

УДК 004.4

DOI: 10.15827/0236-235X.124.808-813

Дата подачи статьи: 15.06.18

2018. Т. 31. № 4. С. 808–813

## Информационная система обработки и хранения данных о кинетике химических реакций

Н.В. Звягинцев<sup>1</sup>, аспирант, n.zvyagintsev@gmail.com

В.А. Биллиг<sup>1</sup>, к.т.н., доцент, старший научный сотрудник, профессор

Vladimir-Billig@yandex.ru

<sup>1</sup> Тверской государственный технический университет, кафедра программного обеспечения,  
г. Тверь, 170026, Россия

В настоящее время активно развивается хемоинформатика, позволяющая прогнозировать многие свойства веществ, в том числе их реакционную способность, биологическую активность и т.д. В частности, исследуются зависимости «структура–свойство», позволяющие на основе данных о структуре прогнозировать различные свойства.

Построение подобных моделей требует использования представительных выборок данных и интерфейсов для работы с ними, поэтому создание специализированных БД является важной задачей хемоинформатики. В настоящее время создан ряд БД с кинетическими параметрами реакций, содержащих сведения о кинетике химических реакций (константы скорости реакций, энергии активации и т.д.). Однако большинство таких БД с кинетическими параметрами не предоставляют специализированные сервисы (API) для программного анализа данных.

В статье рассматривается информационная система LacertaLab, содержащая БД и специализированные сервисы для обработки и анализа информации о протекании химической реакции. В БД хранятся как кинетические параметры реакций, так и данные о механизме химических реакций. Информационная система позволяет осуществлять поиск взаимосвязей между параметрами реакций, используя метод наименьших квадратов. Программные сервисы системы (API) предоставляют средства доступа к информации из БД, с их помощью формируются модели «структура–свойство». Данная ИС имеет возможность взаимодействовать с внешними сервисами, получать от них кинетические параметры и сохранять в собственной БД.

**Ключевые слова:** химическая кинетика, моделирование химических реакций, БД химических соединений и реакций, хемоинформатика.

Моделирование химических реакций является для исследователей важной задачей. На основе результатов моделирования можно выбирать оптимальный режим и условия проведения реакции, влияние условий на ее протекание и рассчитывать выходы продуктов реакции. Важнейшей характеристикой химической реакции является скорость.

С формальной точки зрения химическая реакция записывается в виде  $A + B + \dots \rightarrow C + D + \dots$

Скорость реакции  $w$  в общем виде описывается уравнением

$$w = k_1 C_A^n C_B^m = k_2 C_C^p C_D^q, \quad (1)$$

где  $k_1$  и  $k_2$  – константы скорости реакции;  $C_A$ ,  $C_B$ ,  $C_C$ ,  $C_D$  – концентрация веществ, принимающих участие в реакции;  $n$ ,  $m$ ,  $p$ ,  $q$  – коэффициенты, влияющие на скорость реакции.

Многие реакции являются обратимыми, то есть их продукты могут взаимодействовать между собой с образованием исходных веществ. В таком случае вводится понятие константы равновесия реакции  $K$ , которая выражается через константы скоростей прямой и обратной реакций:

$$K = k_{\text{прямой}} / k_{\text{обратной}}. \quad (2)$$

Большинство реакций являются сложными, то есть состоят из нескольких стадий, протекающих последовательно или параллельно. Для сложных реакций вводится понятие механизма реакции.

Константы реакций, несмотря на исторически сложившийся термин, константами не являются, а зависят от катализатора, от условий (в зависимости от условий проведения реакции может меняться

механизм реакции) и от температуры. Зависимость от температуры можно выразить уравнением Аррениуса:

$$k = A * e^{-E_a / (R * T)}, \quad (3)$$

где  $A$  – множитель, определяющийся частотой столкновений реагирующих молекул;  $R$  – универсальная газовая постоянная;  $E_a$  – энергия активации.

В настоящее время ряд сервисов предоставляют информацию по кинетике ряда реакций, наиболее популярным сервисом является NIST Chemical Kinetics Database [1]. Он имеет веб-интерфейс для поиска реакций по соединениям, участвующим в реакции.

Разработаны программные библиотеки, содержащие классы для описания структуры химических соединений и химических реакций. Наиболее известны CDK (Chemistry Development Kit) [2] и Cantera [3].

