

ГУАП

КАФЕДРА № 42

ОТЧЕТ
ЗАЩИЩЕН С ОЦЕНКОЙ
ПРЕПОДАВАТЕЛЬ

доцент, канд. техн. наук

должность, уч. степень, звание

подпись, дата

В. А. Миклуш

инициалы, фамилия

ОТЧЕТ О ЛАБОРАТОРНОЙ РАБОТЕ № 3

СЛОВАРНО-ОРИЕНТИРОВАННЫЕ МЕТОДЫ КОДИРОВАНИЯ

по курсу:

ТЕОРИЯ ИНФОРМАЦИИ, ДАННЫЕ, ЗНАНИЯ

РАБОТУ ВЫПОЛНИЛ

СТУДЕНТ гр. № _____ 4326

подпись, дата

Г. С. Томчук

инициалы, фамилия

Санкт-Петербург 2025

1 Цель работы

Цель работы: изучение словарно-ориентированных методов кодирования (LZ77, LZ78, LZW).

2 Краткое описание задания

В соответствии с вариантом, таблица 1, построить алгоритм (блок-схему) и написать программу, реализующую заданный метод кодирования. Провести ручную трассировку и сравнить полученные результаты между собой. Рассчитать среднее число элементарных сигналов.

3 Вариант задания

В таблице 1 представлен вариант задания.

Таблица 1 — Вариант работы

№ п/п	Метод кодирования	Текст
17.	LZ78	Деидеологизировали-деидеологизировали, и додеидеологизировались

4 Ход работы

4.1 Трассировка алгоритма

В таблице 2 приведена ручная трассировка работы алгоритма.

Таблица 2 — Трассировка алгоритма

№ в словаре	Код (index, char)	Добавленная фраза
1	(0, Д)	Д
2	(0, е)	е
3	(0, и)	и
4	(0, д)	д
5	(2, о)	ео
6	(0, л)	л
7	(0, о)	о
8	(0, г)	г
9	(3, з)	из
10	(3, р)	ир
11	(7, в)	ов
12	(0, а)	а
13	(6, и)	ли
14	(0, -)	-

№ в словаре	Код (index, char)	Добавленная фраза
15	(4, е)	де
16	(3, д)	ид
17	(5, л)	еол
18	(7, г)	ог
19	(9, и)	изи
20	(0, р)	р
21	(11, а)	ова
22	(13, ",")	ли,
23	(0, " ")	" "
24	(3, " ")	"и "
25	(4, о)	до
26	(15, и)	деи
27	(15, о)	део
28	(6, о)	ло
29	(8, и)	ги
30	(0, з)	з
31	(10, о)	иро
32	(0, в)	в
33	(12, л)	ал
34	(3, с)	ис
35	(0, б)	б

4.2 Среднее число элементарных сигналов

Среднее число элементарных сигналов при кодировании методом LZ78 определяется выражением:

$$\bar{L} = \frac{M * (\log_2 S + \log_2 A)}{N} \quad (1)$$

Где M=35 — кол-во кодовых слов, S=35 — размер словаря, A=35 — мощность алфавита, и N=63 — длина исходной строки. Подставим имеющиеся данные в (1):

$$\bar{L} = \frac{35 \cdot (6+6)}{63} = \frac{420}{63} \approx 6.67 \text{ бит на символ}$$

4.3 Алгоритм и реализация программы

На рисунке 1 изображена блок-схема алгоритма кодирования.

