

Проект, поддержанный Российским фондом фундаментальных исследований (РФФИ) №19-07-00243 А
Создание методов автоматизированной обработки текстов с целью авторизации, анализа и построения ритмичных текстов на русском, французском и английском языках

Информация о составе проекта



Руководитель проекта

Бойчук Елена Игоревна, доктор филологических наук, доцент кафедры романских языков факультета иностранных языков Ярославского государственного педагогического университета им. К.Д. Ушинского. Область научных интересов: литературоведение, лингвистика текста, стилистика, функциональная грамматика, теория языка и сравнительно-сопоставительные исследования. Опубликовано более 90 трудов в области ритмики текста, литературоведения и компьютерной лингвистики.



Исполнители проекта

Лагутина Надежда Станиславовна, кандидат физ.-мат. наук, доцент кафедры вычислительных и программных систем Ярославского государственного университета им. П.Г. Демидова. Научные интересы: компьютерная лингвистика, автоматическая обработка текста, машинное обучение.



Лагутина Ксения Владимировна, аспирант кафедры теоретической информатики, преподаватель кафедры вычислительных и программных систем Ярославского государственного университета им. П.Г. Демидова. Область научных интересов: компьютерная лингвистика и обработка текстов. Опубликовано 27 статей в области компьютерной лингвистики. Награждена медалью Российской Академии наук в области компьютерных наук.



Воронцова Инна Алексеевна, кандидат филологических наук, доцент кафедры теории и практики перевода Ярославского государственного педагогического университета им. К.Д. Ушинского. Научные интересы: лексикология, лексикография, теория и практика перевода, теория текста и теория дискурса.



Шляхтина Елена Васильевна, кандидат филологических наук, доцент кафедры теории и практики перевода Ярославского государственного педагогического университета им. К.Д. Ушинского. Научные интересы: языковая вежливость, эвфемизация речи, политкорректность, перевод, художественный текст, ритм текста.



Мишенькина Елена Владимировна, кандидат филологических наук, доцент кафедры теории и практики перевода Ярославского государственного педагогического университета им. К.Д. Ушинского. Область научных интересов — этнопсихолингвистика, межкультурная коммуникация, перевод, компьютерная лингвистика и обработка текстов. Опубликовано более 40 статей и 5 учебных пособий.



Беляева Ольга Васильевна, ассистент кафедры теории и практики перевода факультета иностранных языков Ярославского государственного педагогического университета им. К.Д. Ушинского; член совета ЯРОО "Ассоциация учителей английского языка", региональный тренер Британского Совета (2011-2015). Научные интересы: корпусная лингвистика, разработка учебных материалов, подготовка к международным экзаменам.

О проекте

Основной научной задачей, на решение которой направлено данное исследование, является создание новых методов в компьютерной лингвистике для автоматизированной обработки текстов на русском, английском и французском языках с целью их авторизации, анализа и построения ритмичных текстов на этих языках. Данная задача интегрирует компьютерную лингвистику и информационные технологии с такими областями филологического знания, как лингвистика текста, текстология, стилистика текста, риторика, психолингвистика, культура речи, а также с количественной лингвистикой, суггестивной лингвистикой, лингвокультурологией. основополагающим параметром создания данных методов является ритм текста.

Актуальность данной задачи состоит в том, что большое количество существующих теоретических исследований в области ритма, который требует статистического подхода, применения методов количественного анализа, автоматизированной обработки текста, не имеет выхода к осуществлению такого рода анализа. Анализ ритма осуществляется вручную, что снижает эффективность текстовой обработки и не всегда приводит к точным результатам.

В основу предложенных исследователями современных программных приложений для русского языка положены такие факторы определения ритмичности текста, как размер предложений и среднеарифметический показатель безударных слогов (Белоусов, К.И., Дусакова Г.Ф., Леонов Д.В., 2017; Кишалова Л.В., 2017; Андрусенко Т.В., 2017). Для английского и французского языков существуют приложения, позволяющие анализировать текст с точки зрения его лексического состава, размера предложений и программ (Text Analyzer, Sketch Engine, Stylo, Word2vec и др.).

Анализ ритма в рамках данного проекта направлен на определение авторства текстов, сопоставление переводов художественных произведений, а также на анализ и составление суггестивных текстов на основе определения их ритмических характеристик, складывающихся из употребления средств ритмизации, проявляющихся на фонетическом, лексико-грамматическом и структурно-композиционном уровнях. В состав данных средств входят стилистические средства выразительности речи, в основу выделения которых заложена повторяемость элемента (рифма, ассонанс, аллитерация, ономотопея, анаграмма, параномазия, таутизм, деривация, полиптотон, анафора, эпифора, симплока, хиазм, однородные члены, эпаналепсис, анадиплосис, дубликация и др.).