CDK написана на Java, ориентирована на биоинформатику и поддерживает многие форматы химических данных. Данная библиотека содержит большое количество классов, абстрагирующих основные понятия хемоинформатики.

Cantera написана на C++, имеет интерфейсы для работы на Python и нацелена на термодинамические расчеты.

Для хранения химической информации разработано большое количество БД, которые используются при поиске новых веществ с заданными свойствами. Большинство подобных исследований

нацелены на разработку новых лекарственных препаратов [4–6], а также на хранение спектральных данных [7].

### Общая структура системы

Описываемая информационная система (ИС) LascertaLab разработана на языке Java. Общая схема взаимодействия элементов ИС показана на рисунке 1. В структуре прослеживаются три основных слоя: слой API, предоставляющий сторонним программам доступ к данным на программном уровне, слой взаимодействия с источниками данных и слой непосредственно данных. Данные могут храниться в БД (реализация для PostgreSQL) либо в файлах (форматы XML, JSON, YML) с аналогичной структурой.

Важным элементом ИС является блок анализа данных, который позволяет искать количественные корреляции «структура–свойство», а также формировать динамические модели с помощью языка Modelica [8].

### Структура БД для хранения кинетической информации

Разработанная БД является развитием идей, предложенных в [9].

Предложенная БД кинетических данных универсальна, позволяет хранить как информацию о механизмах химических реакции, так и формальные параметры реакции. В ней предусмотрена гибкая система хранения свойств химических соединений и реакций.

### Кодирование информации о химических соединениях

При проектировании химических БД важным вопросом является кодирование химических со-

единений. В настоящее время для кодирования химических соединений наиболее популярен CAS-номер [10], уникальный численный идентификатор химических соединений, внесенных в реестр Chemical Abstracts Service. Данный идентификатор записывается в виде трех чисел, разделенных дефисами. Также в химии приняты несколько систем кодирования веществ: SMILES (Simplified Molecular Input Line Entry System) [11, 12], InChI (International Chemical Identifier – международный текстовый химический идентификатор) [13].

Однако при исследовании могут возникнуть ситуации, когда конкретному веществу CAS-номер не присвоен. Кроме того, в случае многих веществ встречается явление изомерии, когда несколько соединений с одинаковым составом могут иметь различную структуру и обладать существенно отличающимися свойствами. По этим соображениям принято решение использовать в БД уникальный идентификатор, никак не связанный со структурой или составом соединений.

### Представление информации о реакции в БД

Формально химическую реакцию можно кодировать совокупностью исходных веществ и продуктов, однако, как было указано выше, большинство реакций в различных условиях обладают различными свойствами, так что, требуется определять константы скоростей и другие параметры в зависимости от условий проведения эксперимента. На рисунке 2 показана взаимосвязь всех таблиц, созданных для хранения кинетической информации.

Основными таблицами являются **compound** (данные о химических соединениях), **reaction** (данные о реакциях), **catalyst** (данные о катализаторах), **solvent** (данные о среде протекания реакций). В таблице **compound** обязательным полем является только идентификатор соединений (первичный ключ), все остальные поля (идентификатор имени химического соединения, CAS-номер, SMILES- и InChi-коды) необязательные. Таблица **reaction** хранит идентификатор самой реакции (он гарантированно уникальный в рамках конкретной БД), имя реакции (не является обязательным, поскольку многие реакции не имеют названий) и признак обратимости реакций (по умолчанию реакция необратима). Перечень продуктов и исходных веществ каждой реакции хранится в таблице **reaction\_component**, корректность идентификаторов реакции и участвующих в ней веществ обеспечивается внешними ключами с таблицами **reaction** и **compound**. Для продуктов реакции поле **component\_type** имеет значение 1, для исходных веществ – 0, для переходных состояний в данном поле хранится значение 2. Очевидно, что в случае обратимости химической реакции понятие продуктов и субстратов условно, поэтому при выборках из БД данных об обратимых

