

Моделирование и анализ информационных систем. Т. 24, № 6 (2017), с. 811–815
Modeling and Analysis of Information Systems. Vol. 24, No 6 (2017), pp. 811–815

©Бойков В. Н., Каряева М. С., 2017

DOI: 10.18255/1818-1015-2017-6-811-815

УДК 519.7 + 025.4.06

Поэтология: задачи построения тезауруса и спецификации стихового текста

Бойков В. Н., Каряева М. С.

получена 15 октября 2017

Аннотация. Представлено краткое изложение доклада авторов на семинаре “Моделирование и анализ информационных систем”, Ярославль, 17 мая 2017 г. В нём рассматривается взаимосвязь задач по автоматизации построения тезауруса и спецификации текста стихотворного произведения в поэтологии.

Ключевые слова: стих, поэтология, тезаурус, информационно-аналитическая система

Для цитирования: Бойков В. Н., Каряева М. С., "Поэтология: задачи построения тезауруса и спецификации стихового текста", *Моделирование и анализ информационных систем*, **24:6** (2017), 811–815.

Об авторах:

Бойков Владимир Николаевич, математик,
Ярославский государственный университет им. П.Г. Демидова,
ул. Советская, 14, г. Ярославль, 150003 Россия, e-mail: boykov_bh@bk.ru

Каряева Мария Сергеевна, orcid.org/0000-0003-4466-1735, аспирант,
Ярославский государственный университет им. П.Г. Демидова,
ул. Советская, 14, г. Ярославль, 150003 Россия,
e-mail: mari.karyaeva@gmail.com

Благодарности:

Работа выполнена при поддержке Российского фонда фундаментальных исследований (РФФИ), проекты № 16-07-01180 и № 16-06-00497.

Под поэтологией понимается группа дисциплин, ориентированная на всестороннее теоретическое и историческое изучение поэзии (как способа организации речи), ее текстов и стиха (как сегмента поэтического текста).

Задачей изучения поэтического текста как объекта поэтологии является его спецификация, т.е., с одной стороны, задание параметров спецификации – определенной матрицы (формуляра) характерных атрибутов (признаков, свойств, особенностей) некоторого класса объектов исследования; с другой стороны, процедура идентификации объекта – выявление специфических значений этих атрибутов для конкретного объекта.

Такого рода изучение предполагает метаописание поэтического текста в терминах тезауруса по поэтологии и, следовательно, создание такого тезауруса. Под тезаурусом понимается совокупность терминов (слов или словосочетаний), представляющих понятия (концепты) данной предметной области с соответствующими

им определениями, а также отношениями и связями между ними, составляющими их спецификацию [3]. Таким образом, объектами исследования и соответственно установления их спецификации (метаописания) в поэтологии являются как поэтические произведения (стиховые тексты), так и сами термины тезауруса по поэтологии. Комплексное решение этих задач представляет поэтологию как информационно-аналитическую систему [1–2].

Тезаурус по поэтологии в ранее проведенных исследованиях задается как инструмент лингво-стиховедческого и историко-литературного метаописания поэзии и поэтического текста [5–9, 11–14]. Исходный терминологический словарь тезауруса по поэтологии составлен из 1544 слов и словосочетаний. Создан иерархический указатель для верхних уровней иерархии терминов с кластеризацией нижних уровней, т.е. рубрикация терминов тезауруса по поэтологии на подобласти:

1. Стиховедение;
2. Стилистика;
3. Поэтика;
4. Риторика;
5. История литературы;
6. Переводоведение и литературная компаративистика;
7. Текстология;
8. Герменевтика;
9. Теоретические школы и направления;
10. Логика и методология науки.