Для реализации цели проекта необходимо создать автоматизированные алгоритмы, позволяющие проводить сравнительно-сопоставительное исследование текстов с точки зрения ритмических характеристик; определить статистические характеристики ритма текста и создать банк данных, включающий тексты с выявленными характеристиками ритма; разработать алгоритм анализа и составления ритмичного текста с целью активного суггестивного воздействия на реципиента; создать метод атрибуции на основе ритмических средств текста и программный прототип, позволяющий хранить и обрабатывать тексты, проводить сравнительно-сопоставительный анализ ритма в переводе и осуществлять атрибуцию.

В данном исследовании впервые предлагается использование перечисленных выше ритмических средств в качестве основы создания автоматизированного метода атрибуции текстов. Разработанные в настоящее время алгоритмы составления ритмичных текстов основаны на абзацном членении и размере предложения. Предлагаемый метод позволит расширить возможности суггестивного воздействия на реципиента при помощи представленных ритмических характеристик. Новизна предлагаемого метода заключается также в том, что он позволяет осуществлять ритмический анализ текстов, оценивать адекватность перевода с точки зрения ритмических

характеристик, осуществлять атрибуцию текстов на трех языках русском, английском и французском. В разработке задействованы новые методы: ритмический анализ текста, сравнительно-сопоставительный ритмический анализ, которые позволят получить более точные результаты обработки ритма текста.

Автоматизированная обработка текста с точки зрения ритмических характеристик позволит более успешно реализовывать задачу авторизации. Областью практического применения автоматизированной обработки текста с позиций реализации ритмических характеристик может являться публицистика, в рамках которой эффективным с точки зрения воздействия на реципиента является создание суггестивных текстов.

Основным результатом проекта должно стать создание программного прототипа, включающего банк текстов на трех языках (корпус электронных художественных текстов на русском, английском и французском языках для осуществления анализа ритма и корпус художественных текстов с выявленными статистическими характеристиками ритма), позволяющего осуществлять автоматизированный анализ ритма, авторизацию текста и проводить сравнительно-сопоставительный анализ переводов текстов.

Публикации по проекту

Монографии

1. Бойчук Е. И. Анализ ритма прозы (на материале французского языка): монография / Е. И. Бойчук. – Ярославль: Канцлер, 2019. – 232 с.
2. Идиостиль и ритм текста: коллективная монография / Бойчук Е.И., Воронцова И.А., Шляхтина Е.В., Беляева О.В., Ярославль: РИО ЯГПУ, 2019. 183 с.
3. Бойчук Е.И. «La France est en guerre»: анализ политического дискурса Э. Макрона в период пандемии коронавирусной инфекции COVID-19: коллективная монография «Антропоцентрическая направленность лингвистических исследований поликультурного военного и политического дискурса». Ярославль, ООО «Цифровая типография», 2020. С. 133-151.
4. Автоматизированный анализ текстовых характеристик. Учебное пособие / Бойчук Е.И., Лагутина К.В., Лагутина Н.С., Воронцова И.А., Шляхтина Е.В., Мишенькина Е.В., Беляева О.В.. - Ярославль: Канцлер. 2020. 88с.

Публикации в SCOPUS и Web of Science:

1. Lagutina K., Lagutina N., Boychuk E., Vorontsova I., Shliakhtina E., Belyaeva O., Paramonov I. A Survey on Stylometric Text Features // 26 Conference of Open Innovation Association FRUCT: Helsinki, Finland, 5-8 November 2019. P. 184-195.
<https://fruct.org/publications/fruct25/files/Lag.pdf>