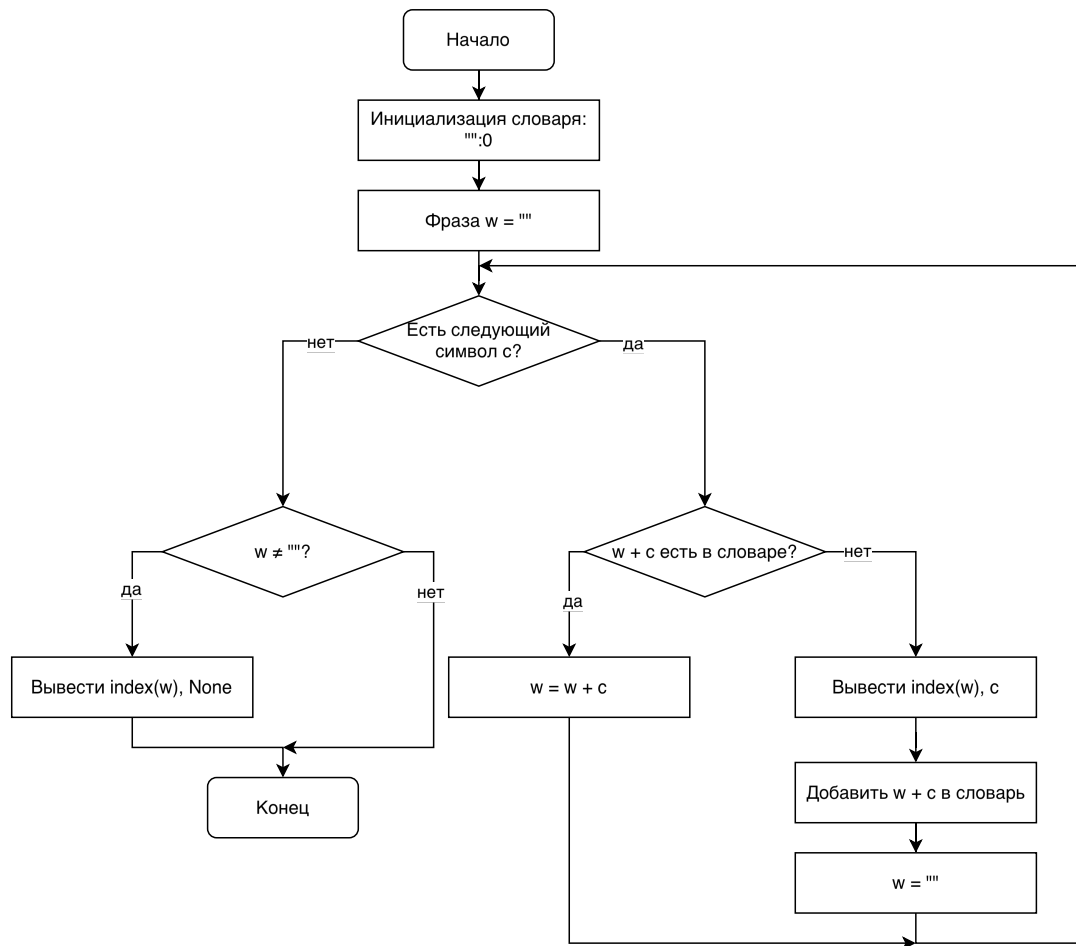


Рисунок 1 — Блок-схема

В Приложении А приведен листинг реализованной программы на языке Python. На рисунке 2 изображен результат выполнения программы.

```
→ ~/Desktop/suai/data-theory/lab03 (git:main) % py lab03.py
(0, 'д')
(0, 'е')
(0, 'и')
(0, 'д')
(2, 'о')
(0, 'л')
(0, 'о')
(0, 'г')
(3, 'э')
(3, 'р')
(7, 'в')
(0, 'а')
(6, 'и')
(0, '-')
(4, 'е')
(3, 'д')
(5, 'л')
(7, 'г')
(9, 'и')
(0, 'р')
(11, 'а')
(13, ',')
(0, ' ')
(3, ' ')
(4, 'о')
(15, 'и')
(15, 'о')
(6, 'о')
(8, 'и')
(0, 'з')
(10, 'о')
(0, 'в')
(12, 'л')
(3, 'с')
(0, 'ь')
```

Рисунок 2 — Результат выполнения программы

Как видно из результатов выполнения программы, полученные коды полностью совпадают с кодами, полученными с помощью ручной трассировки.

5 Вывод

В ходе выполнения работы был изучен словарно-ориентированный метод кодирования LZ78. Был построен алгоритм и блок-схема кодирования, реализована программа, выполнена ручная трассировка процесса сжатия заданного текста. Проведено сравнение длины исходного сообщения и кодированной последовательности, а также рассчитано среднее число элементарных сигналов. Метод LZ78 показал эффективность за счёт повторяющихся фрагментов текста и динамического формирования словаря.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Миклуш, В.А. Основы теории информации: Учебно-методическое пособие / В.А. Миклуш, В.А. Ушаков. — СПб: ГУАП, 2024. — 41 с.
2. Шеннон, К. Э. Работы по теории информации и кибернетике / К. Э. Шеннон. — М.: Иностранная литература, 1963. — 830 с.
3. Колодуб, В. Д. Теория информации: учебник для вузов / В. Д. Колодуб. — М.: ФОРУМ, 2019. — 352 с.
4. Ковалёв, В. А. Теория информации и кодирование: учебное пособие / В. А. Ковалёв. — СПб.: Питер, 2020. — 368 с.

ПРИЛОЖЕНИЕ А

Листинг программы

```
def lz78_encode(text):
    dictionary = {"": 0}
    dict_size = 1
    w = ""
    result = []

    for c in text:
        if w + c in dictionary:
            w = w + c
        else:
            result.append((dictionary[w], c))
            dictionary[w + c] = dict_size
            dict_size += 1
            w = ""

    if w != "":
        result.append((dictionary[w], None))

    return result

text = "Деидеологизировали-деидеологизировали, и додеидеологизировались"
encoded = lz78_encode(text)

for pair in encoded:
    print(pair)
```