

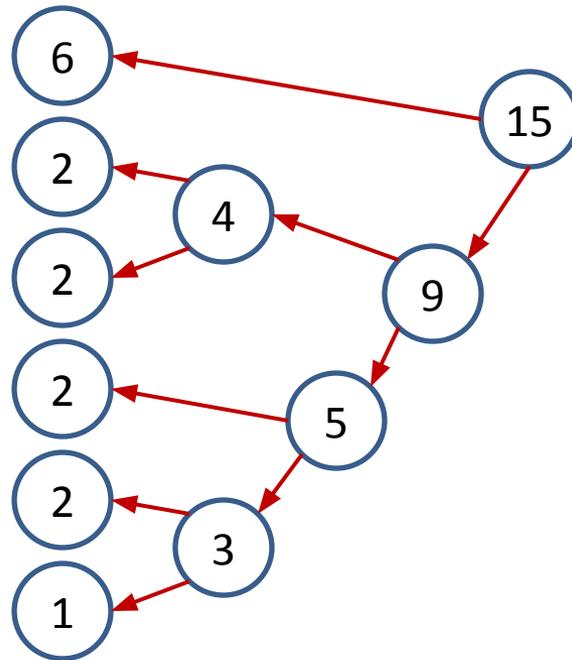
Сжатие
по Хаффману
без дерева
и гвоздей!



Дерево Хаффмана

Исходный текст: AFABCDEABCSAADEA

A	6
B	2
C	2
D	2
E	2
F	1



A	6	0
B	2	100
C	2	101
D	2	110
E	2	1110
F	1	1111

Долой деревьяшку!

Недостатки дерева:

- восстановить код символа – **сложно**.
- сохранить/передать дерево декодеру – **сложно**.

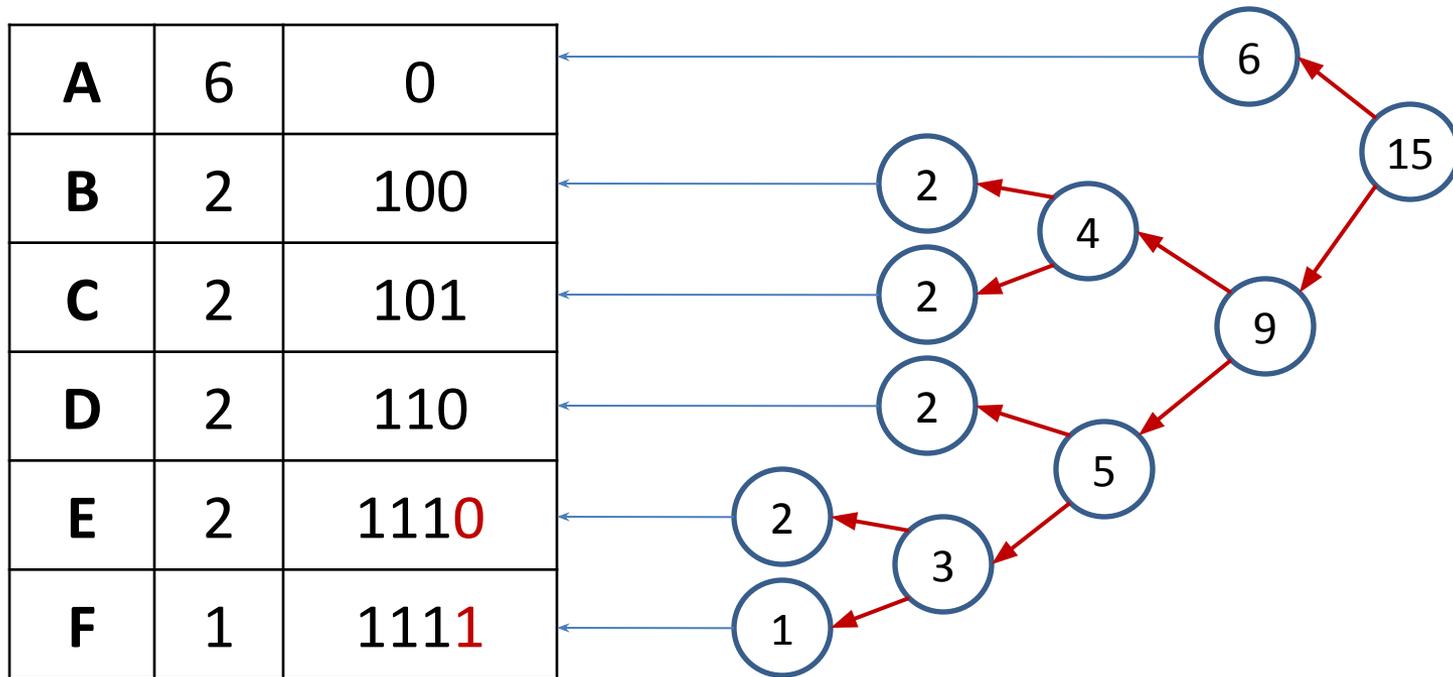
Достоинства дерева:

- декодировать текст – **просто!**

Выводы:

- При декодировании лучше использовать дерево.
- При кодировании дерево лучше чем-то заменить.

Замена дереву



Оказывается, структура дерева и длины кодов символов жёстко связаны.