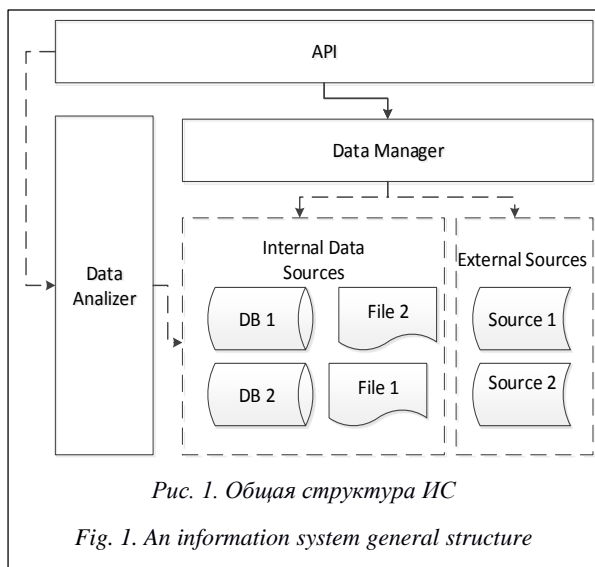


Рис. 1. Общая структура ИС

Fig. 1. An information system general structure

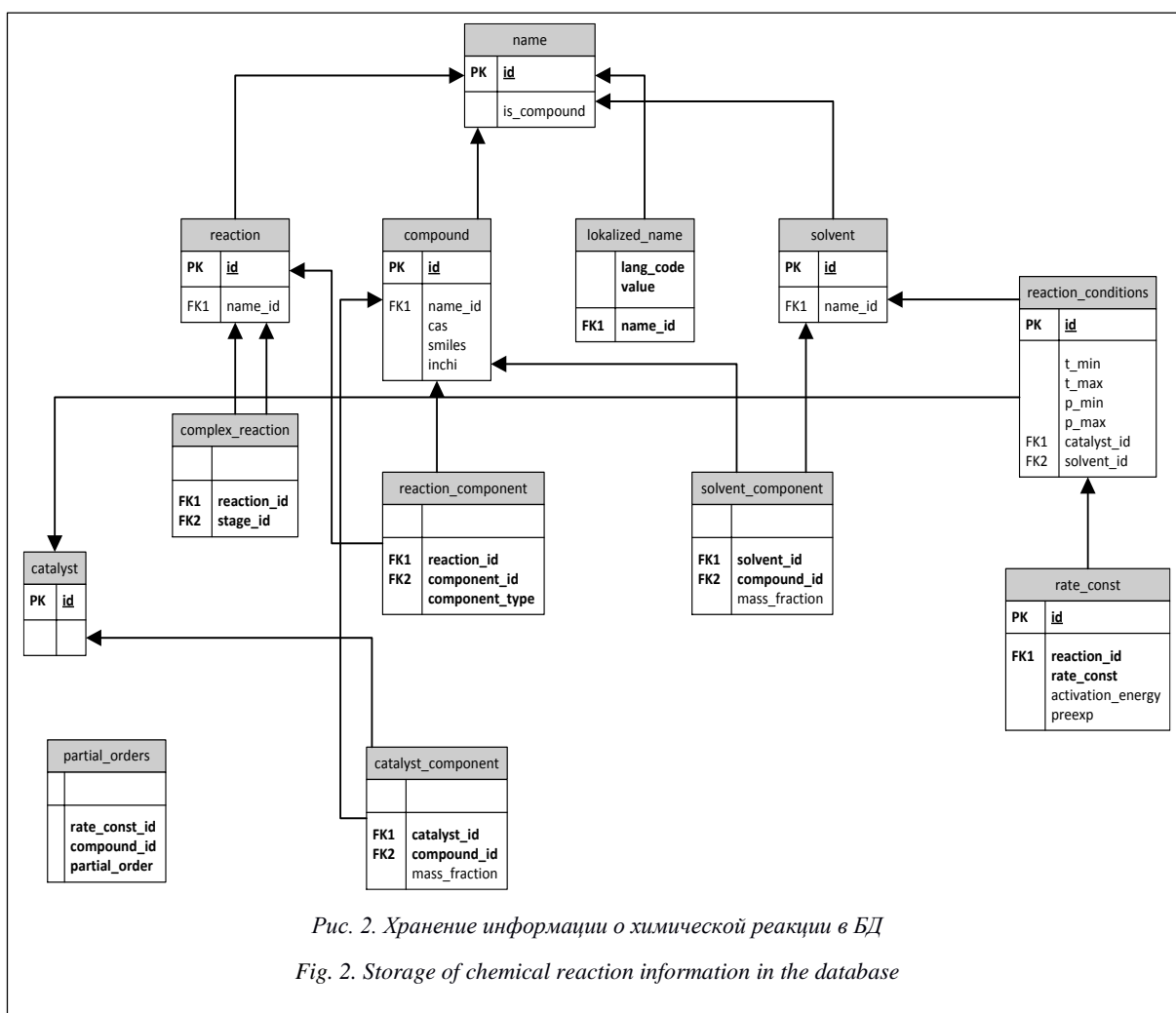


Рис. 2. Хранение информации о химической реакции в БД

Fig. 2. Storage of chemical reaction information in the database

реакциях формируются дополнительные данные для обратных реакций.