2. Boychuk E., Vorontsova I., Shliakhtina E., Lagutina K., Belyaeva O. Automated Approach to Rhythm Figures Search in English Text // AIST, 2019. P. 105-116. <https://ieeexplore.ieee.org/document/6872425>
3. Lagutina K., Lagutina N., Boychuk E. Paramonov I. The Influence of Different Stylometric Features on the Classification of Prose by Centuries // 27 Conference of Open Innovation Association FRUCT: Trento, Italy, 2 October, 2020. P. 108-115. [10.23919/FRUCT49677.2020.9211036](https://www.elibrary.ru/item.asp?id=43968534)
<https://www.elibrary.ru/item.asp?id=43968534>
4. Lagutina K., Poletaev A., Lagutina N., Boychuk E., Paramonov I. Automatic Extraction of Rhythm Figures and Analysis of Their Dynamics in Prose of 19th-21st Centuries // 28 Conference of Open Innovation Association FRUCT: Russia, Yaroslavl, 22-24 April, 2020. P. 247-255. <https://www.elibrary.ru/item.asp?id=42830737>
5. Boychuk E.I., Belyaeva O.V. Multilayeredness Of Literary Prosaic Text: Polyphony Vs Homophony // Conference TILTM 2020 Topical Issues of Linguistics and Teaching Methods in Business and Professional Communication 24-25 april 2020. RUDN University p. 211-223. https://www.europeanproceedings.com/files/data/article/10028/12632/article_10028_12632_pdf_100.pdf DOI: 10.15405/epsbs.2020.12.02.30
6. Vorontsova I.A. Analysis Of Rhythm Reproduction In English-Russian Literary Translation // Conference TILTM 2020 Topical Issues of Linguistics and Teaching Methods in Business and Professional Communication 24-25 april 2020. RUDN University. P. 508-515. https://www.europeanproceedings.com/files/data/article/10028/12671/article_10028_12671_pdf_100.pdf DOI: 10.15405/epsbs.2020.12.02.68
7. Mishenkina E.V. Specifics Of Translating Color Names In English Texts // Conference TILTM 2020 Topical Issues of Linguistics and Teaching Methods in Business and Professional Communication 24-25 april 2020. RUDN University p. 286-295. https://www.europeanproceedings.com/files/data/article/10028/12641/article_10028_12641_pdf_100.pdf DOI: 10.15405/epsbs.2020.12.02.39
8. Shliakhtina E.V. Grammar Rhythm Figures In Literary Prose: Case study Of S.Thomas's Popco // Conference TILTM 2020 Topical Issues of Linguistics and Teaching Methods in Business and Professional Communication 24-25 april 2020. RUDN University p. 382-388. https://www.europeanproceedings.com/files/data/article/10028/12655/article_10028_12655_pdf_100.pdf DOI: 10.15405/epsbs.2020.12.02.52
9. Boychuk E., Lagutina K., Vorontsova I., Mishenkina E., Belyaeva O. Evaluating The Performance Of A New Text Rhythm Analysis Tool // English Studies at NBU. Vol.6, Issue 2. pp.217-232. DOI: <http://doi.org/10.33919/esnbu.20.2.3>
10. N. S. Lagutina, K. V. Lagutina, E. I. Boichuk, I. A. Vorontsova, and I. V. Paramonov, Automated Rhythmic Device Search in Literary Texts Applied to Comparing Original and Translated Texts as Exemplified by English to

Публикации в журналах, рецензируемых ВАК:

1. Лагутина Н.С., Лагутина К.В., Бойчук Е.И., Воронцова И.А., Парамонов И.В. Автоматизированный поиск средств ритмизации художественного текста для сравнительного анализа оригинала и перевода на материале английского и русского языков // Моделирование и анализ информационных систем. Т. 26, №3 (2019), с. 420–440. DOI: 10.18255/1818-1015-2019-3-420-440
2. Бойчук Е.И., Джонсон М.А. Специфика ритмической структуры испанских прозаических текстов // Верхневолжский филологический вестник= VerhnevolzhskiPhilologicalBulletin:научный журнал.– Ярославль: РИО ЯГПУ, 2019.–№2(17)– С. 119-129. DOI: [10.24411/2499-9679-2019-10322](https://doi.org/10.24411/2499-9679-2019-10322)
3. Бойчук Е.И. Количественные наречия как средство авторизации художественных текстов (на материале французской прозы 19в.) // Верхневолжский филологический вестник= VerhnevolzhskiPhilologicalBulletin:научный журнал.–Ярославль: РИО ЯГПУ, 2019.–№1(16)– С. 123-130. DOI: [10.24411/2499-9679-2019-10394](https://doi.org/10.24411/2499-9679-2019-10394) <https://www.elibrary.ru/item.asp?id=37166479>
4. Шляхтина Е.В. Передача ритмических средств при переводе художественной прозы с английского на русский язык (на материале романа К. Аткинсон «Жизнь после жизни») // Вестник МГОУ: Серия Лингвистика, 2019. С. 149-161. <https://vestnik-mgou.ru/Articles/View/13406>
5. Лагутина К., Манахова А. Automated Search and Analysis of the Stylometric Features that Describe the Style of the Prose 19th-21st Centuries // Моделирование и анализ информационных систем, Т. 27, № 3, 2020. С. 330-343. file:///C:/Users/9789~1/AppData/Local/Temp/1352-2225-1-SM.pdf
6. Бойчук Е.И. Джонсон М.А. Испанские количественные наречия в системе элементов авторского идиолекта (на материале романов 19-20 вв.) // Филологические науки. Вопросы теории и практики. Тамбов: Грамота, 2020. № 5. С. 245-250. <https://www.elibrary.ru/item.asp?id=42882614>
7. Певтиев В.И., Бойчук Е.И. Специфика идиолекта Э. Макрона в контексте его политической деятельности // Верхневолжский филологический вестник, № 3, 2020. DOI: [10.20323/2499-9679-2020-3-22-103-111](https://doi.org/10.20323/2499-9679-2020-3-22-103-111) С. 104-113. <https://www.elibrary.ru/item.asp?id=44094724>
8. Бойчук Е.И., Мишенькина Е.В. Автоматизированный анализ ритмических характеристик русскоязычного художественного текста как метод установления принадлежности определенной литературной

