Проектирование информационной системы представления результатов комплексного анализа поэтических текстов

В. Б. Барахнин^{1,2}, О. Ю. Кожемякина¹, Ю. С. Борзилова¹

¹ *Институт вычислительных технологий СО РАН*

² *Новосибирский государственный университет
Новосибирск, Россия*

Аннотация

В рамках проекта по автоматизации работы с поэтическими текстами, реализуемого на базе Института вычислительных технологий СО РАН, ведется комплекс исследований, связанных с анализом поэтических текстов. Каждый компонент проекта относится к одному из структурных уровней анализа текста: структурный, семантический, прагматический. Структурный анализ поэтического текста связан с выделением его метроритмических характеристик. В рамках семантического анализа ведутся исследования по выделению смысловых конструкций из произведения. Прагматический уровень включает в себя исследования по автоматическому определению высокоуровневых характеристик поэтического текста, таких как жанр и стиль. В данной работе описан процесс проектирования и реализации информационной системы представления результатов анализа поэтических текстов. На этапе проектирования сформулированы задачи, которые призвана решать информационная система, а также изложены требования в порядке приоритета для общего проекта. Представленная информационная система объединяет в себе разнородную информацию о результатах анализа поэтических текстов, полученных на каждом из уровней представления. Исходя из потребностей потенциальных пользователей, выполнено описание внешних взаимодействующих элементов системы. Разработан тестовый интерфейс для доступа к хранилищу информационной системы. Реализация информационной системы обеспечит существенное упрощение исследований поэтических текстов.

Ключевые слова

многоуровневая модель информации, автоматизация анализа поэтических текстов, проектирование информационной системы

Благодарности

Работа выполнена при частичной поддержке РФФИ, проект 18-07-01457 «Модели и методы создания информационных систем поддержки научных исследований, интегрированных в открытое семантическое пространство»

Для цитирования

Барахнин В. Б., Кожемякина О. Ю., Борзилова Ю. С. Проектирование информационной системы представления результатов комплексного анализа поэтических текстов // Вестник НГУ. Серия: Информационные технологии. 2019. Т. 17, № 1. С. 5–17. DOI 10.25205/1818-7900-2019-17-1-5-17

The Development of the Information System of the Representation of the Complex Analysis Results for the Poetic Texts

V. B. Barakhnin^{1,2}, O. Yu. Kozhemyakina¹, Yu. S. Borzilova¹

¹ *Institute of Computational Technologies SB RAS*

² *Novosibirsk State University
Novosibirsk, Russian Federation*

Abstract

The project on automation of work with poetic texts, implemented in the Institute of Computational Technologies SB RAS, includes a complex of studies related to the analysis of poetic texts is carried out. Each component of the project

belongs to one of the structural levels of text analysis: structural, semantic, pragmatic. The structural analysis of the poetic text is associated with the definition of its metrorhythmic characteristics. In the context of semantic analysis, the research on the extraction of semantic structures from the poetic texts is carried out. The pragmatic level includes the research on the automatic identification of high-level characteristics of poetic text, such as genre and style. This paper describes the process of designing and implementing of the creation of an information system for presenting the results of the analysis of poetic texts. At the design stage, the tasks to be solved by the information system are formulated, as well as the requirements in order of priority for the overall project. The presented information system combines heterogeneous information about the results of the analysis of poetic texts obtained at each level of representation. Based on the needs of potential users, the description of the external interacting elements of the system is performed. The test interface for the access to the information system storage was developed. The implementation of the information system will provide a significant simplification of the research of poetic texts.