Таблица *solvent* предназначена для хранения информации о растворителях. Так как растворитель может быть комбинированным, его состав задается в таблице *solvent\_component*. Данная таблица связана внешними ключами с таблицами *compound* и *solvent* и хранит массовую долю каждого компонента в конкретном растворителе (поле *mass\_fraction*).

Катализатор также может быть комплексным, поэтому для хранения информации о катализаторах также предложены связанные таблицы *catalyst* и *catalyst\_component*. Информация о механизме химической реакции хранится в таблице *complex\_reaction*.

Условия протекания реакции хранятся в *reaction\_conditions*. Данная таблица связана с *solvent* и *catalyst* внешними ключами, а также хранит диапазоны температур и давлений.

Непосредственно кинетические параметры связаны с таблицей *reaction\_conditions*. Кроме констант реакции, в данной таблице могут храниться энергия активации и множитель, введенные ранее в (3).

## Представление свойств в БД

На этапе проектирования трудно сформировать полный перечень параметров соединений и реакций, которые необходимо хранить в БД. По этой причине в БД реализована модификация модели «сущность–атрибут–значение» (Entity–Attribute–Value). На рисунке 3 показана часть схемы, хранящая информацию о свойствах веществ и химических реакций. Предполагается, что можно хранить данные целочисленных типов (их значения в таблице *int\_property*), вещественных типов (*number\_property*) и текстовых свойств (*txt\_property*). Атрибутами являются записи в таблицах *compound\_property* (атрибуты соединений) и *reaction\_property* (атрибуты реакций).

## API для работы с информацией о кинетике химических реакций

Разработан набор Java-классов, методы которых позволяют получать информацию о структуре и свойствах реагирующих веществ, а также данные о механизмах и кинетике химических реакций. Основными классами в библиотеке являются Com-