- эпохе // Вестник Череповецкого государственного университета. №6. (99). – С. 73–85. DOI: 10.23859/1994-0637-2020-6-99-6
9. Воронцова И.А. Прецедентность как маркер авторского стиля // Верхневолжский филологический вестник, № 2. 2020. С.123-130. <https://vv.yspu.org/publication/2-2020/>
10. Шляхтина Е.В. Редупликация в художественном тексте (на материале произведений британских и американских писателей) // Верхневолжский филологический вестник, № 4. 2020. С.142-150. <https://vv.yspu.org/wp-content/uploads/sites/4/2020/12/VFV-МАКЕТ-%E2%84%964-2020-V-РЕСНАТ.pdf>
11. Бойчук Е.И., Богданова Т.Ф. Основные теоретические проблемы, связанные с понятием тональности текста // Верхневолжский филологический вестник, № 4. 2020. С. 136-141. <https://vv.yspu.org/wp-content/uploads/sites/4/2020/12/VFV-МАКЕТ-%E2%84%964-2020-V-РЕСНАТ.pdf>

Публикации в сборниках РИНЦ и тезисы докладов на конференциях:

1. Бойчук Е.И. К проблеме автоматизированной обработки ритма текста // Романские языки в синхронии и диахронии : межвузовский сборник научных трудов. - М.: ИИУ МГОУ, 2019. - С. 52-59. <https://elibrary.ru/item.asp?id=38528201>
2. Бойчук Е.И., Трофимова Е.А. Лексико-грамматический аспект реализации ритма романа Шарлотты Бронте «Джейн Эйр» // Язык и общество: Диалог культур и традиций: сборник материалов международной научной конференции «Чтения Ушинского». – Вып. 17. – Ярославль: РИО ЯГПУ, 2019. - С. 40-52. http://yspu.org/images/e/eb/%D0%91%D0%BE%D0%B9%D1%87%D1%83%D0%BA_%D0%A2%D1%80%D0%BE%D1%84%D0%B8%D0%BC%D0%BE%D0%B2%D0%B0.pdf
3. Бойчук Е.И. Особенности лексико-грамматического аспекта ритмизации романа А. Гавальда «Просто вместе» // Материалы Всероссийской научной конференции с международным участием «Европейское литературное наследие в кросскультурном пространстве». – Ярославль: Канцлер, 2019. - С. 199-204. <https://elibrary.ru/item.asp?id=37415517>
4. Бойчук Е.И. Специфика ритма современной французской прозы (на материале романа А. Нотомб «Кодекс принца») // Материалы конференции Филологические чтения: Человек. Текст. Дискурс. - Изд. Ярославский государственный университет им П.Г. Демидова, 2019г. - С.6-13. <https://www.elibrary.ru/item.asp?id=39135521>
5. Boychuk E., Belyaeva O. La technique de stylométrie réalisée à la base de l'analyse informatique du rythme du texte // 10-ièmes Journées Internationales de Linguistique de Corpus (JLC). Université Grenoble-

