

## Арифметическое сжатие (арифметическое кодирование)

Исходный текст: **БВВБААГДББ** (10 символов)

Статистика встречаемости (частоты) символов в тексте: A = 2; B = 5; B = Г = Д = 1

Итерация 0. Выбор начального диапазона:  $[0 \div 1 \cdot 10^6]$ .

Определение ширины поддиапазона  $\Delta$ : выбранный начальный диапазон (верхняя граница диапазона (минус) нижняя граница диапазона)/количество символов в тексте,  $(1 \cdot 10^6 - 0)/10 = 1 \cdot 10^5$ .

Визуальное представление выбранного диапазона с учетом частоты появления символов в исходном тексте

А		Б					В	Г	Д	1·10 <sup>6</sup>
0	1·10 <sup>5</sup>	2·10 <sup>5</sup>	3·10 <sup>5</sup>	4·10 <sup>5</sup>	5·10 <sup>5</sup>	6·10 <sup>5</sup>	7·10 <sup>5</sup>	8·10 <sup>5</sup>	9·10 <sup>5</sup>	

Итерация 1. Кодирование символа "Б".

Текущий диапазон:  $[2 \cdot 10^5 \div 7 \cdot 10^5]$ .

Определение ширины поддиапазона  $\Delta$ :  $(7 \cdot 10^5 - 2 \cdot 10^5)/10 = 0,5 \cdot 10^5$ .

А		Б					В	Г	Д	7·10 <sup>5</sup>
2·10 <sup>5</sup>	2,5·10 <sup>5</sup>	3·10 <sup>5</sup>	3,5·10 <sup>5</sup>	4·10 <sup>5</sup>	4,5·10 <sup>5</sup>	5·10 <sup>5</sup>	5,5·10 <sup>5</sup>	6·10 <sup>5</sup>	6,5·10 <sup>5</sup>	

Итерация 2. Кодирование символа "Б".

Текущий диапазон:  $[3 \cdot 10^5 \div 5,5 \cdot 10^5]$ .

Определение ширины поддиапазона  $\Delta$ :  $(5,5 \cdot 10^5 - 3 \cdot 10^5)/10 = 25 \cdot 10^3$ .

А		Б					В	Г	Д	550·10 <sup>3</sup>
300·10 <sup>3</sup>	325·10 <sup>3</sup>	350·10 <sup>3</sup>	375·10 <sup>3</sup>	400·10 <sup>3</sup>	425·10 <sup>3</sup>	450·10 <sup>3</sup>	475·10 <sup>3</sup>	500·10 <sup>3</sup>	525·10 <sup>3</sup>	

Итерация 3. Кодирование символа "В".

Текущий диапазон:  $[475 \cdot 10^3 \div 500 \cdot 10^3]$ .

Определение ширины поддиапазона  $\Delta$ :  $(500 \cdot 10^3 - 475 \cdot 10^3)/10 = 2,5 \cdot 10^3 = 2500$ .

А		Б					В	Г	Д	500000
475000	477500	480000	482500	485000	487500	490000	492500	495000	497500	

Итерация 4. Кодирование символа "Б".

Текущий диапазон:  $[480 \cdot 10^3 \div 492,5 \cdot 10^3]$ .

Определение ширины поддиапазона  $\Delta$ :  $(492,5 \cdot 10^3 - 480 \cdot 10^3)/10 = 1250$ .

А		Б					В	Г	Д	492500
480000	481250	482500	483750	485000	486250	487500	488750	490000	491250	

Итерация 5. Кодирование символа "А".

Текущий диапазон: [480000 ÷ 482500).

Определение ширины поддиапазона  $\Delta$ :  $(482500-480000)/10 = 250$ .

А		Б					В	Г	Д	
480000	480250	480500	480750	481000	481250	481500	481750	482000	482250	482500

Итерация 6. Кодирование символа "А".

Текущий диапазон: [480000 ÷ 480500).

Определение ширины поддиапазона  $\Delta$ :  $(480500-480000)/10 = 50$ .

А		Б					В	Г	Д	
480000	480050	480100	480150	480200	480250	480300	480350	480400	480450	480500

Итерация 7. Кодирование символа "Г".

Текущий диапазон: [480400 ÷ 480450).

Определение ширины поддиапазона  $\Delta$ :  $(480450-480400)/10 = 5$ .

А		Б					В	Г	Д	
480400	480405	480410	480415	480420	480425	480430	480435	480440	480445	480450

Итерация 8. Кодирование символа "Д".

Текущий диапазон: [480445 ÷ 480450).

Определение ширины поддиапазона  $\Delta$ :  $(480450-480445)/10 = 0,5$ .

А		Б					В	Г	Д	
480445,0	480445,5	480446,0	480446,5	480447,0	480447,5	480448,0	480448,5	480449,0	480449,5	480450,0

Итерация 9. Кодирование символа "Б".

Текущий диапазон: [480446,0 ÷ 480448,5).