Не для всех разделов первого уровня тезауруса может применяться автоматическое построение. Большинство терминов тезауруса представляют предметную подобласть Стиховедение, подрубрика которого – первая рубрика второго уровня «Стих» – разбивается на 6 подрубрик третьего уровня:

- 1.1.1 Метрика;
- 1.1.2. Явления начала и конца стихотворной строки;
- 1.1.3. Ритмика;
- 1.1.4. Строфика;
- 1.1.5. Рифмика;
- 1.1.6. Лингвистика стиха.

Основные атрибуты этих подрубрик могут поддаваться автоматическому анализу и задавать соответствующую спецификацию стихового текста.

В широком смысле предназначением рассматриваемой информационно-аналитической системы является установление трех уровней спецификации текста:

С1. Метрико-строфическая спецификация конкретного поэтического произведения (стихового текста);

С2. Поэтико-стилистическая спецификация поэтического творчества конкретного автора на основе метрико-строфической спецификации его произведений;

С3. Историко-литературная спецификация направлений, школ и периодов на основе статистической близости поэтико-стилистической спецификации различных авторов.

В процессе анализа необходимо различать поэтическое произведение как эстетическую целостность и стиховой текст произведения как объект изучения в Стиховедении. В связи с этим следует различать лингво-стиховедческую разметку стиха и

спецификацию стихового текста поэтического произведения. Разметка стиха отражает в основном числовые (количественные и порядковые) характеристики стихового текста, а спецификация представляет собой его метаописание на основе понятий тезауруса по поэтологии.

Разметка стиха (строки) позволяет получить формально-языковое слоговое представление стиха [4,10]:

$$b^{r(1)} A_1 d^{s(1)} b^{r(2)} A_2 d^{s(2)} \dots b^{r(k)} A_k d^{s(k)},$$

где в цепочке символов A означает икты (метрически ударные слоги), k – количество иктов, b и d означают безударные соответственно препозитивные и постпозитивные слоги между иктами, r и s длину соответствующих подцепочек безударных слогов. Подцепочки $b^{r(1)}$ (анакруза), $d^{s(k)}$ (клаузула), $d^{s(j)} b^{r(j+1)}$ (междуиктовый интервал), $j = 1, 2, \dots, (k-1)$, вместе с числом иктов k и числом слогов в строке $(k + \sum_{j=1,2,\dots,(k-1)} (r(j) + s(j+1)))$ представляют собой регулятивы, определяющие в различных сочетаниях их значений репертуар метро-ритмических форм стиха. Полный репертуар метрических схем должен включать во всех формах всевозможные пропуски схемных ударений и сверхсхемные (внеметрические) ударения. Определение метро-ритмической формы стиха является важнейшей характеристикой его спецификации.

С течением времени появляются новые авторы поэтических текстов, подчас приносящие новшества в стихосложение, а в стиховедении появляются новые исследователи, открывающие новые явления в поэтологии, что вносит как дополнения терминов в тезаурус, так и изменения и дополнения спецификации его терминов. Это, в свою очередь, вносит коррективы в матрицу спецификации поэтического текста. Таким образом, поэтология как информационно-аналитическая система не может, да и не должна иметь завершения.

