электронной техники. 2022;25(4):283-287. https://doi.org/10.17073/1609-3577-2022-4-283-287
3. Сеченых П.А. Моделирование кристаллической структуры шпинелей с применением реляционной схемы хранения данных // Математическое моделирование в материаловедении электронных компонентов. МММЭК-2023 : Материалы V Международной конференции, Москва, 23–25 октября 2023 года. – Москва: ООО "МАКС Пресс", 2023. – С. 96-101. – DOI 10.29003/m3597.MMMSEC-2023/98-101

**PROBLEMS OF DESCRIPTION AND ANALYSIS OF DATA SETS FOR INFORMATION SUPPORT OF RESOURCE-INTENSIVE APPLICATIONS**

Sechenykh P.A.

**Abstract:** The work presents an approach to preparing data sets used for information support of resource-intensive applications.

**Keywords:** semantic requirements, data analysis, structuring of information content, data specification.