

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ “ЛЬВІВСЬКА ПОЛІТЕХНІКА”
Інститут комп’ютерних наук та інформаційних технологій
Кафедра систем штучного інтелекту

Звіт до лабораторної роботи №7
з дисципліни «*Теорія інформації*»

Виконав:

ст. гр. КН-211

Головень Ростислав

Викладач:

д.т.н. Косаревич Р. Я.

Лабораторна робота №7

Метод LZ78

Завдання:

1. Реалізувати програмно алгоритм кодування послідовностей символів на основі методу LZ78
2. Здійснити кодування:
 - а) власного прізвища записаного два рази без пропуску.
 - б) довільного фрагмента тексту довжиною від 20 до 30 слів.
 - в) випадково згенерованого набору символів у вигляді 20- 30 слів.

Код програми:

```
import pandas as pd

def compress(data):
    comp_data = []
    dictionary = ['']
    word = ''
    i = 0
    for char in data:
        i += 1
        word += char
        if not word in dictionary:
            dictionary.append(word)
            comp_data.append([dictionary.index(word[:-1]), word[-1]])
            word = ''
        elif i == len(data):
            comp_data.append([dictionary.index(word), ''])
            word = ''
    lapki = []
    for elem in dictionary:
        k = '">'+elem+'"'
        lapki.append(k)
    print(pd.Series(lapki))
    for elem in comp_data:
        print(elem)
    print("_____")
    print(dictionary)
    return comp_data

if __name__ == '__main__':
    data = "HolovenHoloven"
    comp_data = compress(data)
    print(comp_data)
```

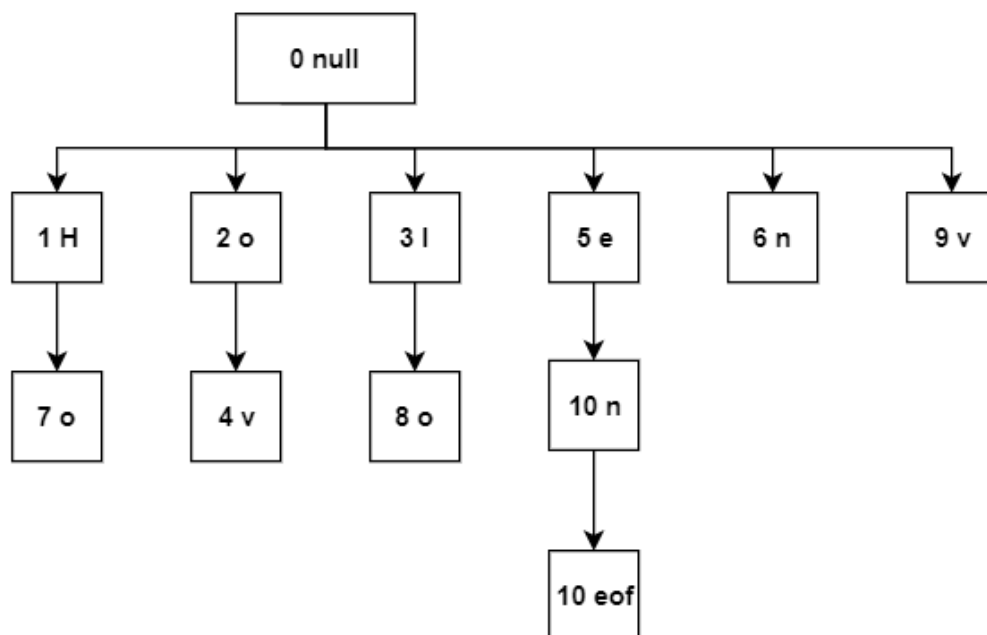
1. Здійснимо кодування власного прізвища записаного два рази без пропуску.

- Вхідна послідовність: *HolovenHoloven*

Кодовий словник:

№	Словник	Мітка
0	0 null	
1	"H"	[0, 'H']
2	"o"	[0, 'o']
3	"l"	[0, 'l']
4	"ov"	[2, 'v']
5	"e"	[0, 'e']
6	"n"	[0, 'n']
7	"Ho"	[1, 'o']
8	"lo"	[3, 'o']
9	"v"	[0, 'v']
10	"en(eof)"	[5, 'n(eof)']

- Кодова послідовність: 0H0o0l2v0e0n1o3o0v5n
- Довжина отриманого коду: 120 біт
- Дерево словника:



2. Здійснимо кодування довільного фрагмента тексту довжиною від 20 до 30 слів.