Keywords

multilevel information model, automation of the analysis of poetic texts, development of information system

Acknowledgements

The work is partly supported by RFBR, project 18-07-01457 "Models and methods of creating information systems supporting scientific research, integrated into open semantic space"

For citation

Barakhnin V. B., Kozhemyakina O. Yu., Borzilova Yu. S. The Development of the Information System of the Representation of the Complex Analysis Results for the Poetic Texts. *Vestnik NSU. Series: Information Technologies*, 2019, vol. 17, no. 1, p. 5–17. (in Russ.) DOI 10.25205/1818-7900-2019-17-1-5-17

Введение

Одной из актуальных задач в области использования информационных технологий для гуманитарных исследований является создание программного приложения для автоматизации комплексного анализа поэтических текстов, призванного облегчить труд филологов. Комплекс таких разработок ведется в Институте вычислительных технологий (ИВТ) СО РАН, концепция создаваемой информационной системы изложена в работах [1; 2].

Среди аналогичных исследований других авторов следует назвать, прежде всего, статью [3], в которой была намечена большая программа исследований метрических, ритмических и фонетических (включая рифму) характеристик русских поэтических текстов. Эта программа опиралась, в свою очередь, на систему STARLING¹, которая являлась частью проекта «Автоматизированный лингвостиховедческий анализ русских поэтических текстов» (после ухода из жизни руководителя проекта С. А. Старостина работы в этом направлении были прекращены). Именно на основе названных исследований нами реализовано действующее в настоящее время программное средство анализа метrorитмических характеристик поэтических текстов², описанное в работе [2]. Однако алгоритм из работы [3] носит полуэмпирический характер, что снижает его точность в случаях сложной акцентуации, поэтому для дальнейших исследований нами реализуется более строго обоснованный алгоритм из статьи [4], модифицируемый с учетом неоднозначной акцентуации текстов на русском языке.

Больших успехов достиг коллектив авторов под руководством И. А. Пильщикова и А. С. Старостина: начиная с 2008 г. выполнен ряд работ, посвященный автоматическому распознаванию метра в силлабо-тонических стихах [5–7]. В 2016 г. в устном докладе «Инструментальная среда для работы с русскоязычными стихотворными корпусами и их специализированной разметкой»³ продемонстрирована инструментальная компьютерная среда РМС («Рабочее место стиховеда»); описана возможность эвристической акцентуации несловарных слов, создание интерфейса редактора шаблонов, визуализация результатов автоматического анализа метрики и ритмики стиха. Однако нам неизвестны работы, в которых авторы проводили исследо-

¹ Проект этимологической базы данных «Вавилонская башня». URL: <http://starling.rinet.ru/indexru.htm>.

² Анализ поэтических текстов онлайн. URL: <http://poem.ict.nsc.ru/>.

³ Запись доклада А. С. Старостина «Инструментальная среда для работы с русскоязычными стихотворными корпусами и их специализированной разметкой». URL: <https://youtu.be/TUWlzxteNY>.

вание автоматизации анализа характеристик более высокого уровня (например, определение жанра).

Так как задачи анализа фонетических и лексических характеристик текста менее специфичны и решаются намного чаще, чем характерная именно для поэтических текстов задача анализа метроритмических характеристик, то для решения этих задач нами используются более или менее стандартные алгоритмы, изложенные в статье [2].

Наконец, для определения стилей и жанровых характеристик текстов используются наиболее известные приемы ансамблирования базовых алгоритмов в композиции, такие как взвешенное голосование, бустинг и стекинг [8] (аналогичные подходы используются для других текстов в [9; 10]).

Рассмотрим некоторые из существующих информационных систем, предназначенных для исследования тех или иных характеристик поэтических текстов: проект «Конкорданс к текстам Ломоносова», система SPARSAR, веб-приложение Metricalizer.

Проект «Конкорданс к текстам Ломоносова»⁴ начат в 2009 г. и строится на основе корпуса авторских текстов, снабженных структурной, филологической и грамматической разметкой. Проект доступен через веб-интерфейс⁵ и представляет собой как алфавитно-частотный конкорданс к текстам, так и сборник изданий текстов Ломоносова. Технологическая цепочка работы с корпусом включает в себя немалую долю ручной разметки корпуса и сегментации текста; с помощью парсера выполняется морфологический разбор с последующей постобработкой (снятие омонимии, исправление ошибок). Практическая реализация упрощала взаимодействие с созданным конкордансом – пользователь может в интерактивном режиме выбрать подходящий термин для работы с ним. Проект не был завершен авторами.

