

МИНОБРНАУКИ РОССИИ
Санкт-Петербургский государственный
электротехнический университет «ЛЭТИ»
им. В.И. Ульянова (Ленина)
Кафедра САПР

ОТЧЁТ
ЛАБОРАТОРНАЯ РАБОТА №2
по дисциплине «Алгоритмы и структуры
данных»
Тема: Алгоритмы кодирования (вариант 2)

Студент гр. 9892 _____ Лескин К.А.

Преподаватель _____ Тутуева А.В.

Санкт-Петербург
2022

1 Описание алгоритма и структур данных

В данной работе реализуется алгоритм кодирования Шенона-Фано. Работа выполнена на языке Python версии 3.8.

ShanonFanoCoder — главный класс, позволяющий кодировать и декодировать строки с помощью методов **encode** и **decode** соответственно. Помимо этого, класс сохраняет полезную информацию после кодирования/декодирования:

- **string** — человекочитаемая строка.
- **tree** — дерево, с помощью которого происходит кодирование/декодирование.
- **codes_table** — маппинг символов и их кодов, созданный с помощью дерева.
- **code_sep** — код в виде списка кодов.
- **code** — код в виде единой непрерывной строки.
- **string_size** — битовый размер строки (с допущением, что один символ — 8 бит).
- **code_size** — битовый размер кода.
- **wrapping_coefficient** — коэффициент сжатия.

Для кодирования/декодирования используется класс **ShanonFanoTree**. Объект класса может быть создан из таблицы вероятностей или из таблицы кодов. В обоих случаях построение дерева происходит с помощью класса **TreeBuilder** схожим образом:

1. Создаётся корневой узел и вносится в очередь.
2. Пока очередь не пуста

Из очереди вытаскивается таблица и узел дререва,

Если узел - лист (узлу назначен символ) - итерация пропускается.

Таблица делится на 2 части — левую и правую. Для таблицы вероятностей разделение идёт по сумме вероятностей, чтобы они были приблизительно похожи. Для таблицы кодов разделение идёт по первому биту кода (1 или 0).

Для каждой части создаётся свой узел. Если возможно, узлам назначается символ.

Созданные узлы со своими таблицами отправляются обратно в очередь.

1.1 Кодирование

Метод `encode` принимает на вход строку, которую нужно закодировать. Из входной строки формируется маппинг символов и их количества в тексте. Например, для строки "aaa aaa bbbb" маппинг будет следующим:

```
1 {  
2   'a': 6,  
3   'b': 4,  
4   ' ': 2,  
5 }
```

Из таблицы вероятностей формируется дерево. С помощью дерева создаётся таблица кодов. С помощью таблицы кодов исходная строка кодируется в двух видах: непрерывный код и код в виде списка. Затем подсчитываются размер входной строки, размер кода и коэффициент сжатия.

1.2 Декодирование

Метод `decode` принимает на вход код, который нужно декодировать, в виде непрерывной строки или списка отдельных кодов и декодирующий объект – дерево или таблицу кодов.

Сначала загружается декодер. Если это таблица кодов, то из неё формируется дерево. Сама таблица сохраняется отдельно. Если это дерево – то оно сохраняется и из него достаётся таблица кодов.

После загружается код. Если это непрерывная строка, то код разделяется с помощью метода `separate` класса `ShanonFanoTree`. Сам непрерывный код сохраняется отдельно. Если это раздельный код, то он сохраняется и соединяется в непрерывный.

Потом создаётся инвертированная таблица код-символ из таблицы символ-код. По этой таблице раздельный код переводся в декодированную строку.

Затем подсчитываются размер входной строки, размер кода и коэффициент сжатия.

2 Оценка временной сложности

Оценка временной сложности представлена в таблице 1

Метод	Оценка временной сложности
encode	$O(n)$
decode	$O(n)$

Таблица 1 – Оценка временной сложности методов класса ShanonFanoCoder

3 Примеры работы

3.1 Примеры кодирования

Пример кодирования 1

Входная строка: *It's easy to quit smoking. I've done it hundreds of times.*

Пример выполнения представлен на рис. 1

```
Encode example 1

String (length: 58, size: 464):
It's easy to quit smoking. I've done it hundreds o
f times.

Code (size: 239):
11010011011001010000001111111010111110000011010000
00111101110000111011000001010111100011110001111010
11101110110000110101100111101000100010011000101000
10000111011000011100111000101010011110000011001010
000100011011000011001111011100101010110

Code separated:
('11010', '0110', '11001', '010', '000', '001', '111111')
('010', '111110', '000', '0110', '1000', '000', '111101')
('11000', '0111', '0110', '000', '010', '10111', '1000')
('111100', '0111', '1010', '111011', '10110', '000', '11010')
('11001', '111010', '001', '000', '1001', '1000', '1010')
('001', '000', '0111', '0110', '000', '111001', '11000')
('1010', '1001', '111000', '001', '1001', '010', '000')
('1000', '11011', '000', '0110', '0111', '10111', '001')
('010', '10110')

Wrapping coefficient: 1.9414225941422594

Codes table:
(('a', '111111'), ('y', '111110'), ('q', '111101'), ('k', '111100'))
(('g', '111011'), ('v', '111010'), ('h', '111001'), ('r', '111000'))
(('f', '11011'), ('I', '11010'), ('"', '11001'), ('u', '11000'))
(('m', '10111'), ('.', '10110'), ('n', '1010'), ('d', '1001'))
(('o', '1000'), ('i', '0111'), ('t', '0110'), ('s', '010'))
(('e', '001'), (' ', '000'))
```