Список литературы / References

- [1] Захаров В.Е., Бойков В.Н., Вигурский К.В., Пильщиков И.А., *Русская поэзия: проблемы консолидации и анализа в электронном формате*, Москва, 2004; [Zakharov V.E., Boykov V.N., Vigurskiy K.V., Pil'shchikov I.A., *Russkaja poezija: problemy konsolidacii i analiza v elektronnom formate*, Moskva, 2004., (in Russian).]
- [2] Бойков В.Н., Захаров В.Е., Пильщиков И.А., Сысоев Т.М., “Тезаурус как инструмент поэтологии”, *Модел. и анализ информ. систем*, **17:1** (2010), 5–23; [Boikov V.N., Zakharov V.E., Pil'shchikov I.A., Sysoev T.M., “Thesaurus as a poetological tool”, *Model. Anal. Inform. Sist.*, **17:1** (2010), 5–23, (in Russian).]
- [3] Лукашевич Н.В., *Тезаурусы в задачах информационного поиска*, Издательство Московского университета, М., 2011; [Lukashevich N.V., *Tezaurusy v zadachah informacionnogo poiska*, Izdatelstvo Moskovskogo universiteta, Moscow, 2011, (in Russian).]
- [4] Бойков В.Н., “Контекстно-свободная грамматика одной ритмической модели русского стиха”, *Модел. и анализ информ. систем*, **19:4** (2012), 154–167; [Boikov V.N., “A Context-Free Grammar of One Rhythmic Model of Russian Verse”, *Model. Anal. Inform. Sist.*, **19:4** (2012), 154–167, (in Russian).]
- [5] Бойков В.Н., Пильщиков И.А., “Семантическая модель "Тезауруса по поэтологии" в составе информационно-аналитической системы”, *Интернет и современное общество*, Труды XVI Всероссийской объединенной конференции IMS-2013, Санкт-Петербург, 2013, 273–279; [Boikov V.N., Pil'shchikov I.A., “Semanticheskaya model

- ”Tezaurusa po poehtologii” v sostave informacionno-analiticheskoy sistemy”, *Internet i sovremennoe obshchestvo*, Trudy XVI Vserossiyskoy konferencii IMS-2013, Sankt-Peterburg, 2013, 273–279, (in Russian).]
- [6] Бойков В.Н., Захаров В.Е., Каряева М.С., Соколов В.А., “Тезаурус по поэтологии как инструмент для информационного поиска и коллекции знаний”, *Модель и анализ информ. систем*, **20**:4 (2013), 125–135; [Boykov V.N., Zakharov V.E., Karyayeva M.S., Sokolov V.A., “Domain-Specific Thesaurus as a Tool for Information Retrieval and Collection of Knowledge”, *Model. Anal. Inform. Sist.*, **20**:4 (2013), 125–135, (in Russian).]
- [7] Бойков В.Н., Захаров В.Е., Каряева М.С., Соколов В.А., “Предметно-ориентированный тезаурус в открытой информационно-аналитической системе”, *Электронные библиотеки: перспективы, методы и технологии, электронные коллекции*, RCDL’2013, 2013, 70–76; [Boykov V.N., Zakharov V.E., Karyayeva M.S., Sokolov V.A., “Predmetno-orientirovannyj tezaurus v otkrytoj informacionno-analiticheskoy sisteme”, *Ehlektronnye biblioteki: perspektivy, metody i tekhnologii, ehlektronnye kolekcii*, RCDL’2013, 2013, 70–76, (in Russian).]
- [8] Бойков В.Н., Захаров В.Е., Каряева М.С., Соколов В.А., “Об автоматической рубрикации терминов тезауруса открытой информационно-аналитической системы”, *RCDL-2014*, Дубна, 2014; [Boykov V.N., Zakharov V.E., Karyayeva M.S., Sokolov V.A., “Ob avtomaticheskoy rubrikacii terminov tezaurusa otkrytoj informacionno-analiticheskoy sistemy”, *RCDL-2014*, Dubna, 2014, (in Russian).]
- [9] Бойков В.Н., Захаров В.Е., Каряева М.С., Соколов В.А., “Информационно-аналитический блок тезауруса как инструмент обучения”, *Образование, наука и экономика в вузах и школах. Интеграция в международное образовательное пространство*, Труды международной научной конференции, Цахкадзор, 2014, 205–209, http://nuclphys.sinp.msu.ru/math/conf/tsaghkadzor_conference_part1.pdf; [Boykov V.N., Zakharov V.E., Karyayeva M.S., Sokolov V.A., “Informacionno-analiticheskij blok tezaurusa kak instrument obuchenija”, *Education, science and economics at universities and schools*, Publications International Scientific Conference Integration to international educational area, Tsaghkadzor, 2014, 205–209, http://nuclphys.sinp.msu.ru/math/conf/tsaghkadzor_conference_part1.pdf, (in Russian).]
- [10] Бойков В.Н., Каряева М.С., Соколов В.А., Пильщиков И.А., “Об автоматической спецификации стиха в информационно-аналитической системе”, *DAMDID/RCDL-2015*, Обнинск, 2015; [Boykov V.N., Karyayeva M.S., Sokolov V.A., Pil’schikov I.A., “Ob avtomaticheskoy specifikacii stiha v informacionno-analiticheskoy sisteme”, *DAMDID/RCDL-2015*, Obninsk, 2015, (in Russian).]
- [11] Каряева М.С., “Лингвостатистический анализ терминологии для построения тезауруса предметной области”, *Модель и анализ информ. систем*, **22**:6 (2015), 834–851; [Karyayeva M.S., “Linguistic and statistical analysis of the terminology for constructing the thesaurus of a specified field”, *Model. Anal. Inform. Sist.*, **22**:6 (2015), 834–851, (in Russian).]
- [12] Каряева М.С., Соколов В.А., “Обзор методов извлечения гипо/гиперонимов из коллекций документов”, *Заметки по информатике и математике*, Сборник научных статей, **7**, ЯрГУ им. П.Г. Демидова, Ярославль, 2015, 53–58, <http://www.lib.uniyar.ac.ru/edocs/iuni/20150406.pdf>; [Karyayeva M.S., Sokolov V.A., “Obzor metodov izvlechenija gipo/giperonimov iz kolekcij dokumentov”, *Zametki po informatike i matematike*, Sbornik nauchnyh statej, **7**, YarGU im. P.G. Demidova, Jaroslavl, 2015, 53–58, <http://www.lib.uniyar.ac.ru/edocs/iuni/20150406.pdf>, (in Russian).]
- [13] Каряева М.С., Соколов В.А., “Об информационных технологиях решения задачи извлечения терминов предметной области”, *Математика, физика, информатика и их приложения в науке и образовании*, Сборник тезисов докладов Международной школы-конференции молодых ученых, МИРЭА, М., 2016, 143–145, <https://elibrary.ru/item.asp?id=28373830>; [Karyayeva M.S., Sokolov V.A., “Ob informacionnyh tehnologijah reshenija zadachi izvlechenija terminov predmetnoj oblasti”, *Mathematics, physics, informatics and their applications in science and education*, Moskva, 2016, 143–145, <https://elibrary.ru/item.asp?id=28373830>, (in Russian).]

