

УДК 004.65

## ПРОЕКТИРОВАНИЕ ГРАФОВОЙ ХИМИЧЕСКОЙ БАЗЫ ДАННЫХ И РАЗРАБОТКА АЛГОРИТМОВ ПОИСКА

Литвак И.Г. (Университет ИТМО)

Научный руководитель – Пантюхин И.С.

(Университет ИТМО)

**Введение.** На данный момент существует несколько десятков баз данных химических соединений и реакций с публично доступным пользовательским интерфейсом. Эти сервисы позволяют получать подробную информацию о свойствах отдельных химических соединений или реакций, но их использование для практико-ориентированных и научных задач, требующих сложного анализа, затруднено из-за ограниченных возможностей используемых поисковых систем. Использование графовой модели базы данных позволяет представить в ее структуре большое количество связей между объектами [1], что в свою очередь позволяет разработать более сложные алгоритмы поиска и анализа данных.

**Основная часть.** Структура предложенной базы данных состоит из двух типов вершин графа: химического соединения и химической реакции. Ребра графа направленные, имеют множество типов, разделенных на две категории:

1) Связь между химическим соединением и реакцией (реагент, продукт, катализатор).

2) Связь между двумя химическими соединениями (стереоизомер).

Для описания химической структуры соединения используется формат SMILES [2], позволяющий уникальным и единственным образом представить ее в формате текстовой строки. В связи с этим свойством строка SMILES также используется как уникальный идентификатор для химического соединения.

Для демонстрации возможностей графовой базы данных реализованы три типа поиска:

1) Поиск по прямому соответствию идентификаторов химического соединения (SMILES, InChI) с использованием простого индекса вершин.

2) Полнотекстовый поиск по названиям или описаниям химических веществ с использованием полнотекстового индекса вершин.

3) Обход графа для рекурсивного поиска химических реакций по заданным параметрам с началом в вершине, соответствующей заданному соединению.

**Выводы.** Спроектирована оптимизированная для поисковых алгоритмов модель графовой базы данных, описывающая параметры вершин графа и структуру связей между ними. Разработаны прототипы алгоритмов поиска для различных практических сценариев использования базы данных пользователями.

### Список использованных источников:

1. Засядко Г. Е., Карпов А. В. Проблемы разработки графовых баз данных // Инженерный вестник Дона. – 2017. – Т. 44. – №. 1 (44). – С. 24.

2. Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules // Journal of chemical information and computer sciences. – 1988. – Т. 28. – №. 1. – С. 31-36.