Описанная в работе Р. Дельмонте [11] система SPARSAR предполагает автоматический комплексный анализ поэтических текстов с целью изучения их стиля. SPARSAR⁶ выполняет анализ каждого стихотворения на разных уровнях: на уровне предложения, на уровне строки и на уровне строфы. Такая информационная система была бы полезна авторам статьи, однако детальное описание внутренней структуры уделяется модулю, связанному с автоматическим чтением текста (TextToSpeech [12]).

Веб-приложение Metricalizer⁷, разработанное К. Боббенхаузенем и Б. Хаммерихом, позволяет производить автоматический анализ метрических характеристик немецких стихов [13]. В системе предусмотрены метрический анализ стихотворения, создание XML-документов по результатам проводимого анализа, разбор текстов по акцентуации и рифмовке, фонетический разбор слов в формате X-SAMPA (Extended Speech Assessment Methods Phonetic Alphabet, «расширенный фонетический алфавит методов оценки речи»). Упомянутая система наиболее релевантно соотносится со структурой, которую можно применить для информационной системы в ИВТ СО РАН, однако в работе [13] детальная структура системы Metricalizer не представлена.

Разнообразие алгоритмов обработки поэтических текстов, несомненно, является важным фактором успешного решения описываемой задачи. Однако во многих случаях остается актуальным вопрос реализации алгоритма в виде отчуждаемого программного продукта и его использование без непосредственного участия авторов. В силу этого программное приложение должно представлять собой некую информационную подсистему со специфическими входными и выходными данными. Важно настроить взаимодействие подсистемы как с конечным пользователем, так и с другими подсистемами.

Таким образом, необходимо разработать подход к проектированию и реализации системы анализа поэтических текстов, учитывающий ее сложную модульную структуру. Цель данной

⁴ Конкорданс к текстам Ломоносова – концепция и реализация. URL: <http://www.dialog-21.ru/digests/dialog2009/materials/html/61.htm>.

⁵ Поляков А. Е., Пильщиков И. А., Бергельсон М. Б. Конкорданс к текстам Ломоносова. ФЭБ, 2009. URL: <http://feb-web.ru/feb/lomoconc/abc/>.

⁶ SPARSAR, an Expressive Poetry Reader. URL: <https://sparsar.wordpress.com>.

⁷ Metricalizer. URL: <https://metricalizer.de>.

статьи заключается в описании процесса проектирования информационной системы автоматизации комплексного анализа поэтических текстов.

Постановка задачи

Процесс анализа поэтических текстов сводится к следующим последовательным шагам (рис. 1), на которых осуществляется анализ характеристик текста:

- инициализация – формирование корпуса поэтических текстов и его предобработка для последующего анализа;
- структурный анализ – определение низкоуровневых характеристик поэтического текста (фонетика и метроритмика стихотворения);
- семантический анализ – определение смысловых конструкций с учетом поэтической синонимии;
- прагматический анализ – экспертная оценка принадлежности к определенным стилевым характеристикам для поэтического текста (жанр, стиль и др.);
- синтез проведенных исследований – определение влияния низших уровней поэтических текстов на более высокие, а также объединение результатов в удобном для восприятия и поиска виде.

Рис. 1. Взаимосвязи в процессе анализа поэтических текстов
Fig. 1. The interrelations in the process of the analysis of poetic texts

Концептуальное проектирование включало в себя формирование возможностей создаваемой информационной системы. Создаваемая система должна обладать следующими возможностями.