- Вхідна послідовність:

*Well_somebody_told_me_you_had_a_boyfriend_who_looked_like_a_girlfriend
_that_I_had_in_February_of_last_year_it_is_not_confidential_I_have_got_pot
ential*

Кодовий словник:

№	Словник	Мітка
0	0 null	
1	"W"	[0, 'W']
2	"e"	[0, 'e']
3	"l"	[0, 'l']
4	"l_"	[3, '_']
5	"s"	[0, 's']
6	"o"	[0, 'o']
7	"m"	[0, 'm']
8	"eb"	[2, 'b']
9	"od"	[6, 'd']
10	"y"	[0, 'y']
11	"_"	[0, '_']
12	"t"	[0, 't']
13	"ol"	[6, 'l']
14	"d"10	[0, 'd']
15	"_m"	[11, 'm']
16	"e_"	[2, '_']
17	"yo"	[10, 'o']
18	"u"	[0, 'u']
19	"_h"	[11, 'h']
20	"a"	[0, 'a']
21	"d_"	[14, '_']
22	"a_"	[20, '_']
23	"b"	[0, 'b']
24	"oy"	[6, 'y']
25	"f"	[0, 'f']
26	"r"	[0, 'r']
27	"i"	[0, 'i']
28	"en"	[2, 'n']
29	"d_w"	[21, 'w']
30	"h"	[0, 'h']
31	"o_"	[6, '_']
32	"lo"	[3, 'o']
33	"ok"	[6, 'k']
34	"ed"	[2, 'd']
35	"_l"	[11, 'l']
36	"ik"	[27, 'k']
37	"e_a"	[16, 'a']
38	"_g"	[11, 'g']
39	"ir"	[27, 'r']
40	"lf"	[3, 'f']