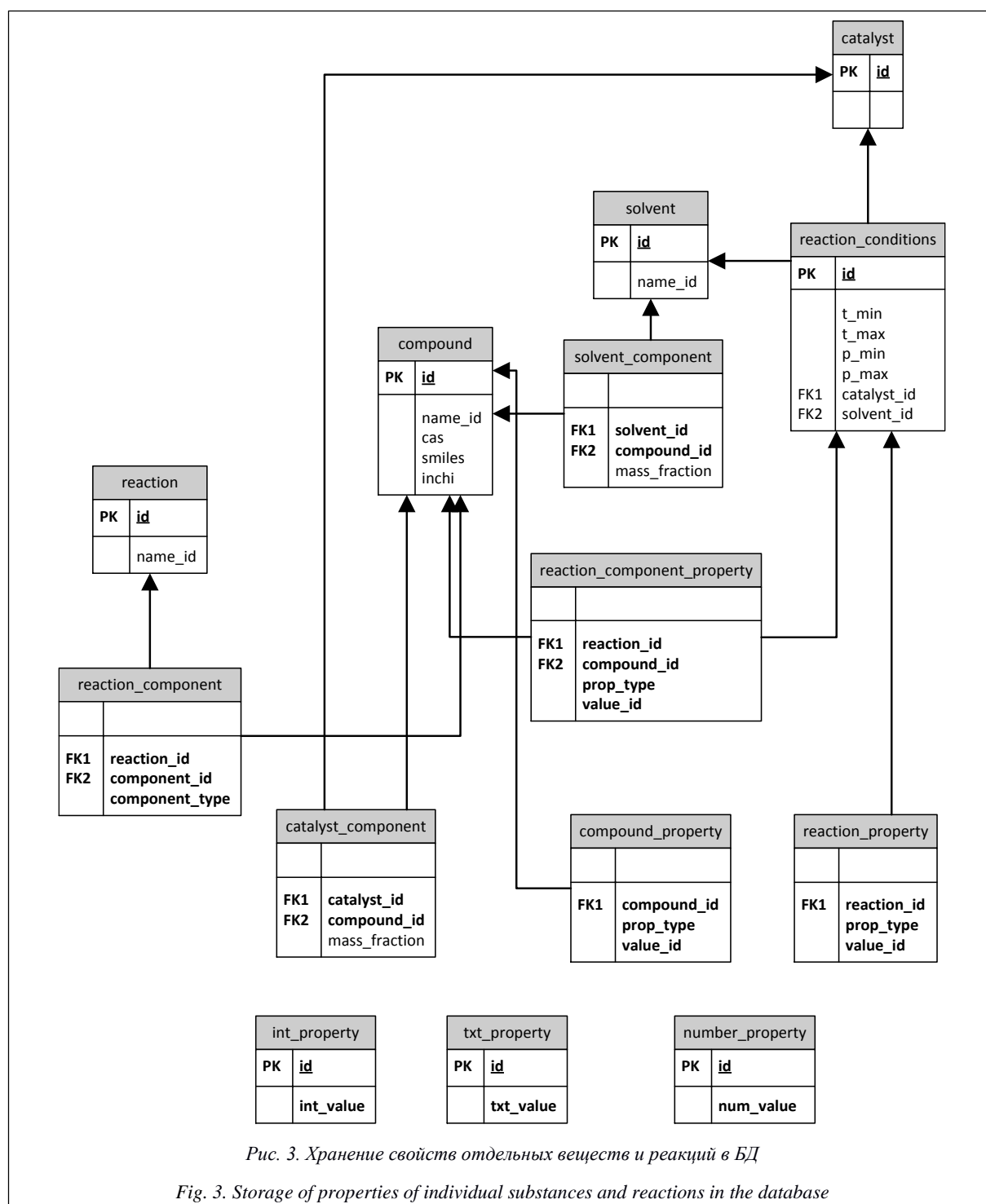


Рис. 3. Хранение свойств отдельных веществ и реакций в БД

Fig. 3. Storage of properties of individual substances and reactions in the database

ound, Reaction, ComplexReaction, Properties, Property. Эти классы соответствуют, с одной стороны, сущностям проблемной области, с другой – сущностям БД. При проектировании API максимальное внимание уделено возможности унификации источников информации о химических веществах и химических реакциях.

Основные возможности API:

- загрузка информации о соединениях и реакциях; разрабатываемые классы имеют возмож-

ность кодировать состав реакции, условия протекания (катализаторы, давление, температура и другие характеристики);

- извлечение данных из внешних БД и сохранение их в БД системы LacertaLab;

- расчет скорости реакции для заданной концентрации и при изменении концентрации с течением времени;

- построение модели «структура–свойство».

Ряд классов в библиотеке соответствуют табли-

цам БД (Compound, Reaction, ComplexReaction, Properties, Property). Для взаимодействия с источниками данных имеется интерфейс DataSource, реализации которого взаимодействуют с источником данных (БД, файл):

```
public interface DataSource {
    public Reaction[] searchReactions(Compound ...
compounds);
    public Compound searchCompound(String code);
    public Compound searchCompoundBy-
Cas(CASNumber cas);
    public void saveCompound(Compound com-
pound);
    public void saveReaction(Reaction reaction);
}
```

Изменение концентрации оксида азота (V) при протекании реакции  $N_2O_5 \rightarrow NO_2 + O_2$  рассчитывается следующим образом:

```
DataSource ds = new DBDataSource("jdbc:post-
gresql://localhost:5432/kineticsdb");
Compound N2O5 = ds.searchCompound("N2O5");
Compound NO2 = ds.searchCompound("NO2");
Compound O2 = ds.searchCompound("O2");
// Проверяем, все ли данные о соединениях счи-
таны корректно
if(O2 == null || NO2 == null || N2O5 == null)
    throw new Exception("Data Error: compounds not
found");
Reaction[] reactions = ds.searchReactions(N2O5,
NO2, O2);
// Должна быть найдена только одна реакция !
if(reactions.length != 1)
    throw new Exception("Data Error: a reaction N2O5
-> NO2 + O2 is double in DB");
Compound[] compounds = {N2O5};
double[] Co = {0.4};
ReactionProcessor processor = new ReactionProces-
sor(reactions[0]).
    setBeginConcentrations(compounds, Co).
    setTimeGrid(10, 40, 5);
processor.calculateConcentrations();
```

В приведенном примере описывается подключение к источнику данных (в рассматриваемом случае источником является БД), далее из БД извлекаются данные о веществах и инициализируются объекты, соответствующие этим веществам. После этого происходит поиск в базе самой реакции и формирование объекта класса Reaction, соответствующего данной реакции. Затем создается объект класса ReactionProcessor, который позволяет задавать параметры реакции (задаются начальные концентрации, интервал времени в секундах) и производить расчет (главным образом, изменение концентраций в процессе реакции). В результате работы данной программы происхо-

дит расчет изменения концентраций (см. <http://www.swsys.ru/uploaded/image/2018-4/2018-4-dop/4.jpg>).