- Alpes, 26-28.11.19. P. 163-167.
https://jlc2019.sciencesconf.org/data/pages/JLC_Xe.pdf
6. Лагутина Н.С., Лагутина К.В. Обзор инструментов для анализа ритма текста // Заметки по информатике и математике, Вып. 11. - Ярославль, ЯрГУ, 2019. С. 100-107. <https://www.elibrary.ru/item.asp?id=39160332>
 7. Туманова А.Д., Лагутина Н.С. Алгоритмы автоматического поиска лексических аспектов анализа ритма Семьдесят вторая всероссийская научно-техническая конференция студентов, магистрантов и аспирантов высших учебных заведений с международным участием. Ярославль: ЯГТУ. С. 212-215. <https://www.ystu.ru/72%20%D0%9D%D0%A2%D0%9A%20%D1%82%D0%BE%D0%BC%203.pdf>
 8. Ратников Е.С., Лагутина Н.С. Программное приложение для автоматического определения лексических аспектов ритмики текста // Семьдесят вторая всероссийская научно-техническая конференция студентов, магистрантов и аспирантов высших учебных заведений с международным участием. Ярославль: ЯГТУ. С. 169-172. <https://www.ystu.ru/72%20%D0%9D%D0%A2%D0%9A%20%D1%82%D0%BE%D0%BC%203.pdf>
 9. Шляхтина Е. В., Туленцова Е. А. Особенности перевода художественного текста в жанре «хоррор» с английского на русский язык (на материале Р. Л. Стайна «Мечь садовых гномов») // Язык и общество: диалог культур и традиций. Сборник материалов 73-й международной научной конференции «Чтения Ушинского». Вып 17. Ярославль: РИО ЯГПУ, 2019. С. 193-196. <https://www.elibrary.ru/item.asp?id=41482824>
 10. Воронцова И.А., Маценко Л.В. Стратегии и приемы межъязыковой трансляции культурного компонента системы художественного текста (на материале поэмы В.Ерофеева «Москва-Петушки» и его перевода на английский язык) // Европейское литературное наследие в кросскультурном пространстве: сборник материалов Всероссийской научной конференции с международным участием (23-24 мая 2019г.). Москва: Научная библиотека, 2019. С. 99-106.
 11. Лагутина К.В., Манахова А.М. Автоматизированный поиск и сравнение стилометрических характеристик во французских художественных текстах XIX—XXI веков // Заметки по информатике и математике, № 12, 2020, С. 189-197. <https://www.elibrary.ru/item.asp?id=43163415>
 12. Лагутина Н.С., Миронов А.Р. Определение параметров ритмической структуры прозаических текстов на основе чередования ударных и безударных слогов // Заметки по информатике и математике, № 12, 2020, С. 198-204. <https://www.elibrary.ru/item.asp?id=43163416>
 13. Лагутина Н.С., Попов Е.В. Определение стилометрических характеристик коротких текстов и их применение в задачах классификации // Заметки по информатике и математике, № 12, 2020, С. 254-261. <https://www.elibrary.ru/item.asp?id=43163422>

14. Лагутина Н.С., Кузьмина А.Н., Самарина А.А., Автоматизация поиска фонетических аспектов анализа ритма прозаических текстов на французском языке // Семьдесят третья всероссийская научно-техническая конференция студентов, магистрантов и аспирантов высших учебных заведений с международным участием. ЯГТУ. Ч.1., 2020. С. 991-994.
<https://www.ystu.ru/files/nauka/73%20%D0%9D%D0%A2%D0%9A%201%20%D1%87%D0%B0%D1%81%D1%82%D1%8C.pdf>
15. Лагутина К.В., Галкина Д.А. Автоматический поиск аллитерации и ассонанса в испанской прозе. Семьдесят третья всероссийская научно-техническая конференция студентов, магистрантов и аспирантов высших учебных заведений с международным участием. ЯГТУ. Ч.1., 2020. С. 1034-1037.
<https://www.ystu.ru/files/nauka/73%20%D0%9D%D0%A2%D0%9A%201%20%D1%87%D0%B0%D1%81%D1%82%D1%8C.pdf>
16. Бойчук Е.И. Специфика ритма современной французской прозы (на материале романов Ж. Легардинье, М. Леви, Г. Мюссо) // «Романские языки: взаимодействие литературы и культуры народов» 23-24 июня 2020 г. С. 242-253. <https://www.elibrary.ru/item.asp?id=43943454>
17. Бойчук Е.И. Многослойность художественного прозаического текста: полифония VS гомофония // Сборник материалов IX Международной научной конференции «Актуальные проблемы лингвистики и лингводидактики иностранного языка делового и профессионального общения». Москва, РУДН, 24–25 сентября 2020 г. С. 259-261. <https://www.elibrary.ru/item.asp?id=44062254>
18. Воронцова И.А. Ритмический анализ текста романа А. Мердок «The Black Prince» и его перевод на русский язык с применением автоматизированного инструмента ProseRhythmDetector // Сборник материалов IX Международной научной конференции «Актуальные проблемы лингвистики и лингводидактики иностранного языка делового и профессионального общения». Москва, РУДН, 24–25 сентября 2020 г. С. 214-216. <https://www.elibrary.ru/item.asp?id=44062254>
19. Шляхтина Е.В. Грамматические средства ритмизации художественного текста (на материале романа с. Томас «Корпорация Попс» и его перевода на русский язык) // Сборник материалов IX Международной научной конференции «Актуальные проблемы лингвистики и лингводидактики иностранного языка делового и профессионального общения». Москва, РУДН, 24–25 сентября 2020 г. с. 150-152. <https://www.elibrary.ru/item.asp?id=44062254>
20. Мишенькина Е.В. Специфика перевода цветоименований в англоязычных текстах // Сборник материалов IX Международной научной конференции «Актуальные проблемы лингвистики и лингводидактики иностранного языка делового и профессионального общения». Москва, РУДН, 24–25 сентября 2020 г. с. 137-139. <https://www.elibrary.ru/item.asp?id=44062254>