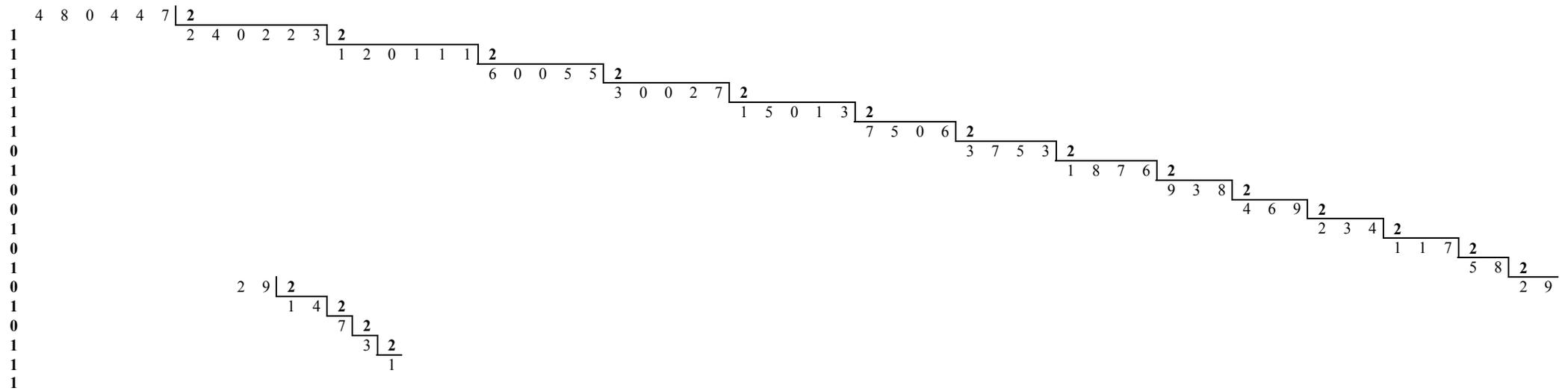
Определение ширины поддиапазона  $\Delta$ :  $(480448,5-480446,0)/10 = 0,25$ .

А		Б					В	Г	Д	
480446,0	480446,25	480446,5	480446,75	480447,0	480447,25	480447,5	480447,75	480448,0	480448,25	480448,5

Итерация 10. Кодирование символа "Б".

Текущий диапазон: [480446,5 ÷ 480447,75).

Поскольку был закодирован последний символ исходного текста, то определение ширины поддиапазона  $\Delta$  больше не требуется. Результат арифметического кодирования текста – любое число (почти любое: желательно выбрать число с наименьшим числом знаков) из последнего определенного диапазона, в нашем случае это диапазон [480446,5 ÷ 480447,75), в который входит число 480447. Теперь необходимо только перевести это число в двоичный вид, что и проделано ниже:



Двоичная последовательность, полученная в результате арифметического сжатия: 1110101010010111111 (19 бит)

Для сравнения степени сжатия, закодируем тот же исходный текст посредством статического алгоритма Хаффмана

На основании статистики встречаемости символов в тексте построим дерево Хаффмана

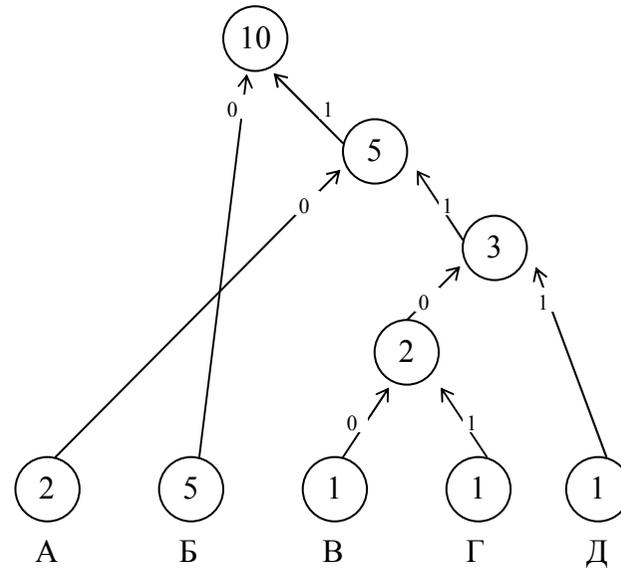


Таблица кодов постоянной длины  $m$  (3 бита) и кодов Хаффмана переменной длины  $k$  (2-4 бит)

символ	исходный код			код Хаффмана			
<b>А</b>	0	0	0			1	0
<b>Б</b>	0	0	1				0
<b>В</b>	0	1	0	1	1	0	0
<b>Г</b>	0	1	1	1	1	0	1
<b>Д</b>	1	0	0		1	1	1

Объем исходного (несжатого) текста в равномерной кодировке  $V_{исх} = \text{количество символов} \times \text{разрядность кода } m$ .  $V_{исх} = 10 \times 3 = 30$  (бит)

Объем сжатого текста по Хаффману  $V_{сж} = \text{частота (вес) символа} \times \text{разрядность кода } k$ .  $V_{сж} = 2 \times 2 + 5 \times 1 + 1 \times 4 + 1 \times 4 + 1 \times 3 = 20$  (бит)

Объем сжатого текста посредством арифметического сжатия - 19 бит