## Заключение

Разработанная ИС LacertaLab для хранения и анализа данных о кинетике химических реакций позволяет хранить как формальные данные о протекании химических реакций, так и информацию о механизме реакции. Разработано API (на языке Java), позволяющее проводить программный анализ кинетических данных, в том числе осуществлять построение моделей «структура–свойство». Разработан функционал, позволяющий в режиме online импортировать данные из наиболее популярных кинетических БД.

## Литература

1. NIST Chemical Kinetics Database. URL: <https://kinetics.nist.gov/kinetics/index.jsp> (дата обращения: 13.06.2018).
2. Steinbeck C., Han Y., Kuhn S., Horlacher O., Luttmann E., Willighagen E. The chemistry development kit (CDK): An open-source java library for chemo- and bioinformatics. J. of Chemical Information and Computer Sciences, 2003, vol. 2, no. 43, pp. 493–500. DOI: 10.1021/ci025584y.
3. Goodwin D.G., Moffat H.K., and Speth R.L. Cantera: An object-oriented software toolkit for chemical kinetics, thermodynamics, and transport processes. URL: <http://www.cantera.org>, 2017. Version 2.3.0 (дата обращения: 13.06.2018). DOI: 10.5281/zenodo.170284.
4. Modelica. URL: <https://www.modelica.org/documents/ModelicaSpec34.pdf> (дата обращения: 13.06.2018).
5. Miller M.A. Chemical database techniques in drug discovery. Nature reviews (drug discovery), 2002, vol. 1, pp. 220–227.
6. Wang Hao. Design of a structure search engine for chemical compound database. Thesis PhD, Georgia State Univ., 2008. URL: [http://scholarworks.gsu.edu/cs\\_diss/33](http://scholarworks.gsu.edu/cs_diss/33) (дата обращения: 13.06.2018).
7. Papadakis E., Anantpinijwatna A., Woodley J.M. and Gani R. A reaction database for small molecule pharmaceutical processes integrated with process information. Processes, 2017, vol. 5, 58 p.
8. Numeric databases for chemical analysis. J. of Research of the National Institute of Standards and Technology, 1989, vol. 94, no. 1, pp. 25–35.
9. Звягинцев Н.В., Гордеев Р.Н., Бурилин А.В. Разработка универсальной базы данных структуры и свойств химических соединений для построения моделей «структура–свойство» на основе эвристических алгоритмов // Инженер. вестн. Дона. 2013. № 4. URL: [http://www.ivdon.ru/uploads/article/pdf/IVD\\_37\\_zvyaginev.pdf\\_1967.pdf](http://www.ivdon.ru/uploads/article/pdf/IVD_37_zvyaginev.pdf_1967.pdf) (дата обращения: 13.06.2018).
10. Support. URL: <http://support.cas.org/content/chemical-substances/faqs> (дата обращения: 13.06.2018).
11. Weininger D. SMILES, a chemical language and information system. Part 1: Introduction to methodology and encoding rules. J. Chem. Inf. Comput. Sci., 1988, vol. 28, no. 1, pp. 31–36.
12. Weininger D., Weininger A., Weininger J.L. SMILES. Part 2. Algorithm for generation of unique SMILES notation. J. Chem. Inf. Comput. Sci., 1989, vol. 29, no. 2, pp. 97–101.
13. Heller S., McNaught A., Stein S., Tchekhovskoi D., Pletnev I. InChI – the worldwide chemical structure identifier standard. J. of Cheminformatics, 2013, vol. 1, no. 5, pp. 32–41. DOI: 10.1186/1758-2946-5-7. PMID 23343401.