21. Беляева О.В. Современные исследования в области идиолекта британских писателей // Сборник материалов IX Международной научной конференции «Актуальные проблемы лингвистики и лингводидактики иностранного языка делового и профессионального общения». Москва, РУДН, 24–25 сентября 2020 г. С. 201-203. <https://www.elibrary.ru/item.asp?id=44062254>
22. Бойчук Е.И. Обзор корпусов текстов, представленных на конференции Linguistique de Corpus в университете Гренобль-Альпы (Франция), 25-28 ноября, 2019 г. // Язык и общество: диалог культур и традиций. Сборник материалов 74-й международной научной конференции «Чтения Ушинского». Вып 18. Ярославль: РИО ЯГПУ, 2020. С.111-119. <https://www.elibrary.ru/item.asp?id=43946695>
23. Воронцова И.А., Петрова Е.Е. Транслатологические особенности народной сказки (на материале ирландской народной сказки “Teig O’Kane and the Corpse”) // Язык и общество: диалог культур и традиций. Сборник материалов 74-й международной научной конференции «Чтения Ушинского». Вып 18. Ярославль: РИО ЯГПУ, 2020. С. 184-189. <https://www.elibrary.ru/item.asp?id=43946685>
24. Мишенькина Е.В., Черногорова С.Е. Специфика перевода юридической лексики в кинотексте // Язык и общество: диалог культур и традиций. Сборник материалов 74-й международной научной конференции «Чтения Ушинского». Вып 18. Ярославль: РИО ЯГПУ, 2020. С. 203-209. <https://www.elibrary.ru/item.asp?id=43946677>
25. Бойчук Е.И., Лагутина К.В., Лагутина Н.С. Автоматизированный поиск и анализ ритмических характеристик художественных прозаических произведений 19-21 веков // Международный Научный Конгресс по применению современных IT решений в фундаментальных научных исследованиях и промышленных технологиях FIT-M, 17-19 декабря 2020 г. - С. 83-87. https://fit-m.org/uploads/tezis_FIT-M_2020.pdf

Охранные документы на результаты интеллектуальной деятельности

Свидетельство о государственной регистрации программы для ЭВМ № 2019619380 «Программа, реализующая автоматизированный алгоритм анализа ритма текста на основе фонетических, лексико-грамматических и структурно-композиционных параметров ритма для текстов на русском, английском и французском языках». Правообладатели: Ратников Егор Сергеевич, Туманова Анастасия Дмитриевна, Бойчук Елена Игоревна, Лагутина Надежда Станиславовна, Лагутина Ксения Владимировна. Дата регистрации 16.07.2019г.

Свидетельство о государственной регистрации программы для ЭВМ № 2020618659 «Программа для поиска и визуализации ритмических характеристик текста на русском, английском, французском и испанском

языках». Правообладатели: Ларионов Владислав Дмитриевич, Петряков Владислав Леонидович, Полетаев Анатолий Юрьевич, Бойчук Елена Игоревна. Дата регистрации 13.07.2020 г.