1. Обеспечение доступа к корпусам поэтических текстов. При этом к текстам могут предъявляться дополнительные требования. Например, в процессе обработки важно учитывать особенности старой (времени создания текста) орфографии.
2. Автоматизированная обработка корпуса поэтических текстов, хранящихся в БД:
 - а) определение фонетических характеристик текста;

- б) исследование метроритмических характеристик: метрика, стопность; рифмовка строфики и др. с указанием неоднозначностей, которые не могут быть разрешены автоматически;
- в) определение лексических характеристик текста;
- г) определение жанрово-стилевых характеристик текста.

3. Занесение полученных характеристик в хранилище (базу данных).

4. Статистическая обработка полученных характеристик и их представление в удобном для исследователя виде.

5. Возможность импорта корпусов поэтических текстов из баз данных и файлов и их экспорта в другие информационные системы для дальнейшей автоматизированной обработки.

Алгоритмы, реализуемые при прохождении перечисленных выше этапов анализа (структурного, семантического, прагматического), изложены в работах авторов под руководством В. Б. Барахнина [2; 8; 14]. Однако остается актуальной задача синтеза проводимых исследований в пределах общей информационной системы, пригодной для использования экспертами-филологами. В рамках текущего исследования все подсистемы проекта будут рассматриваться в виде «черного ящика» – ведущую роль играют данные, подаваемые на вход и получаемые на выходе. В качестве базового инструмента проектирования на начальном этапе использовались интеллект-карта и диаграмма прецедентов. На рис. 2 процесс анализа поэтических текстов представлен в нотации IDEF0. Выделяются следующие бизнес-процессы верхнего уровня.

1. Преобработка текста (процесс A1): на вход поступает необработанный текст (возможно, в дореформенной орфографии) с информацией об этом тексте. Под управлением правил формирования дореволюционной и современной орфографии происходит преобразование текста в соответствии с современной орфографией с занесением результатов в систему хранения. Дальнейшая обработка текста осуществляется в современной орфографии.

2. Структурный анализ (процесс A2): на основании формализованных правил построения метра и ритма поэтического текста выполняется извлечение его структурных характеристик: тип рифмы, количество женских и мужских окончаний и др.

3. Семантический анализ (процесс A3): на основании формализованных правил построения словосочетаний выполняется извлечение синтаксических конструкций из поэтического текста и их идентификация с использованием базового словаря, учитывающего синонимию.

4. Прагматический анализ (процесс A4): с использованием системы классификации жанров и стилей на выходе формируется гипотеза о принадлежности поэтического текста к определенному жанру и стилю.

Информационная система (ИС) должна учитывать этапы анализа поэтических текстов. Структура системы состоит из компонентов, перечисленных при описании постановки задачи. Связи в структуре показаны на рис. 3. Компонент фонетического анализа включен в систему как готовый модуль, ранее разработанный в ИВТ СО РАН, он осуществляет акцентуацию (расстановку ударений) и транскрипцию слов. Включенные в модуль элементы метроритмики используют входные данные, получаемые на этапе структурного анализа. Учет вхождений слов и словосочетаний в те или иные поэтические тексты связан с задачей составления частотных справочников и конкордансов. При прагматическом анализе определяются жанр и стиль произведения. Компоненты пользовательского интерфейса представляют собой ключевые возможности по работе с системой: просмотр характеристик и сравнение поэтических текстов по ним, а также поиск произведений.

Технические требования, предъявляемые к разрабатываемой ИС, отражены в таблице (см. далее). Для каждого требования выставлен приоритет его выполнения:

- высокий – выполнение таких требований необходимо;
- средний – выполнение таких требований желательно;
- низкий – выполнение таких требований необязательно.

