Министерство образования Республики Беларусь

Учреждение образования «Белорусский государственный университет   
информатики и радиоэлектроники»

Факультет компьютерных систем и сетей

Кафедра информатики

Дисциплина «Операционные среды и системное программирование»

|  |  |
| --- | --- |
|  |  |

**ОТЧЕТ**

к лабораторной работе № 2

на тему «Обработка текстовой информации. Регулярные выражения»

Выполнил             Я. Ю. Прескурел

Проверил                          Н. Ю. Гриценко

Минск 2024

**СОДЕРЖАНИЕ**

[1 Постановка задачи 3](#_Toc157722973)

[2 Краткие теоретические сведения 4](#_Toc157722974)

[3 Результаты выполнения лабораторной работы 6](#_Toc157722975)

[Выводы 7](#_Toc157722976)

[Список использованных источников 8](#_Toc157722977)

[Приложение А (обязательное) Листинг исходного кода 9](#_Toc157722978)

# **1 ПОСТАНОВКА ЗАДАЧИ**

Целью выполнения данной лабораторной работы является изучение методов и средств обработки текстовой информации, включая регулярные выражения, и использующих их утилит.

# **2 КРАТКИЕ ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ**

Командная оболочка обеспечивает взаимодействие между пользователем и средой операционной системы. Она является специализированным программным продуктом, которые обеспечивает выполнение команд и получение результатов их выполнения. Примером оболочки может быть, например, оболочка *bash* операционных систем *Unix*.

Скрипты *shell* – это текстовые файлы, содержащие команды *Unix*, которые интерпретируются оболочкой операционной системы.   
Они используются для автоматизации задач, управления системными ресурсами и выполнения различных операций. Вместо того, чтобы вручную вводить команды, используются *shell*-скрипты. Достоинство *shell*-скриптов   
в том, что в одном скрипте можно объединить большое количество стандартных задач, что обеспечивает экономию времени и сокращение количества возможных ошибок при многократном выполнении однотипных задач.

Для выполнения данной лабораторной работы были использованы следующие сведения и концепции:

1 Утилиты *sed* и *awk* являются мощными средствами для обработки текстовых данных в командной строке. *Sed* (*stream editor*) предназначен для выполнения текстовых преобразований, в то время как *awk* является языком программирования и утилитой для обработки и анализа текста. Обе утилиты поддерживают использование регулярных выражений, что делает их особенно полезными при обработке больших объемов текстовой информации.[1]

2 Регулярные выражения представляют собой мощный инструмент для обработки текстовой информации. Они используются для поиска и манипулирования строк, соответствующих определенным шаблонам. Регулярные выражения предоставляют гибкие и эффективные методы поиска и замены в текстовых данных.[2]

3 Цель обработки текстовой информации заключается в создании скрипта для *sed*, *awk* и т.д., а также скрипта *shell*, способного обрабатывать входные данные в соответствии с заданным вариантом задания. Основные задачи включают в себя поиск, фильтрацию и трансформацию данных в соответствии с определенными критериями.

4 Скрипт должен быть спроектирован таким образом, чтобы корректно обрабатывать ошибочные входные данные. Это включает в себя проверку наличия необходимых файлов, корректность формата данных и предусмотрение механизмов обработки ошибок для уведомления пользователя о возможных проблемах.

5 Скрипт должен взаимодействовать с утилитами *sed*, *awk* и другими необходимыми программами для достижения цели обработки текстовой информации. Это включает передачу данных, выполнение необходимых преобразований и получение результата для дальнейшего использования.

6 Сценарий должен включать в себя механизмы обработки неочищенных данных, обеспечивая корректное выполнение задач даже в случае наличия во входных данных некорректных или непредвиденных элементов.

7 Полученные результаты обработки текстовой информации должны быть использованы для достижения конечной цели, описанной в варианте задания. Это может включать в себя формирование отчетов, выделение ключевых данных или другие действия в соответствии с требованиями задачи.

Таким образом, при использовании всех вышеперечисленных концепций был разработан *shell*-скрипт, который реализует автокорректор, меняющий строчные буквы на заглавные в начале предложений приведенного текстового документа.

