

Темпы прироста счета текущих операций у стран бывшего G8 с 1997 по 2023 годы демонстрируют значительные колебания, отражая изменения в глобальной экономике и внутренней политике. В то время как некоторые страны, такие как Германия и Канада, показали стабильный рост, другие, например, Италия и Великобритания, столкнулись с вызовами, препятствующими устойчивому развитию. Общая тенденция указывает на необходимость адаптации экономической политики для устойчивого повышения конкурентоспособности в условиях глобальных изменений.

После проведения корреляционного анализа получены следующие выводы: исходя из значений коэффициента корреляции, сделан вывод о том, что средняя обратная взаимосвязь присутствует только у темпов прироста счета текущих операций Великобритании и Канады; также прямая связь средней силы присутствует у темпов прироста Великобритании и России. У остальных стран связь между темпами прироста счета текущих операций почти полностью отсутствует, что говорит о независимости решений, принимаемых странами в области экспорта и импорта товаров и услуг.

После проведения однофакторного дисперсионного анализа среди стран бывшего G8 по показателю «Счет текущих операций» можно сделать вывод о том, что на уровне значимости 0,05 выдвигаемая гипотеза о равенстве средних значений темпов прироста счета текущих операций для выбранных стран была принята, что говорит об их статистическом равенстве.

Проведённое исследование позволило выявить ключевые тенденции в динамике доли резервных валют (доллара США, евро и йены) и ВВП соответствующих стран в мировой экономике, а также проанализировать счет текущих операций стран G8.

Литература

- 1 Кобзарь, А. И. Прикладная математическая статистика. Для инженеров и научных работников / А. И. Кобзарь. – М. : ФИЗМАТЛИТ, 2006. – 816 с.
- 2 Кремер, Н. Ш. Теория вероятностей и математическая статистика / Н. Ш. Кремер. – М. : ЮНИТИ-ДАНА, 2004. – 573 с.
- 3 Андронов, А. М. Теория вероятностей и математическая статистика : учебник для вузов / А. М. Андронов, Е. А. Копытов, Л. Я. Гринглаз. – СПб. : Питер, 2004. – 461 с.
- 4 Уилкс, С. Математическая статистика / С. Уилкс. – М. : Наука, 1967. – 632 с.
- 5 SMART-LAB: Статистика доллара, евро и йены : [сайт]. – Санкт-Петербург, 2010–2025. – URL: <https://smart-lab.ru/blog/897348.php> (дата обращения: 15.11.2024).
- 6 The Global Economy: Счет текущих операций : [сайт]. – Атланта, 2012–2025. – URL: <https://www.theglobaleconomy.com> (дата обращения: 11.11.2024).

УДК 004.5:519.2:811.161

А. С. Корж

РАЗРАБОТКА ПРИЛОЖЕНИЯ ДЛЯ СТАТИСТИЧЕСКОГО АНАЛИЗА ТЕКСТОВ НА БЕЛОРУССКОМ ЯЗЫКЕ

Статья посвящена разработке приложения на языке программирования Python для статического анализа текстов на белорусском языке. Приложение позволяет проводить частотный анализ букв, буквенных пар, слов, длин слов, предложений, определяет энтропию и информационную избыточность текстов. С помощью приложения на основе статических показателей можно сравнивать произведения на белорусском языке различных авторов.

Сравнение стилей и жанров текстов произведений различных авторов является важной задачей в области компьютерной лингвистики. Статистический подход к анализу текстов позволяет автоматизировать процесс сравнения текстов. Автоматизация анализа текстов с применением различных алгоритмов широко используется для анализа текстов на английском, русском, болгарском, чешском и других языках [1]. Разработаны программные комплексы, которые предназначены для выполнения полного цикла анализа текстов [2], однако инструментов для статистического анализа белорусских текстов нет. Целью работы является разработка приложения на языке программирования Python для статистического анализа текстов на белорусском языке. В работе использовались произведения следующих авторов: Владимира Короткевича, Леонида Дайнеко, Людмилы Рублевской.

Разработана архитектура приложения, проведена демонстрация его работы на основе одного из произведений автора текстов на белорусском языке. Работа приложения осуществляется в три этапа (рисунок 1): передача пользователем текстового файла в приложение; обработка файла, переданного пользователем; передача пользователю обработанных данных.