Рис. 2. Процесс анализа поэтических текстов в нотации IDEF0
 Fig. 2. The process of the analysis of poetic texts in IDEF0 notation

ИС "Поэтические тексты и их анализ"

Рис. 3. Структура информационной системы

Fig 3. The structure of the information system

Требования, предъявляемые к ИС
The requirements for the information system

Требование	Приоритет
Учитывать вышеописанные этапы комплексного анализа	Высокий
Отображать метроритмические характеристики	Высокий
Отображать результаты семантического анализа по каждому отдельному произведению	Средний
Разработать функционал ручной корректировки данных экспертом	Средний
Реализовать доступ к ИС для конечного пользователя	Высокий
Предусмотреть логику работы системы для повторяющихся значений	Средний
Спроектировать логику работы системы для дальнейшего ее масштабирования	Высокий
Разработать API (программный интерфейс приложения) для доступа к ИС	Низкий

На рис. 4 представлена диаграмма прецедентов. Система проектируется с учетом того, что в процессе работы она будет использоваться различными категориями пользователей и взаимодействовать с другими системами. В качестве акторов были выбраны:

- пользователь – филолог, проводящий исследование поэтических текстов;
- программист-исследователь – лицо, ответственное за наиболее полную реализацию функционала и бесперебойную работу системы, а также за математическую обработку результатов;
- администратор системы – лицо, ответственное за использование особых технических средств, добавление нового контента, актуализацию данных и т. п.;
- внешняя система – та или иная подсистема проекта по автоматизации анализа поэтических текстов, упомянутая ранее; экспортирует результаты проведенных исследований в разрабатываемую систему для комплексного отображения результатов.

Роли программиста-исследователя и пользователя могут быть объединены в зависимости от специфики выполняемой филологической задачи. Кроме того, пользователь, как эксперт, может проводить оценку данных, выводимых системой, и в случае возникновения спорных моментов выполнить редактирование данных в ИС. Все действия в системе должны протоколироваться администратором – например, при импорте данных от подсистем проекта могут возникнуть непредвиденные технические проблемы, которые необходимо устранить вручную.

Далее приведем некоторый сценарий взаимодействия возможных участников системы (согласно диаграмме прецедентов). Пользователь ИС имеет возможность просматривать информацию о произведениях, после проведения экспертной оценки и в случае нахождения спорных данных пользователь-филолог отправляет запрос на доступ к редактированию или информирует администратора БД о результатах экспертной оценки и решении об изменении данных. Администратор БД проверяет экспортируемые данные от внешних систем, которые проводят анализ поэтических текстов на технические ошибки (в случае если внешняя система получила ошибки при результате анализа).

Программист-исследователь будет иметь возможность с помощью API организовать пакетную выгрузку данных о проанализированных поэтических текстах. Поскольку текущий функционал отражен в требованиях к системе как «требование с низким приоритетом», работа над API находится на стадии проектирования.

Внешние системы (подсистемы) с помощью заранее установленных протоколов взаимодействия загружают результаты анализа лирических произведений, полученные при работе отдельных алгоритмов (в том числе из подсистемы, описанной в [2]). В перспективе разраба-

тывается вариант взаимодействия, при котором подсистемы в качестве входных параметров используют данные описываемой информационной системы.

Рис. 4. Диаграмма прецедентов ИС
Fig. 4. The use case diagram of the information system

Рис. 5. Концептуальная ER-модель информационной системы
Fig. 5. The conceptual ER-model of the information system

Болдинская осень
А. С. Пушкин

- ☐ Бесы
- ☐ Безумных лет угасшее веселье...
- ☐ Труд
- ☐ Ответ анониму
- ☐ Царскосельская статуя
- ☐ к переводу "Иллиады"
- ☐ Румяный критик мой, насмешник толстопузы
- ☐ Глухой глухого звал...
- ☒ **Дорожные жалобы**
- ☐ Прощание
- ☐ Паж, или Пятнадцатый год
- ☐ Я здесь, Инезилья
- ☐ Пред испанкой благородной

То ли дело рюмка рома,
 Ночью сон, поутру чай;
 То ли дело, братцы, дома!..
 Ну, пошёл же, погоняй!..