# **3 РЕЗУЛЬТАТЫ ВЫПОЛНЕНИЯ ЛАБОРАТОРНОЙ РАБОТЫ**

В ходе выполнения данной лабораторной работы был разработан *bash*-скрипт, который реализует автокорректор, заменяющий строчные буквы на заглавные в начале предложений. Результат работы *bash*-скрипта представлен на рисунке 3.1.

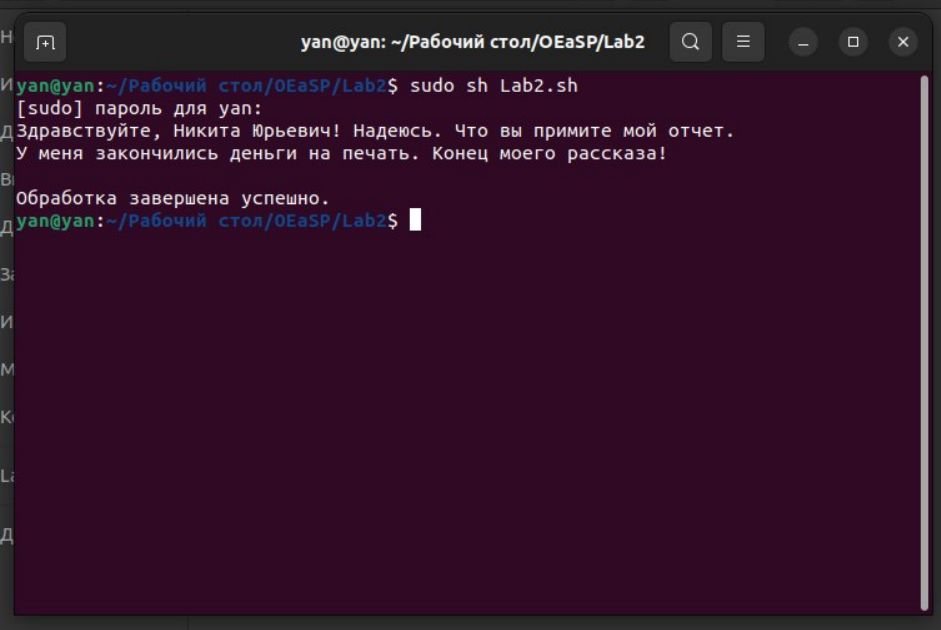


Рисунок 3.1 – Результат работы *bash*-скрипта

В итоге, при вызове *bash*-скрипта выводится в консоль исправленный текст из файла *autocorrector*.*txt*, который поддерживает текст русского и английского языка.

# **ВЫВОДЫ**

В ходе лабораторной работы были изучены методы и средства обработки текстовой информации, также произошло ознакомление с регулярными выражениями и использующими их утилитами. Кроме того, в ходе лабораторной работы был разработан скрипт для оболочки *bash*, который реализует автокорректор, меняющий строчные буквы на заглавные в начале предложений приведенного текстового документа.

# **СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ**

[1] Метод sed [Электронный ресурс]. – Режим доступа: https://losst.pro/komanda-sed-linux – Дата доступа: 08.02.2024

[2] Регулярные выражения [Электронный ресурс]. – Режим доступа: https://medium.com/introduction-into-bash/bash-scripts-regular-expressions-7d1a0473a902 – Дата доступа: 08.02.2024

[3] BashTutorials [Электронный ресурс]. – Режим доступа:   
https://prog /bash-tutorials/. – Дата доступа: 08.02.2024.

## **ПРИЛОЖЕНИЕ А**

## **(обязательное)**

## **Листинг исходного кода**

Листинг 1 – Программный код разработанного приложения

#!/bin/bash

# проверка наличия файла

if [ ! -f "autocorrector.txt" ]; then

echo "Ошибка: Файл autocorrector.txt не найден."

exit 1

fi

sed -r 's/(^|[.!?]\s)([а-яА-Яa-zA-Z])/\1\U\2/g' autocorrector.txt

# проверка успешного выполнения sed

if [ $? -ne 0 ]; then

echo "Ошибка: Возникла проблема при обработке текста."

exit 1

fi

echo "\n\nОбработка завершена успешно."