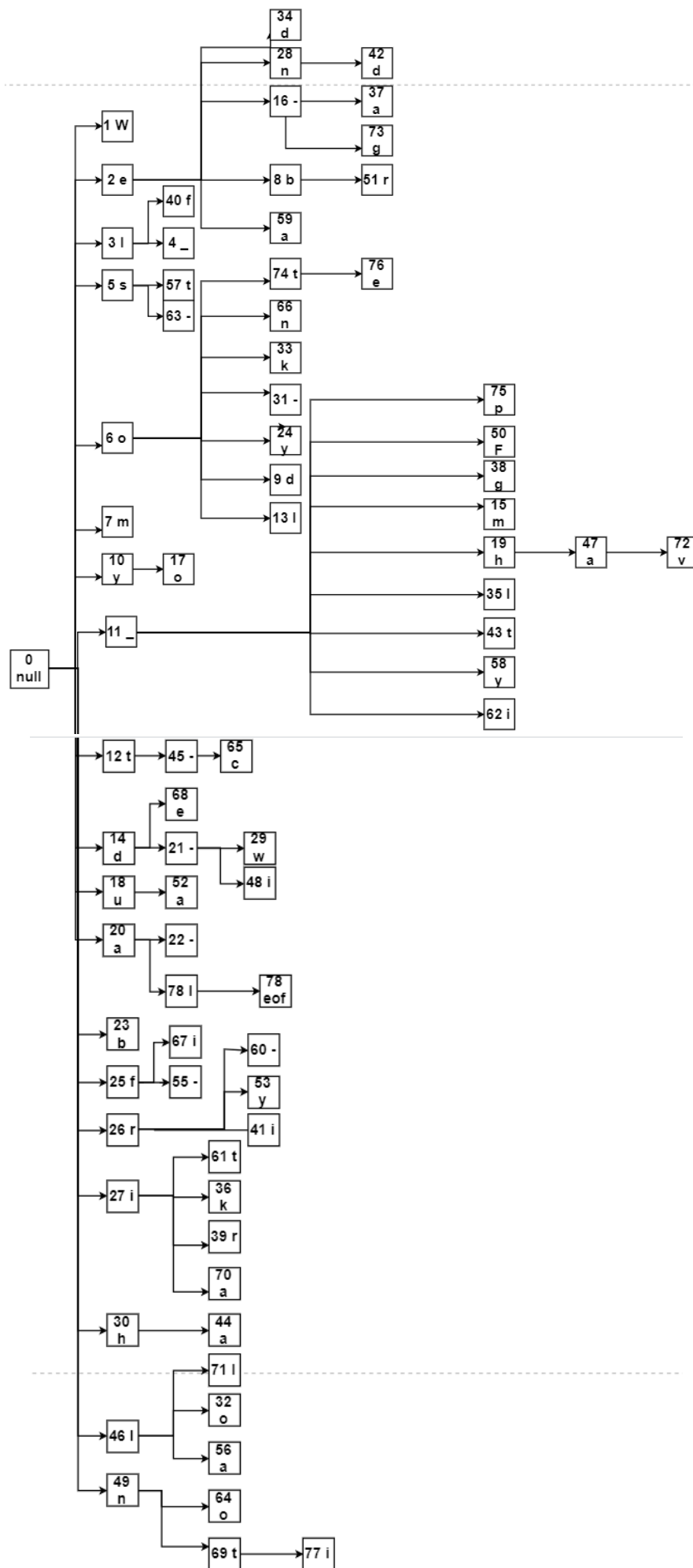
41	"ri"	[26, 'i']
42	"end"	[28, 'd']
43	"_t"	[11, 't']
44	"ha"	[30, 'a']
45	"t_"	[12, '_']
46	"I"	[0, 'I']
47	"_ha"	[19, 'a']
48	"d_i"	[21, 'i']
49	"n"	[0, 'n']
50	"_F"	[11, 'F']
51	"ebr"	[8, 'r']
52	"ua"	[18, 'a']
53	"ry"	[26, 'y']
54	"_o"	[11, 'o']
55	"f_"	[25, '_']
56	"la"	[3, 'a']
57	"st"	[5, 't']
58	"_y"	[11, 'y']
59	"ea"	[2, 'a']
60	"r_"	[26, '_']
61	"it"	[27, 't']
62	"_i"	[11, 'i']
63	"s_"	[5, '_']
64	"no"	[49, 'o']
65	"t_c"	[45, 'c']
66	"on"	[6, 'n']
67	"fi"	[25, 'i']
68	"de"	[14, 'e']
69	"nt"	[49, 't']
70	"ia"	[27, 'a']
71	"l_I"	[4, 'I']
72	"_hav"	[47, 'v']
73	"e_g"	[16, 'g']
74	"ot"	[6, 't']
75	"_p"	[11, 'p']
76	"ote"	[74, 'e']
77	"nti"	[69, 'i']
78	"al(eof)"	[20, 'l(eof)']

- Кодова послідовність:

0W0e0l3_0s0o0m2b6d0y0_0t6l0d11m2_10o0u11h0a14_20_0b6y0f0r0i2n21w
0h6_3o6k2d11l27k16a11g27r3f26i28d11t30a12_0l19a21i0n11F8r18a26y11o2
5_3a5t11y2a26_27t11i5_49o45c6n25i14e49t27a4I47v16g6t11p74e69i20l

- Довжина отриманого коду: 1170 біт

- Дерево словника:



3. Здійснимо кодування випадково згенерованого набору символів у вигляді 20 - 30 слів.

- Вхідна послідовність:

*fsgfd_erty_cvbn_gkijki_qwerwq_xcvbxn_adsfs_tyhrt_kioko_zxcvz_mgmhjg_hfh
bg_kiyuki_uioluiol_zxvczd_trgrtg_tyujtyj_xcvbxvc_nmjghm_lkjend*

Кодовий словник:

№	Словник	Мітка
0	0 null	
1	"f"	[0, 'f']
2	"s"	[0, 's']
3	"g"	[0, 'g']
4	"fd"	[1, 'd']
5	"_ "	[0, '_']
6	"e"	[0, 'e']
7	"r"	[0, 'r']
8	"t"	[0, 't']
9	"y"	[0, 'y']
10	"_c"	[5, 'c']
11	"v"	[0, 'v']
12	"b"	[0, 'b']
13	"n"	[0, 'n']
14	"_g"	[5, 'g']
15	"k"	[0, 'k']
16	"i"	[0, 'i']
17	"j"	[0, 'j']
18	"ki"	[15, 'i']
19	"_q"	[5, 'q']
20	"w"	[0, 'w']
21	"er"	[6, 'r']
22	"wq"	[20, 'q']
23	"_x"	[5, 'x']
24	"c"	[0, 'c']
25	"vb"	[11, 'b']
26	"x"	[0, 'x']
27	"n_ "	[13, '_']
28	"a"	[0, 'a']
29	"d"	[0, 'd']
30	"sf"	[2, 'f']
31	"s_ "	[2, '_']
32	"ty"	[8, 'y']
33	"h"	[0, 'h']
34	"rt"	[7, 't']
35	"_k"	[5, 'k']
36	"io"	[16, 'o']
37	"ko"	[15, 'o']
38	"_z"	[5, 'z']
39	"xc"	[26, 'c']
40	"vz"	[11, 'z']
41	"_m"	[5, 'm']
42	"gm"	[3, 'm']
43	"hj"	[33, 'j']

44	"g_"	[3, '_']
45	"hf"	[33, 'f']
46	"hb"	[33, 'b']
47	"g_k"	[44, 'k']
48	"iy"	[16, 'y']
49	"u"	[0, 'u']
50	"ki_"	[18, '_']
51	"ui"	[49, 'i']
52	"o"	[0, 'o']
53	"l"	[0, 'l']
54	"uio"	[51, 'o']
55	"l_"	[53, '_']
56	"z"	[0, 'z']
57	"xv"	[26, 'v']
58	"cz"	[24, 'z']
59	"d_"	[29, '_']
60	"tr"	[8, 'r']
61	"gr"	[3, 'r']
62	"tg"	[8, 'g']
63	"_t"	[5, 't']
64	"yu"	[9, 'u']
65	"jt"	[17, 't']
66	"yj"	[9, 'j']
67	"_xc"	[23, 'c']
68	"vbx"	[25, 'x']
69	"vc"	[11, 'c']
70	"_n"	[5, 'n']
71	"m"	[0, 'm']
72	"jg"	[17, 'g']
73	"hm"	[33, 'm']
74	"_l"	[5, 'l']
75	"kj"	[15, 'j']
76	"en"	[6, 'n']
77	"d(eof)"	[29, '(eof)']

- Кодова послідовність:
0f0s0g1d0_0e0r0t0y5c0v0b0n5g0k0i0j15i5q0w6r20q5x0c11b0x13_0a0d2f2_8
y0h7t5k16o15o5z26c11z5m3m33j3_33f33b44k16y0u18_49i0o0l51o53_0z26v
24z29_8r3r8g5t9u17t9j23c25x11c5n0m17g33m5115j6n29
- Довжина отриманого коду: 1155 біт

Висновок: на даній лабораторній роботі я ознайомився з кодуванням Лімпеля-Зіва, а також програмно реалізував алгоритм кодування LZ78.

Ключовим для розміру одержуваних кодів є розмір словника у фразах, тому що кожен код при кодуванні за методом LZ78 містить номер фрази в словнику. З останнього випливає, що ці коди мають постійну довжину, рівну округленому в більшу сторону бінарного логарифму розміру словника +8 (це кількість біт в байт-коді розширеного ASCII).

Очевидно, що чим більше словник, тим (за інших рівних умов) більший ступінь стиснення можна досягти. З іншого боку, важливим практичним моментом є

швидкість упаковки, цей параметр теж залежить від пристрою словника. Основні операції при упаковці: 1) пошук в словнику фрагмента; 2) вставка в словник нових фрагментів. Необхідно, щоб ці дві операції були максимально швидкими.

А також при великих обсягах даних середнє число біт, необхідне для кодування прямує до ентропії.

Ці алгоритми широко застосовуються в архіваторах, що свідчить про його ефективність при роботі з великим обсягом даних.

Основним недоліком алгоритмів сімейства Лімпеля-Зіва є мала ефективність при кодуванні незначного обсягу даних. Крім цього, для кодування алгоритмами Лімпеля-Зіва встановлено, що символи, які повторюються і часті ланцюги символів кодуються ефективно, але на кодування початкових фраз витрачається відносно велика кількість біт.