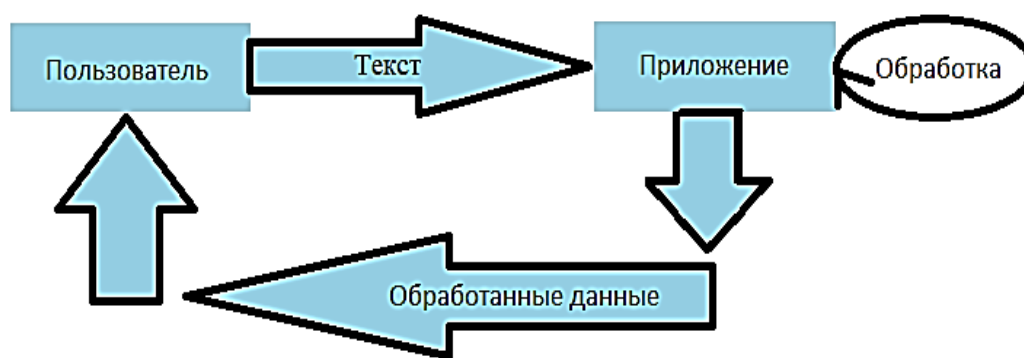


Рисунок 1 – Схема работы программы

Реализация приложения осуществлена на языке программирования Python [3–6]. В первом этапе происходит передача в программу файла с текстом от пользователя. Приложение работает с файлами типа: “pdf”, “docx”, “txt”, “png”, “jpg”, “jpeg”, “tiff”, “bmp”.

Второй этап представляет собой обработку заданного файла: текст преобразовывается в строку и переводится в нижний регистр; из строки удаляются все спецсимволы и переносы; обрабатывается текст, данные передаются пользователю в текстовой и визуализированной форме.

На третьем этапе пользователь принимает обработанную информацию. Пользователь получает график частоты символов белорусского языка, вероятность их встречи, количество гласных, согласных, всего символов (гласные + согласные + апостроф), величина информационной энтропии, величина информационной избыточности, наиболее часто встречаемые пары букв, наиболее часто встречаемые пары букв, где первая буква выбрана пользователем, наиболее часто встречаемые пары букв, где вторая буква выбрана пользователем, коммутативность выбранной буквы, абсолютная коммутативность выбранной буквы, сколько всего расстояний от выбранной буквы до следующей встречи с ней, минимальное, максимальное, среднее расстояние, общее количество слов, слово минимальной, максимальной длины, минимальная, максимальная, средняя длина слов, минимальное, максимальное, среднее количество слов в предложении, самое длинное и короткое предложение, десять самых часто встречаемых слов.

Рассмотрим работу приложения на примере произведения Леонида Дайнеко «Жалезныя жалуды». После запуска программы выдаётся первое изображение, отображающее частоту букв в тексте (рисунок 2).

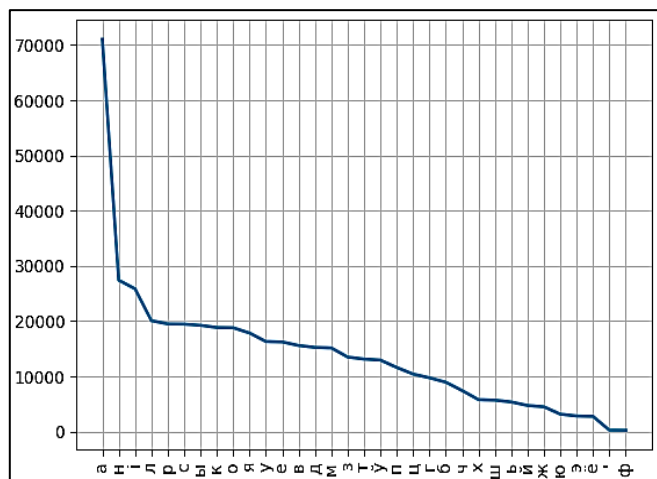


Рисунок 2 – Частоты букв в тексте

Далее выводится график наиболее часто встречающихся пар букв в тексте (рисунок 3).

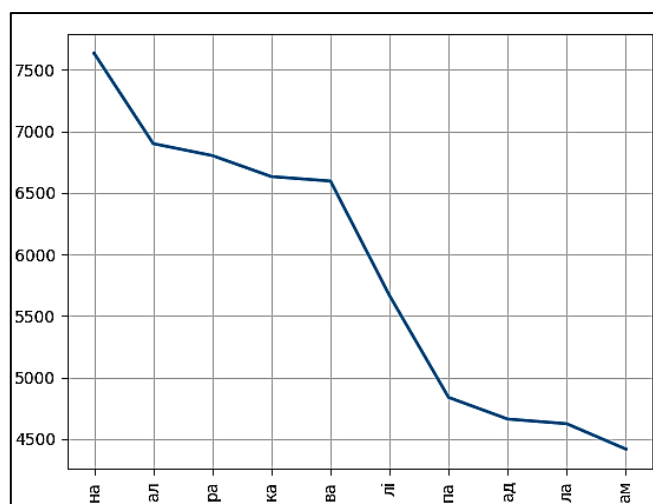


Рисунок 3 – Наиболее часто встречаемые пары букв в тексте

Определены длины слов и их частоты (рисунок 4), количество слов в предложении и их частоты (рисунок 5).