Автор: Пушкин Александр Сергеевич
[Ссылка на оригинал](#) (откроется в новом окне)
[\[Словосочетания\]](#)
[\[Анализ поэтических текстов онлайн \(ИВТ СО РАН\)\]](#)

Количество строк в строфе	4
Тип рифмовки	перекрестная (ABAB);
Количество слогов в строках	
Жанр	элегия
Образец (А - женские окончания; а - мужские)	

Метро-ритмическая статистика:

Размерность	хорей
Стопность	
Число строк	32

Рис. 6. Фрагмент интерфейса ИС
 Fig. 6. A fragment of the information system interface

Основная часть ИС реализована в виде базы данных, хранящей в себе как сами стихотворения, так и их характеристики. Концептуальная ER-модель системы показана на рис. 5. Для стандартизации описания метаданных использовался ГОСТ Р 7.0.10-2010⁸, опирающийся на набор элементов «Дублинского ядра»⁹ – норматива для общесистемного описания информационных ресурсов.

Сущность POEMS включает в себя:

- 1) фонетические характеристики (сущность PHONETICS), на основании которых определяются метроритмические характеристики (сущность MRSTATISTICS); последняя, в свою очередь, включает в себя метаданные, приведенные на рис. 2;
- 2) словарь слов (сущность DICTIONARY) связан с сущностями «словосочетания» (COLLOCATIONS), которые извлекаются из текста; ключевыми метаданными сущности COLLOCATIONS являются статистика упоминания в произведении (PoemStatistics) и у данного автора (AuthorStatistics) вообще.;
- 3) характеристики высокого уровня (сущность HIGH LEVEL CHARACTERISTICS) включают в себя данные о жанре (GENRE) и стиле (STYLE);
- 4) другие характеристики (сущность OTHERS) являются вспомогательными и служат для хранения разнородной внешней информации о произведениях, например источник (Source), сведения об издании (Publisher) и др.

Для организации доступа к ИС был спроектирован и разработан интерфейс пользователя (рис. 6), доступный через веб-браузер клиента. Запросы к базе данных организованы в формате SQL, что повышает универсальность использования системы. Интерфейс ИС доступен по адресу: http://db4.sbras.ru/~poems_user/IS. При тестовой реализации системы в качестве

⁸ ГОСТ Р 7.0.10-2010 (ИСО 15836:2003). Набор элементов метаданных «Дублинское ядро» [Текст]. Введ. 2011-07-01. М.: Стандартинформ, 2011. 12 с. (Система стандартов по информации, библиотечному и издательскому делу).

⁹ The Dublin Core Metadata Initiative. URL: <http://dublincore.org>.

экспериментального корпуса текстов было решено использовать цикл лирических произведений А. С. Пушкина, написанных в 1830 г., в так называемую «Болдинскую осень».

Выбирая доступные произведения автора (меню слева), пользователь получает возможность просмотреть само произведение и ознакомиться с его характеристиками. Для некоторых произведений приведена ссылка на рукописный оригинал – использовались ресурсы Института русской литературы¹⁰ (ИРЛИ) РАН. Дополнительное окно словосочетаний открывается при нажатии на одноименную ссылку в режиме просмотра произведения: пользователь имеет возможность просмотреть выделенные словосочетания и сочетания слов и при наличии соответствующих прав доступа изменить их (после проведения экспертной оценки). Реализована функциональность сравнения нескольких произведений – пользователь выбирает необходимые для анализа произведения: для просмотра отображается сводная таблица произведений с ключевыми характеристиками.

Заключение

Таким образом, в результате выполнения работы осуществлено проектирование и реализация информационной системы, осуществляющей интеграцию созданных ранее в ИВТ СО РАН модулей, предназначенных для решения отдельных задач комплексного анализа поэтических текстов. Разработан прототип информационной системы. С помощью диаграммы прецедентов показано, какие действия может выполнять каждый участник системы. Реализован графический интерфейс информационной системы. Результаты работы показали принципиальную возможность интеграции упомянутых выше компонентов системы, что открывает перспективы создания весьма востребованного филологами инструментария автоматизации комплексного анализа поэтических текстов на русском языке.