Свидетельство о государственной регистрации программы для ЭВМ № 2020618648 «Программа для статистического анализа стилометрических и ритмических характеристик текстов на русском, английском, французском и испанском языках. Правообладатели: Манахова Алла Михайловна, Лагутина Ксения Владимировна, Лагутина Надежда Станиславовна. Дата регистрации 15.07.2020 г.

ProseRhythmDetector (PRD)

Программа предназначена для поиска ритмических характеристик текста, основанных на лексических повторениях, и их визуализации. Она представляет собой десктопное приложение, способное работать с русскими, английскими, французскими и испанскими текстами.

В программу можно загрузить txt-файл в формате UTF-8 с текстом, затем указать язык (русский, английский, французский, испанский). Программа обрабатывает текст: разделяет его на слова и предложения, нумерует слова, находит слова и предложения, в которых имеются ритмические характеристики. Затем она выводит текст на экран, а ритмические характеристики подсвечивает различными цветами. Также она выводит общий список ритмических характеристик.

Найденные ритмические характеристики программа может экспортировать в json-файл.

Программой также осуществляется статистический анализ стилометрических и ритмических характеристик текстов на естественном языке.

В json-файле, в котором содержится текст и номера слов с ритмическими характеристиками, программа считает статистические характеристики для ритмических: количество появлений на 100 предложений, доля уникальных слов, доля разных частей речи и т.п. Для самого текста программа считает стилометрические характеристики: количество предложений и их средний размер, количества наиболее часто встречающихся слов и т. д.

Полученные характеристики сохраняются в виде csv-таблиц (Рис. 1) и визуализируются с помощью графиков (Рис. 2) и тепловых карт (Рис. 3):

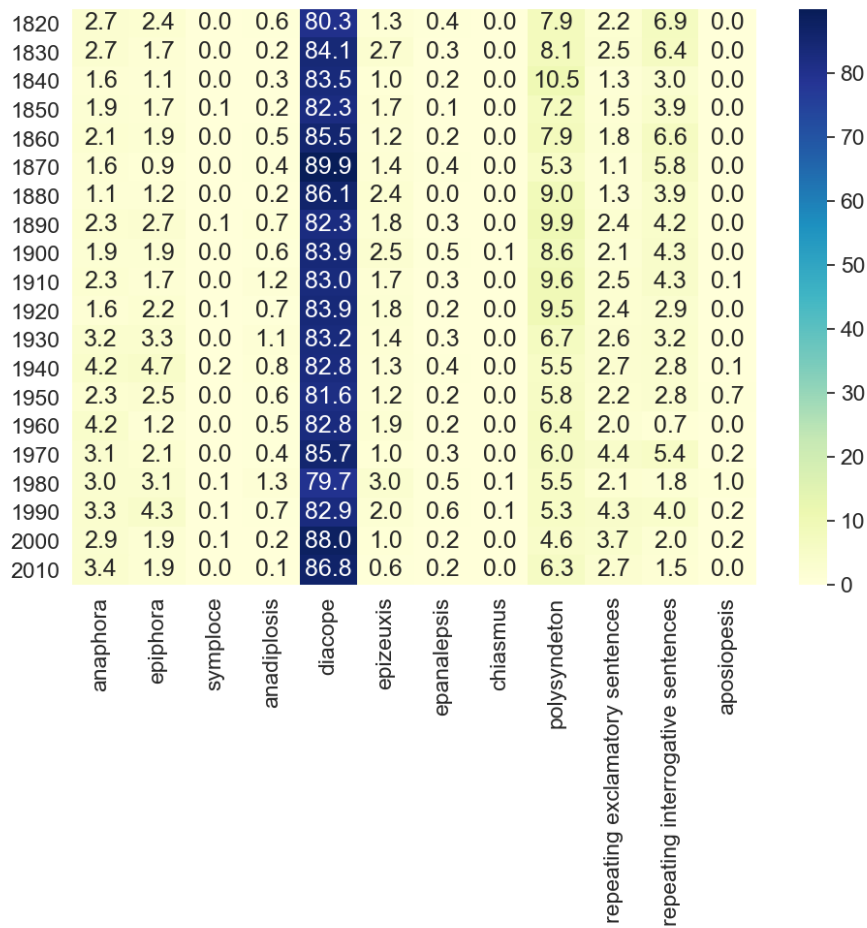


Рис. 1. CSV-таблицы данных по ритмическим средствам.