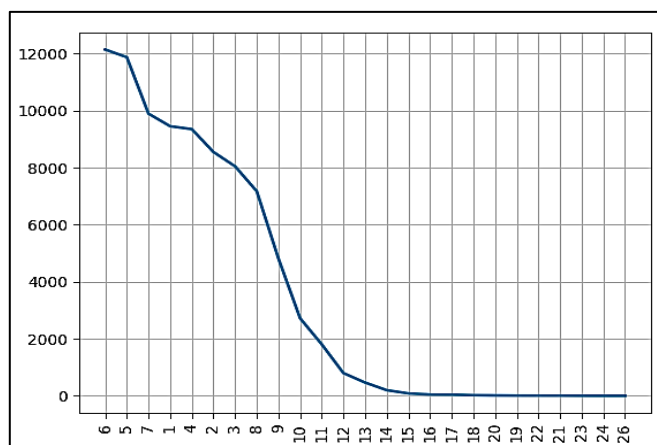


Рисунок 4 – Средняя длина слов

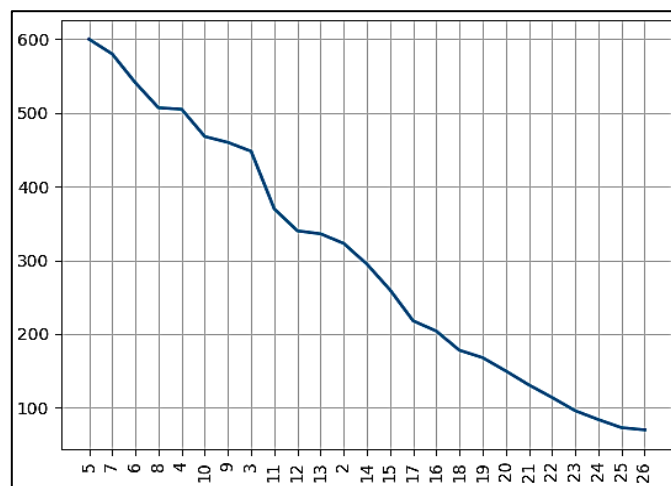


Рисунок 5 – Среднее количество слов в предложении

На рисунке 6 представлены десять самых частых слов и их частота.

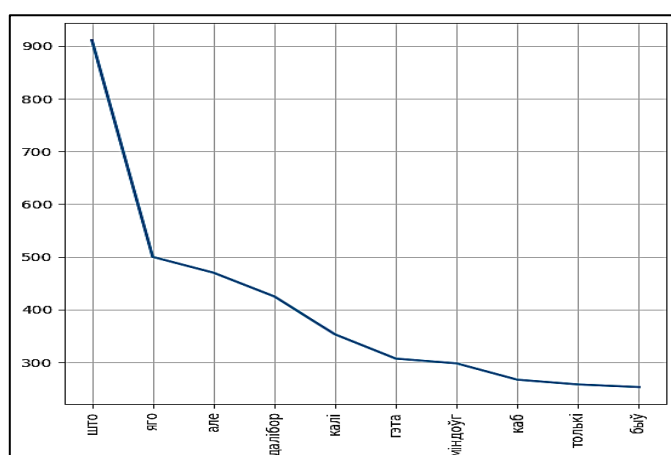


Рисунок 6 – Частота слов

Проведено сравнение произведений В. Короткевича, Л. Дайнеко и Л. Рублевской. Для сравнения были выбраны произведения “Цыганскі кароль”, “Сівая легенда” В. Короткевича, “Жалезныя жалуды”, “Меч князя Вячкі” Л. Дайнеко и “Сутарэнні Ромула”, “Скокі смерці” Л. Рублевской. Сравнение проводилось по следующим параметрам: наиболее часто используемые буквы в тексте; информационная энтропия, информационная избыточность; наиболее часто встречаемые пары букв; статистика по длине слов; статистика по количеству слов в предложении (рисунок 7); наиболее часто встречаемые слова (рисунок 8).