Рис. 1 – Пример кодирования строки 1

Пример кодирования 2

Входная строка: *Many of life's failures are people who did not realize how close they were to success when they gave up.*

Пример выполнения представлен на рис. 2

```

Encode example 2

String (length: 104, size: 832):
Many of life's failures are people who did not rea
lize how close they were to success when they gave
up.

Code (size: 431):
11111111001011101111000001101101100100010111110110
101111110011100011011100101011110001101011001001
11000100101011001000110010100110110011000010001010
0111101100011110010111111000011101011010011001011
0010100101000101111111010100001111011010100011000
10000110011100100010011011110101110000101001010110
01000100110110000111011010110001100001001110011100
01010011110101110100100110111101011100001111100100
1011110110100011010110011111010

Code separated:
('1111111', '10010', '11101', '11100', '00', '0110', '11011')
('00', '1000', '10111', '11011', '010', '1111110', '01110')
('00', '11011', '10010', '10111', '1000', '11010', '10110')
('010', '01110', '00', '10010', '10110', '010', '00')
('11001', '010', '0110', '11001', '1000', '010', '00')
('1010', '01111', '0110', '00', '111100', '10111', '111100')
('00', '11101', '0110', '10011', '00', '10110', '010')
('10010', '1000', '10111', '1111101', '010', '00', '01111')
('0110', '1010', '00', '11000', '1000', '0110', '01110')
('010', '00', '10011', '01111', '010', '11100', '00')
('1010', '010', '10110', '010', '00', '10011', '0110')
('00', '01110', '11010', '11000', '11000', '010', '01110')
('01110', '00', '1010', '01111', '010', '11101', '00')
('10011', '01111', '010', '11100', '00', '1111100', '10010')
('1111011', '010', '00', '11010', '11001', '1111010')

Wrapping coefficient: 1.9303944315545243

Codes table:
(('M', '1111111'), ('"', '1111110'), ('z', '1111101'), ('g', '1111100'))
(('v', '1111011'), ('.', '1111010'), ('d', '111100'), ('n', '11101'))
(('y', '11100'), ('f', '11011'), ('u', '11010'), ('p', '11001'))
(('c', '11000'), ('i', '10111'), ('r', '10110'), ('w', '1010'))
(('t', '10011'), ('a', '10010'), ('l', '1000'), ('h', '01111'))
(('s', '01110'), ('o', '0110'), ('e', '010'), (' ', '00'))

```

Рис. 2 – Пример кодирования строки 2

Пример кодирования 3

Входная строка: *May the force be with you*

Пример выполнения представлен на рис. 3

```

Encode example 3

String (length: 25, size: 200):
May the force be with you

Code (size: 93):
11111111101010001000111010001110101101110011011010
0011010010001100101111000111001010011010110

Code separated:
('1111', '11110', '1010', '00', '100', '0111', '010')
('00', '11101', '0110', '11100', '11011', '010', '00')
('11010', '010', '00', '1100', '10111', '100', '0111')
('00', '1010', '0110', '10110')

Wrapping coefficient: 2.150537634408602

Codes table:
(('M', '11111'), ('a', '11110'), ('f', '11101'), ('r', '11100'))
(('c', '11011'), ('b', '11010'), ('w', '1100'), ('i', '10111'))
(('u', '10110'), ('y', '1010'), ('t', '100'), ('h', '0111'))
(('o', '0110'), ('e', '010'), (' ', '00'))

```

Рис. 3 – Пример кодирования строки 3

3.2 Примеры декодирования

Пример декодирования 1

Входной код:

01111111100010001000100011110000111110010011000100
1111011001100101

Входная таблица кодов:

```

1 {
2   "h": "11111",
3   "a": "11110",
4   "e": "1110",
5   "r": "1101",
6   "n": "1100",
7   "g": "101",
8   "i": "100",
9   "t": "011",
10  "s": "010",
11  " ": "00"
12 }

```

Пример выполнения представлен на рис. 4

```

Decode example 1

String (length: 21, size: 168):
this is a test string

Code (size: 66):
01111111100010001000100011110000111110010011000100
1111011001100101

Code separated:
('011', '11111', '100', '010', '00', '100', '010')
('00', '11110', '00', '011', '1110', '010', '011')
('00', '010', '011', '1101', '100', '1100', '101')

Wrapping coefficient: 2.5454545454545454

Codes table:
(('h', '11111'), ('a', '11110'), ('e', '1110'), ('r', '1101'))
(('n', '1100'), ('g', '101'), ('i', '100'), ('t', '011'))
(('s', '010'), (' ', '00'))

