

результат. Выявлены аномалии и выбросы, требующие дополнительного анализа. Основные компоненты показали, что значительная часть вариации данных объясняется несколькими основными факторами. Кластеризация позволила разделить данные на группы со схожими характеристиками.

Для более глубокого понимания потребуется дальнейшее исследование автоматизации и интеграции новых методов анализа с большими данными, а также применение этих методов в различных отраслях.

А. С. Корж
(ГГУ имени Ф. Скорины, Гомель)

РАЗРАБОТКА ПРИЛОЖЕНИЯ ДЛЯ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ ТЕКСТА НА БЕЛОРУССКОМ ЯЗЫКЕ

Для анализа текста на белорусском языке разработано приложение (на языке программирования Python), которое проводит анализ текста на белорусском языке. Приложение выводит информацию, касающуюся информационной избыточности языка, такую как: частота букв, пар букв, слов, длин слов, предложений, коммутативность, абсолютную коммутативность [1].

Архитектура приложения, которая включает в себя: задание текста пользователем; обработку текста приложением; передачу полученных результатов пользователю (рисунок 1).

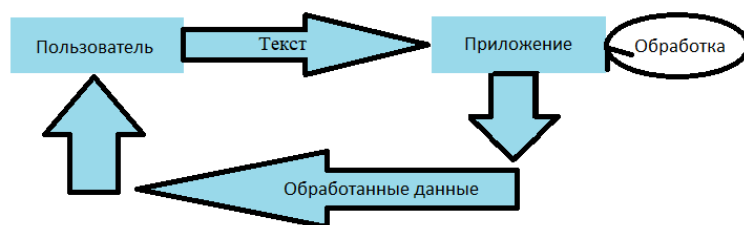


Рисунок 1 – Архитектура приложения

Реализации приложения осуществлялось с помощью программы для распознавания текстов – Tesseract. В работе использовались библиотеки: pytesseract – для считывания текста с изображения;

os – для работы с файлами компьютера пользователя; cv2 – для анализа, классификации и обработки изображений; collections – для подсчета различных данных; string, math, являющиеся встроенными модулями Python; string использовался для передачи текста заданного пользователем, math – для подсчетов различных функций; re – для использования регулярных функций; matplotlib – для вывода полученных данных в виде графика [2, 3].

Приложение работает с файлами типа: „pdf“, „docx“, „txt“, „png“, „jpg“, „jpeg“, „tiff“, „bmp“.

Полученное приложение может быть использовано специалистами в области компьютерной лингвистики.

Литература

1 Васильев, Ю. Обработка естественного языка. Python и spaCy на практике / Ю. Васильев. – СПб. : Питер, 2021. – 256 с.

2 Корж, А. С. Разработка приложения для считывания текста с изображения на языке программирования Python // Творчество молодых: сб. науч. работ студентов, магистрантов и аспирантов: Гомель: ГГУ им. Ф. Скорины. – 2024. – Ч.1. – С. 211–215.

3 Корж, А. С. Разработка приложения для считывания текста на белорусском языке / А. С. Корж // Новые математические методы и компьютерные технологии в проектировании, производстве и научных исследованиях: материалы XXVII Республиканской научной конференции студентов и аспирантов (Гомель, 18–20 марта 2024 г.). – Гомель : ГГУ им. Ф. Скорины, 2024. – С. 132.

К. А. Крюковский

(ГрГУ имени Янки Купалы, Гродно)

ПРОГНОЗИРОВАНИЕ ВАЛЮТ И КРИПТОВАЛЮТ С ПОМОЩЬЮ БИБЛИОТЕКИ NASTRADAMUS

Целью данной работы является разработка прототипа системы прогнозирования валют и криптовалют с использованием библиотеки Nastradamus. В первую очередь, была реализована система для сбора и анализа данных о ценах, используя API различных финансовых платформ и бирж.