



Программная система комплексного анализа русских поэтических текстов

Создана программная система комплексного анализа русских поэтических текстов, предназначенная для автоматизированного извлечения из поэтических текстов их характеристик, относящихся к различным структурным уровням (метрика, ритмика, фонетика, лексика, грамматика, литературный стиль, тематика, литературный жанр), и исследования их взаимозависимости. Программная система позволяет освободить исследователей-стиховедов от рутинной работы, расширить круг анализируемых произведений и получать новые научные результаты, связанные с исследованием взаимозависимостей между структурами различных уровней русских поэтических текстов.

Программная система комплексного анализа русских поэтических текстов находится в открытом доступе в сети Интернет по адресу <http://poem.ict.nsc.ru/system/>

Болдинская осень
А. С. Пушкин

Бесы
 Безумных лет угасшее веселье...
 Труд
 Ответ анониму
 Царскосельская статуя
 к переводу "Илиады"
 Румяный критик мой, насмешник толстопузый
 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 (АВАВ);
Количество слогов в строках	
Жанр	элегия
Образец (А - женские окончания; а - мужские)	

Метро-ритмическая статистика:

	развернуть
Размерность	хорей
Стопность	
Число строк	32

Фрагмент интерфейса системы

Публикации:

- Barakhnin V., Kozhemyakina O., Grigorieva I. Determination of the features of the author's style of A.S. Pushkin's poems by machine learning methods // Applied Sciences (Switzerland). 2022. Vol. 2. Iss. 1674.
- Kozhemyakina O.Yu. Conceptual design of the software system for automated complex analysis of poetic texts // Вычислительные технологии. 2022.– Т.27.– № 2.–С. 122-137.
- Barakhnin V., Kozhemyakina O., Revun A., Shashok N. Technological features of cross-language migration from PHP to Python of software products working with intensive data // CEUR Workshop Proceedings. 2021. V. 3036. P. 411-421.

Программная система комплексного анализа русских поэтических текстов

АВТОРЫ: к.филол.н. О.Ю.Кожемякина, д.т.н. В.Б.Баракнин, Н.А.Шашок, И.В.Кузнецова

Современный подход к исследованию текстовых сообщений предполагает использование многоуровневой модели информации. В художественных текстах, прежде всего, поэтических, уровни структуры произвольного сообщения имеют устойчивую иерархию: метрика, ритмика, фонетика, лексика, грамматика, литературный стиль, тематика, литературный жанр. Процесс анализа каждого поэтического текста состоит из рассмотрения структурных уровней текста как самостоятельных смысловых единиц с последующим связыванием полученных данных с другими элементами структуры: лексика формирует семантику данного конкретного стихотворения, метрика является общим фоном семантической традиции, на которой оно воспринимается.

Одной из основных трудностей, возникающих при решении задач анализа поэтических текстов, является необходимость анализа корпуса поэтических текстов большого объема. Эта задача чрезвычайно трудоемка, поэтому, как правило, в поле зрения исследователя попадает лишь сравнительно небольшой круг стихотворений поэтов классического периода, что существенно снижает полноту анализируемого материала и, следовательно, достоверность полученных результатов. В процессе решения указанной проблемы были получены следующие научные результаты:

— Впервые сформулированы и обоснованы принципы создания программных систем автоматизированного многоуровневого анализа поэтических текстов, предназначенных для хранения поэтических текстов и их представления посредством широкого спектра извлеченных из них метаданных, что позволяет в автоматизированном режиме генерировать словари рифм, метроритмические справочники, словари языка поэтов, конкордансы.

— Обосновано соответствие между уровнями многоуровневой модели информации и уровнями структуры поэтического текста, отличающееся учетом специфики классического русского стихосложения, а также обеспечивающее возможность применения информационных технологий и математических методов в задаче автоматизированного комплексного анализа поэтических текстов.

— Разработана структура описания русского поэтического текста, отличающаяся многоуровневостью и наличием категориальных связей между уровнями, обеспечивающая системный анализ обрабатываемого текста и выявление взаимозависимостей его уровней.

— Разработаны алгоритмы автоматизированного извлечения жанровых и стилевых характеристик, впервые основанные на методах машинного обучения и позволившие обеспечить точность определения жанровых и стилевых характеристик, превышающую 95%.

— Предложены алгоритмы фонетического анализа стихотворно-текстовой информации, отличающиеся использованием информации о морфологической структуре анализируемых лексем, обеспечивающие перевод поэтических текстов из дореформенной орфографии в современную с учетом сохранения фонетических свойств структуры текста.

— Исследованы принципы автоматизированного анализа метроритмических и строфических характеристик текста (фактур), впервые основанные на единстве метроритмики и строфики, обеспечивающие цельное представление о нижних уровнях поэтического текста.

— Исследованы принципы и разработаны алгоритмы составления конкордансов и справочников, извлечения словосочетаний, составления словарей языка поэтов, автоматизирующие снятие проблемы омографии, что обеспечивает семантическую точность составляемых словарей.