```

Рис. 4 – Пример кодирования строки 1

Пример декодирования 2

Входной код:

```

1111111101000100001010111111100011011001101010100
10010111000110011000100101110110111101000001111100
11101111000010101100100110100011101010110011100001
10010111001001000011110101110001101000001000010111
10010010101011111001101011000000011010001101110011
10110

```

Входная таблица кодов:

```

1 {
2     "т": "111111",
3     "v": "111110",
4     ",": "111101",
5     "x": "111100",
6     "p": "111011",
7     "o": "111010",
8     "k": "111001",
9     "c": "111000",
10    "m": "110111",
11    ".": "110110",
12    "h": "11010",
13    "f": "11001",
14    "l": "11000",
15    "d": "10111",
16    "n": "10110",
17    "u": "1010",

```

```

18     "r": "1001",
19     "i": "1000",
20     "s": "0111",
21     "t": "0110",
22     "a": "010",
23     "e": "001",
24     " ": "000"
25 }

```

Пример выполнения представлен на рис. 5

```

Decode example 2

String (length: 61, size: 488):
The adventures first, explanations take such a dre
adful time.

Code (size: 255):
1111111101000100001010111111100011011001101010100
1001011100011001100010010111011011101000001111100
11101111000010101100100110100011101010110011100001
10010111001001000011110101110001101000001000010111
10010010101011111001101011000000011010001101110011
10110

Code separated:
('111111', '11010', '001', '000', '010', '10111', '111110')
('001', '10110', '0110', '1010', '1001', '001', '0111')
('000', '11001', '1000', '1001', '0111', '0110', '111101')
('000', '001', '111100', '111011', '11000', '010', '10110')
('010', '0110', '1000', '111010', '10110', '0111', '000')
('0110', '010', '111001', '001', '000', '0111', '1010')
('111000', '11010', '000', '010', '000', '10111', '1001')
('001', '010', '10111', '11001', '1010', '11000', '000')
('0110', '1000', '110111', '001', '110110')

Wrapping coefficient: 1.9137254901960785

Codes table:
(('T', '111111'), ('v', '111110'), ('.', '111101'), ('x', '111100'))
(('p', '111011'), ('o', '111010'), ('k', '111001'), ('c', '111000'))
(('m', '110111'), ('.', '110110'), ('h', '11010'), ('f', '11001'))
(('l', '11000'), ('d', '10111'), ('n', '10110'), ('u', '1010'))
(('r', '1001'), ('i', '1000'), ('s', '0111'), ('t', '0110'))
(('a', '010'), ('e', '001'), (' ', '000'))

```

Рис. 5 – Пример кодирования строки 2

Пример декодирования 3

Входной код:

```

11111111001011101110101001000011111001010101001111
01110110111001011010111010101000011101000110100101
10001100110011000110100101100001110001000011101000
1101101110110

```

Входная таблица кодов:


```

1 {
2     "S": "111111",
3     "d": "111110",
4     "n": "11110",
5     ",": "111011",
6     "m": "111010",
7     "g": "11100",
8     "i": "11011",
9     "y": "11010",
10    "u": "1100",
11    "c": "1011",
12    "e": "1010",
13    "s": "100",
14    "t": "0111",
15    ".": "0110",
16    "o": "010",
17    " ": "00"
18 }

```

Пример выполнения представлен на рис. 6

```

Decode example 3

String (length: 44, size: 352):
Success doesn't come to you... you go to it.

Code (size: 163):
111111110010111011101010010000111110010101001111
01110110111001011010111010101000011101000110100101
10001100110011000110100101100001110001000011101000
1101101110110

Code separated:
('111111', '1100', '1011', '1011', '1010', '100', '100')
('00', '111110', '010', '1010', '100', '11110', '111011')
('0111', '00', '1011', '010', '111010', '1010', '00')
('0111', '010', '00', '11010', '010', '1100', '0110')
('0110', '0110', '00', '11010', '010', '1100', '00')
('11100', '010', '00', '0111', '010', '00', '11011')
('0111', '0110')

Wrapping coefficient: 2.1595092024539877

Codes table:
(('S', '111111'), ('d', '111110'), ('n', '11110'), ('', '111011'))
(('m', '111010'), ('g', '11100'), ('i', '11011'), ('y', '11010'))
(('u', '1100'), ('c', '1011'), ('e', '1010'), ('s', '100'))
(('t', '0111'), ('.', '0110'), ('o', '010'), (' ', '00'))

```

Рис. 6 – Пример кодирования строки 3

4 Листинг

Исходный код и код отчёта доступны по ссылке: <https://github.com/kira607/2lab-algo-3-2>