Список литературы / References

1. **Баракхнин В. Б., Кожемякина О. Ю.** Об автоматизации комплексного анализа русского поэтического текста // *CEUR Workshop Proceedings*. 2012. Т. 934. С. 167–171.
Barakhnin V. B., Kozhemyakina O. Yu. About the automation of the complex analysis of Russian poetical text. *CEUR Workshop Proceedings*, 2012, vol. 934, p. 167–171. (in Russ.)
2. **Баракхнин В. Б., Кожемякина О. Ю., Забайкин А. В., Хаятова В. Д.** Автоматизация комплексного анализа русского поэтического текста: модели и алгоритмы // *Вестник НГУ. Серия: Информационные технологии*. 2015. Т. 13, № 3. С. 5–18.
Barakhnin V. B., Kozhemyakina O. Yu., Zabaykin A. V., Khayatova V. D. Automation of the complex analysis of Russian poetic texts: models and algorithms. *Vestnik NSU. Series: Information Technologies*, 2015, vol. 13, no. 3, p. 5–18. (in Russ.)
3. **Козьмин А. В.** Автоматический анализ стиха в системе Starling // *Тр. Междунар. конф. «Компьютерная лингвистика и интеллектуальные технологии»*. М.: Изд. центр РГГУ, 2006. С. 265–268.
Kozmin A. V. Automatic analysis of verse into the Starling system. In: *Proc. International conference “Computational linguistics and intellectual technologies”*, 2006, p. 265–268. (in Russ.)
4. **Бойков Н. В., Каряева М. С., Соколов В. А. и др.** Об автоматической спецификации стиха в информационно-аналитической системе // *Тр. XVII Междунар. конф. «Аналитика и управление данными в областях с интенсивным использованием данных»*. Обнинск: ИАТЭ НИЯУ МИФИ, 13–16 окт. 2015 г., С. 144–151.
Boikov V. N., Karyeva M. S., Sokolov V. A., Pilshchikov I. A. On an Automatic Procedure for the Specification of a Poetic Text for an Open Information-Analytical System. In: *Trudy*

¹⁰ Электронная библиотека ИРЛИ РАН. URL: <http://lib.pushkinskiydom.ru>.