На основе перечисленных моделей и алгоритмов была впервые создана программная система комплексного анализа русских поэтических текстов, предназначенная для автоматизированного извлечения из поэтических текстов их характеристик, относящихся к различным структурным

уровням, и исследования их взаимозависимости. Программная система находится в открытом доступе в сети Интернет по адресу <http://poem.ict.nsc.ru/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 (АВАВ);
Количество слогов в строках	
Жанр	элегия
Образец (А - женские окончания; а - мужские)	

Метро-ритмическая статистика:

развернуть	
Размерность	хорей
Стопность	
Число строк	32

Разработанная программная система позволяет освободить исследователей-стихovedов от рутинной работы, расширить круг анализируемых произведений и получать новые научные результаты, связанные с исследованием взаимозависимостей различных структурных уровней русских поэтических текстов.

ПУБЛИКАЦИИ

Статьи (WoS, Scopus):

Barakhnin V., Kozhemyakina O., Grigorieva I. Determination of the features of the author's style of A.S. Pushkin's poems by machine learning methods // Applied Sciences (Switzerland). – 2022. – Vol. 12. – Iss. 1674. – <https://doi.org/10.3390/app12031674> (Q2 WoS, Q2 Scopus)

Kozhemyakina O.Yu. Conceptual design of the software system for automated complex analysis of poetic texts // Вычислительные технологии. – 2022. – Т.27. – № 2. – С. 122-137. – <https://doi.org/10.25743/ICT.2022.27.2.010> (Scopus)

Barakhnin V., Kozhemyakina O., Revun A., Shashok N. Technological features of cross-language migration from PHP to Python of software products working with intensive data // CEUR Workshop Proceedings. – 2021. – V. 3036. – P. 411-421. (Scopus)

Баракхнин В.Б., Кожемякина О.Ю., Кузнецова И.В., Карнова В.А. Модель фактуры русских поэтических текстов // Вычислительные технологии. – 2021. – Т. 26. – № 3. – С. 107-117. – <https://doi.org/10.25743/ICT.2021.26.3.007> (Scopus)

Barakhnin V., Kozhemyakina O., Pastushkov I., Kuznetsova I., Borzilova Y. Improvement of the algorithm of automated definition of rhyme // CEUR Workshop Proceedings. - 2020. - Vol.2667. - P.36-41. - ISSN 1613-0073. (Scopus)

Barakhnin V.B., Kozhemyakina O.Yu., Mukhamediev R.I., Borzilova Yu.S., Yakunin K.O. The design of the structure of the software system for processing text document corpus // *Business Informatics*. – 2019. – Vol.13. – N 4. – P. 60–72. – <http://dx.doi.org/10.17323/1998-0663.2019.4.60.72> (WoS, Scopus)

Barakhnin V.B., Kozhemyakina O.Yu., Kuznetsova I.V. Development and implementation of the algorithm for automatic analysis of metrorhythmic characteristics of Russian poetic texts // *CEUR Workshop Proceedings*. – 2019. – V. 2523. – P. 290-298. (Scopus)

Диссертации:

Кожемякина О.Ю. Программная система комплексного анализа русских поэтических текстов: модели и алгоритмы // Дис. ... доктора технич. наук. – Новосибирск, 2022. – 288 с.

Свидетельства о регистрации программ и баз данных:

Барахнин В.Б., Кожемякина О.Ю., Карпова В.А., Кузнецова И.В. Программа определения фактур русских поэтических текстов // Свидетельство о государственной регистрации программы для ЭВМ № 2021664226 от 01.09.2021 г. – Заявка № 20216163325 от 24.08.2021 г. (Федеральная служба по интеллектуальной собственности). – <https://new.fips.ru/ofpstorage/Doc/PrEVM/RUNWPR/000/002/021/664/226/2021664226-00001/document.pdf>

Барахнин В.Б., Кожемякина О.Ю., Кузнецова И.В. Программа автоматического анализа метроритмических характеристик поэтических текстов на русском языке // Свидетельство о государственной регистрации программы для ЭВМ № 2020663072 от 22.10.2020 г. – Заявка № 2020662151 от 12.10.2020 г. (Федеральная служба по интеллектуальной собственности). – <https://www1.fips.ru/ofpstorage/Doc/PrEVM/RUNWPR/000/002/020/663/072/2020663072-00001/document.pdf>

Барахнин В.Б., Кожемякина О.Ю., Борзилова Ю.С. Русские поэтические тексты и их комплексные стиховедческие метаданные // Свидетельство о государственной регистрации базы данных № 2020621889 от 15.10.2020 г. – Заявка № 2020621862 от 28.09.2020 г. (Федеральная служба по интеллектуальной собственности). – <https://www.fips.ru/ofpstorage/Doc/PrEVM/RUNWDB/000/002/020/621/889/2020621889-00001/document.pdf>