**Information system for processing and storing information  
on chemical reaction kinetics**

N.V. Zvyagintsev<sup>1</sup>, Postgraduate Student, n.zvyagintsev@gmail.com

V.A. Billig<sup>1</sup>, Ph.D. (Engineering), Associate Professor, Senior Researcher, Professor, Vladimir-Billig@yandex.ru

<sup>1</sup> Tver State Technical University, Software Department, Tver, 170026, Russian Federation

**Abstract.** Nowadays, chemoinformatics is actively developing. This science allows predicting many properties of substances including reactivity, biological activity, etc. In particular, many specialists investigate QSAR dependencies that allow predicting various properties based on structure data.

The creation of such models requires representative data samples and interfaces to work with this data. Therefore, the creation of specialized databases is an important task of chemoinformatics. At present, there is a number of databases with kinetic parameters of reactions, containing information on chemical reaction kinetics (reaction rate constants, activation energies, etc.). However, most databases with kinetic parameters do not provide specialized services (API) for data software analysis.

The paper considers the information system LacertaLab containing a database and specialized services for processing and analyzing chemical reactions data. The database allows storing both kinetic parameters of the reactions and the data on a chemical reaction mechanism. The information system allows searching for relationships between reaction parameters using the least squares method. The system software services (API) provide access to information from databases and allow building QSAR models, including a number of mathematical functions for processing data (for example, OLS). The information system LacertaLab is able to interact with external services, obtain kinetic parameters from them and store them in its own database.

**Keywords:** chemical kinetics, modeling of chemical reactions, database of chemical compounds and reactions, chemoinformatics.

**References**

1. NIST Chemical Kinetics Database. Available at: <https://kinetics.nist.gov/kinetics/index.jsp> (accessed June 13, 2018).
2. Steinbeck C., Han Y., Kuhn S., Horlacher O., Luttmann E., Willighagen E. The chemistry development kit (CDK): An open-source java library for chemo- and bioinformatics. *J. of Chemical Information and Computer Sciences*. 2003, vol. 2, no. 43, pp. 493–500. DOI: 10.1021/ci025584y.
3. Goodwin D.G., Moffat H.K., Speth R.L. *Cantera: An Object-Oriented Software Toolkit for Chemical Kinetics, Thermodynamics, and Transport Processes*. 2017. Version 2.3.0. Available at: <http://www.cantera.org> (accessed June 13, 2018). DOI: 10.5281/zenodo.170284.
4. *Modelica*. Available at: <https://www.modelica.org/documents/ModelicaSpec34.pdf> (accessed June 13, 2018).
5. Miller M.A. Chemical database techniques in drug discovery. *Nature Reviews (Drug Discovery)*. 2002, vol. 1, pp. 220–227.
6. Wang H. *Design of a Structure Search Engine for Chemical Compound Database*. Thesis PhD, Georgia State Univ. Publ., 2008. Available at: [http://scholarworks.gsu.edu/cs\\_diss/33](http://scholarworks.gsu.edu/cs_diss/33) (accessed June 13, 2018).
7. Papadakis E., Anantpinijwatna A., Woodley J.M., Gani R. A reaction database for small molecule pharmaceutical processes integrated with process information. *Processes*. 2017, vol. 5, 58 p.
8. Numeric databases for chemical analysis. *J. of Research of the National Institute of Standards and Technology*. 1989, vol. 94, no. 1, pp. 25–35.
9. Zvyagintsev N.V., Gordeev R.N., Burilin A.V. Development of a universal database structure and properties of chemical compounds to build models "structure-property" based on heuristic algorithms. *Engineering J. of Don*. 2013, no. 4. Available at: [http://www.ivdon.ru/uploads/article/pdf/IVD\\_37\\_zvyagincev.pdf\\_1967.pdf](http://www.ivdon.ru/uploads/article/pdf/IVD_37_zvyagincev.pdf_1967.pdf) (accessed June 13, 2018).
10. *Support*. Available at: <http://support.cas.org/content/chemical-substances/faqs> (accessed June 13, 2018).
11. Weininger D. SMILES, a chemical language and information system. Part 1: Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* 1988, vol. 28, no. 1, pp. 31–36.
12. Weininger D., Weininger A., Weininger J.L. SMILES. Part 2. Algorithm for generation of unique SMILES notation. *J. Chem. Inf. Comput. Sci.* 1989, vol. 29, no. 2, pp. 97–101.
13. Heller S., McNaught A., Stein S., Tchekhovskoi D., Pletnev I. InChI – the worldwide chemical structure identifier standard. *J. of Cheminformatics*. 2013, vol. 1, no. 5, pp. 32–41. DOI: 10.1186/1758-2946-5-7. PMID 23343401.