- [14] Karyaeva M. S., Sokolov V.A., “On the Problem of Multi-word Term Extraction from a Domain-specific Document Collection”, *Data Analytics and Management in Data Intensive Domains (DAMDID/RCDL'2017)*, Proceedings of the XIX International Conference, Moscow, 218–221.

Boykov V. N., Karyaeva M. S. , "Poetology: Problems of Constructing a Thesaurus and Verse Text Specification", *Modeling and Analysis of Information Systems*, **24:6** (2017), 811–815.

DOI: 10.18255/1818-1015-2017-6-811-815

Abstract. It is a brief version of the report made by the authors at the seminar “Modeling and Analysis of Information Systems”, Yaroslavl, May 17, 2017. The interrelation of problems of computer-aided constructing the thesaurus and specification of verse texts in poetology is considered.

Keywords: verse, poetology, thesaurus, information-analytical system

On the authors:

Vladimir N. Boykov, mathematician,
P.G. Demidov Yaroslavl State University,
14 Sovetskaya str., Yaroslavl 150003,
Russia, e-mail: boykov_bh@bk.ru

Maria S. Karyaeva, orcid.org/0000-0003-4466-1735, graduate student,
P.G. Demidov Yaroslavl State University,
14 Sovetskaya str., Yaroslavl 150003, Russia,
e-mail: mari.karyaeva@gmail.com

Acknowledgments:

This work was supported by RFBR, projects № 16-07-01180, № 16-06-00497.