- XVII Mezhdunarodnoi konferencii DAMDID/RCDL'2015 [Proc. XVII Int. Conf. "Data Analytics and Management in Data Intensive Domains"], 2015, p. 144–151. (in Russ.)
5. **Пильщиков И. А., Старостин А. С.** Проблемы автоматизации базовых процедур ритмико-синтаксического анализа силлабо-тонических текстов // Национальный корпус русского языка: 2006–2008: Новые результаты и перспективы. СПб., 2009. С. 298–315.
Pilshchikov I. A., Starostin A. S. Problems of automatization for basic procedures of rhythmic and syntactic analysis of sillabo-tonic texts. In: Russian National Corpus: 2006–2008: New results and prospects. St. Petersburg, 2009, p. 298–315. (in Russ.)
 6. **Пильщиков И. А., Старостин А. С.** Автоматическое распознавание стихотворных размеров: теория и практика // Поэтика и фоностилика: Бриковский сборник. М., 2010. Вып. 1: Материалы Междунар. науч. конф. «I Бриковские чтения: Поэтика и фоностилика» (Москва, 10–12 февраля 2010 года). С. 41–49.
Pilshchikov I. A., Starostin A. S. Automatic recognition of verse meters: theory and practice. Brikovskii sbornik. Vyp. 1: Materialy Mezhdunar. nauch. konf. «I Brikovskie chteniya: Poetika i fonostilistika» [Brik's Collection, vol. 1: Proc. Int. Conf. "Brik's reading: Poetics and phonostilistics"]. Moscow, 2010, p. 41–49. (in Russ.)
 7. **Pilshchikov I., Starostin A.** Automated Analysis of Poetic Texts and the Problem of Verse Meter. *Current Trends in Metrical Analysis. Littera: Studies in Language and Literature*, 2011, p. 133–140.
 8. **Барахнин В. Б., Кожемякина О. Ю., Пастушков И. С.** Сравнительный анализ методов автоматической классификации поэтических текстов на основе лексических признаков // CEUR Workshop Proceedings. 2017. Т. 2022. С. 252–257.
Barakhnin V. B., Kozhemyakina O. Yu., Pastushkov I. S. Comparative analysis of methods of automated classification of poetic texts based on lexical signs. *CEUR Workshop Proceedings*, 2017, vol. 2022, p. 252–257. (in Russ.)
 9. **Bulygin M. V., Sharoff S. A.** Using Machine Translation for Automatic Genre Classification in Arabic. In: Proc. International conference "Computational linguistics and intellectual technologies", 2006, p. 153–162.
 10. **Loukachevitch N. V., Rusnachenko N.** Extracting Sentiment Attitudes from Analytical Texts. In: Proc. International conference "Computational linguistics and intellectual technologies", 2018, p. 459–468.
 11. **Delmonte R.** Computing poetry style. *CEUR Workshop Proceedings*, 2013, vol. 1096, p. 148–155.
 12. **Bacalu C., Delmonte R.** Prosodic Modeling for Speech Recognition. In: Proc. Workshop AI*IA, 1999, p. 45–55.
 13. **Bobenhausen K., Hammerich K.** Literary metrics, Linguistic metrics, and the algorithmic analysis of German poetry using Metricalizer. *Languages*, 2015, vol. 199, no. 3, p. 67–87.
 14. **Barakhnin V. B., Kozhemyakina O. Yu., Rychkova E. V., Gladkikh A. S., Pastushkov I. S.** Software for learning to solve problems of classification using of machine learning. In: European Proc. of Social & Behavioural Sciences, 2018, p. 106–112. DOI 10.15405/epsbs.2018.11.02.12.

Материал поступил в редколлегию
Received
22.10.2018

Сведения об авторах / Information about the Authors

Барахнин Владимир Борисович, доктор технических наук, доцент, ведущий научный сотрудник, Институт вычислительных технологий СО РАН (пр. Академика Лаврентьева, 6, Новосибирск, 630090, Россия), профессор, факультет информационных технологий, Новосибирский государственный университет (ул. Пирогова, 1, Новосибирск, 630090, Россия)

Vladimir B. Barakhnin, Doctor of Engineering Sciences, Associate Professor, Leader Researcher, Institute of Computational Technologies SB RAS (6 Academician Lavrentiev Ave., Novosibirsk, 630090, Russian Federation); Professor, Novosibirsk State University (1 Pirogov Str., Novosibirsk, 630090, Russian Federation)

bar@ict.nsc.ru

ORCID 0000-0003-3299-0507

Кожемякина Ольга Юрьевна, кандидат филологических наук, старший научный сотрудник, Институт вычислительных технологий СО РАН (пр. Академика Лаврентьева, 6, Новосибирск, 630090, Россия)

Olga Yu. Kozhemyakina, Candidate of Philology, Senior Researcher, Institute of Computational Technologies SB RAS (6 Academician Lavrentiev Ave., Novosibirsk, 630090, Russian Federation)

olgakozhemyakina@mail.ru

ORCID 0000-0003-3619-1120

Борзилова Юлия Сергеевна, аспирант, Институт вычислительных технологий СО РАН (пр. Академика Лаврентьева, 6, Новосибирск, 630090, Россия)

Yulia S. Borzilova, Graduate Student, Institute of Computational Technologies SB RAS (6 Academician Lavrentiev Ave., Novosibirsk, 630090, Russian Federation)

i.borzilova@alumni.nsu.ru

ORCID 0000-0002-8265-9356