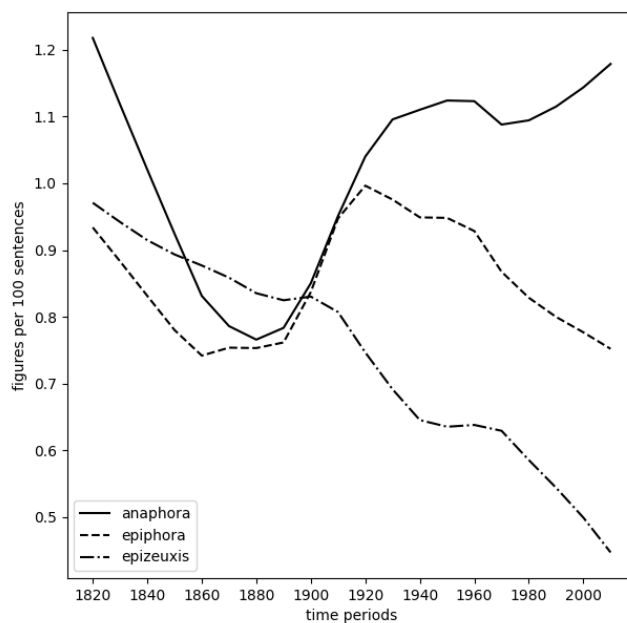


Рис. 2. Графики построения кривой ритма.

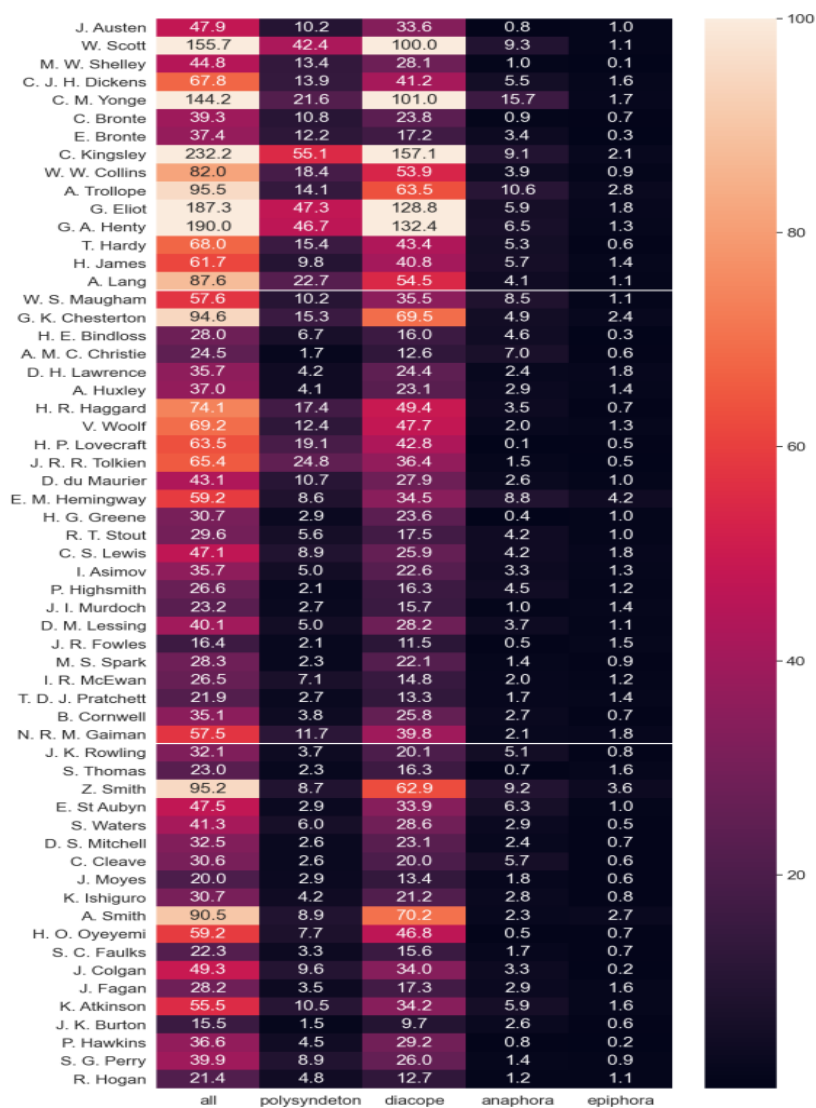


Рис. 3. Тепловые карты ритмических средств по авторам.