	Количество слов в предложении					
	Владимир Короткевич		Леонид Дайнеко		Людмила Рублевская	
	Цыганскі кароль	Сівая легенда	Жалезныя жалуды	Меч князя Вячкі	Скокі смерці	Сутарэнні Ромула
Минимальное	2	2	2	2	2	2
Максимальное	48	54	62	77	71	69
Среднее	8,41	9,04	11,22	10,63	8,73	9,82

Рисунок 7 – Количество слов в предложении

	Самые частые слова, %					
	Владимир Короткевич		Леонид Дайнеко		Людмила Рублевская	
	Цыганскі кароль	Сівая легенда	Жалезныя жалуды	Меч князя Вячкі	Скокі смерці	Сутарэнні Ромула
1	Шго - 1,51%	Шго - 1,62%	Шго - 1,30%	Шго - 1,07%	Шго - 1,84%	Шго - 1,77%
2	Яго - 1,27%	Яго - 0,95%	Яго - 0,71%	Яго - 0,69%	Але - 0,91%	Але - 0,68%
3	Гэта - 0,78%	Гэта - 0,83%	Але - 0,67%	Вячка - 0,58%	Юрась - 0,66%	Корб - 0,60%
4	Яноўскі - 0,75%	Але - 0,72%	Далібор - 0,61%	Але - 0,56%	Мяне - 0,59%	Гэта - 0,60%
5	Кароль - 0,62%	Калі - 0,68%	Калі - 0,50%	Князь - 0,47%	Гэта - 0,57%	Так - 0,53%
6	Але - 0,59%	Яны - 0,53%	Гэта - 0,44%	Быў - 0,38%	Мне - 0,52%	Калі - 0,50%
7	Калі - 0,56%	Толькі - 0,53%	Міндоўт - 0,43%	Было - 0,38%	Нібыта - 0,52%	Ася - 0,50%
8	Толькі - 0,48%	Сказаў - 0,52%	Каб - 0,38%	Толькі - 0,37%	Калі - 0,50%	Яго - 0,47%
9	Яшчэ - 0,46%	Было - 0,47%	Толькі - 0,37%	Яны - 0,35%	Яго - 0,47%	Пра - 0,45%
10	Пад - 0,42%	Яшчэ - 0,44%	Быў - 0,36%	Гэта - 0,34%	Толькі - 0,44%	Варановіч - 0,45%

Рисунок 8 – Частота слов

Определяемые статистики позволяют количественно охарактеризовать произведения авторов и сравнить их стили. Например, количество слов в предложениях рассматриваемых произведений Л. Дайнеко существенно различается с произведениями В. Короткевича и Л. Рублевской на уровне значимости 0.05. По наиболее часто встречаемым словам (рисунок 8) видно разнообразие текста, частота встреч не превышает 2 %, выбираются только слова, длина которых не менее трёх символов.

Таким образом, представленное приложение, разработанное на языке программирования Python, анализирует текст на белорусском языке, заданный пользователем. Приложение визуализирует статистическую информацию о произведении, касающуюся информационной избыточности языка: частоту букв, пар букв, слов, длин слов и предложений. Приложение позволяет сравнивать произведения на белорусском языке различных авторов, выявляя их стилистические особенности.

Полученное приложение может быть использовано специалистами в области компьютерной лингвистики.

Литература

- 1 Орлов, Ю. Н. Методы статистического анализа литературных текстов / Ю. Н. Орлов, К. П. Осминин. – М. : Книжный дом «ЛИБРОКОМ», 2017. – 312 с.
- 2 Шевелёв, О. Г. Разработка и исследование алгоритмов сравнения стилей текстовых произведений: дис. ... канд. техн. наук : 05.13.18 / Шевелев Олег Геннадьевич; Том. гос. ун-т. – Томск, 2006. – 176 с.
- 3 Васильев, Ю. Обработка естественного языка. Python и spaCy на практике / Ю. Васильев. – СПб. : Питер, 2021. – 256 с.
- 4 Златопольский, Д. Основы программирования на языке Python / Д. Златопольский. – М. : ДМК Пресс, 2017. – 284 с.
- 5 Лусиану, Р. Python – к вершинам мастерства: Лаконичное и эффективное программирование / Р. Лусиану. – М. : МК Пресс, 2022. – 898 с.
- 6 Рейтц, К. Автостопом по Python / К. Рейтц, Т. Шлюссер. – СПб. : Питер, 2017. – 336 с.