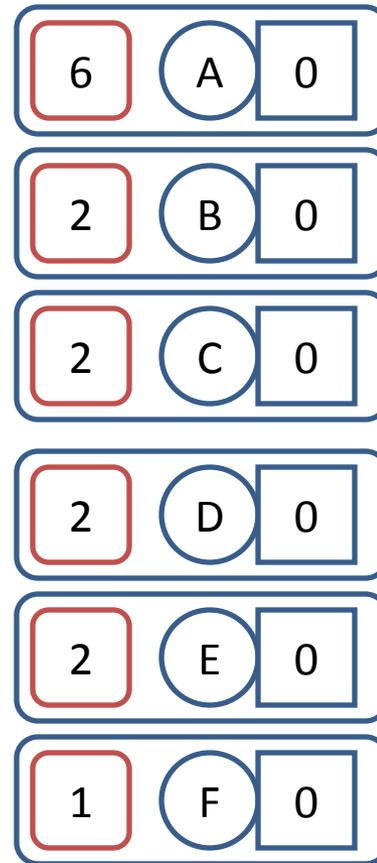
Коды Хаффмана **можно восстановить только по их длине**, при условии, что символы объединяются по правилам упорядоченности!

Алгоритм «без липы»

Исходный текст: AFABCDEABCSAADEA

A	6
B	2
C	2
D	2
E	2
F	1

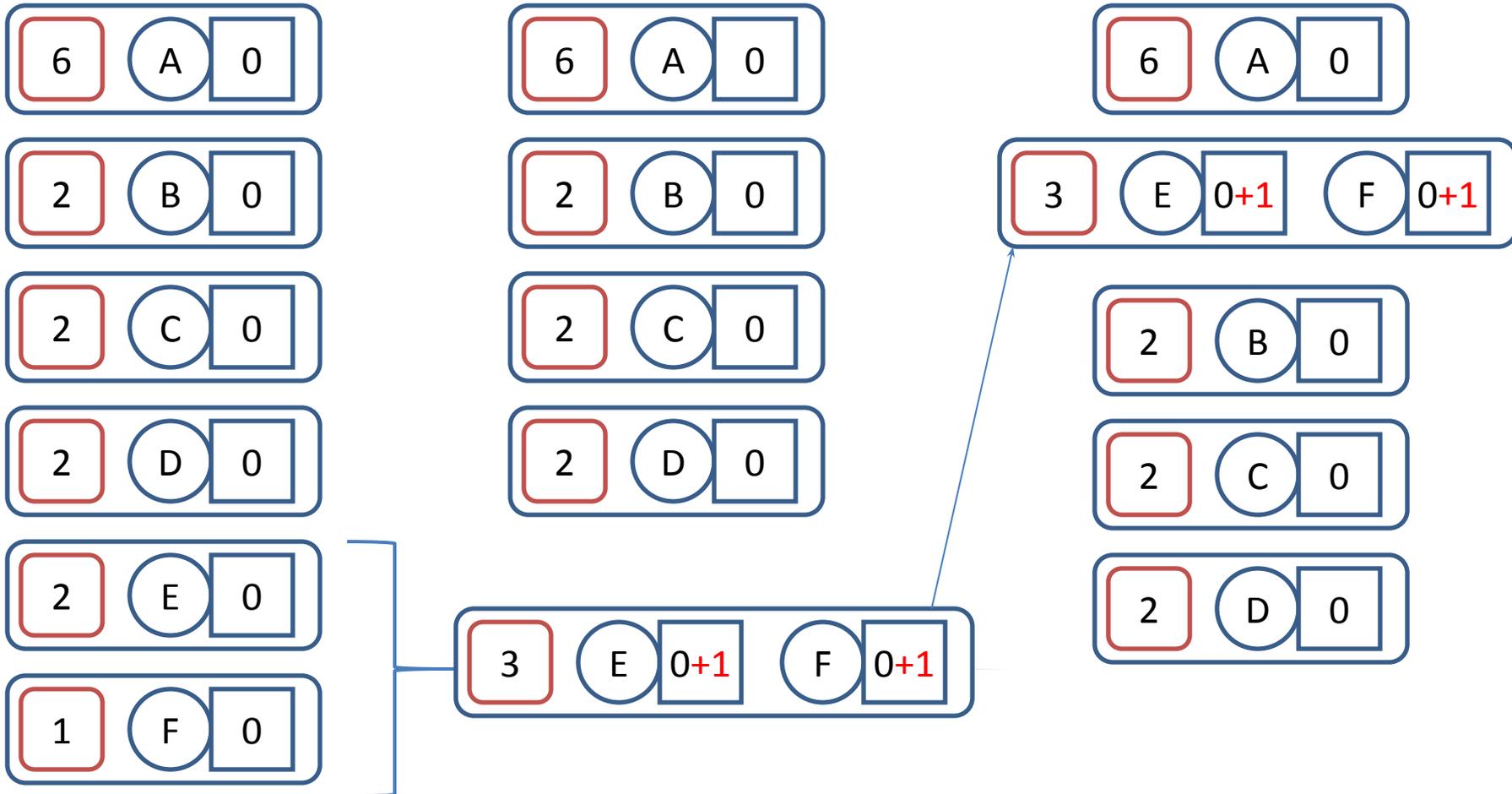
*Формируем
списки*



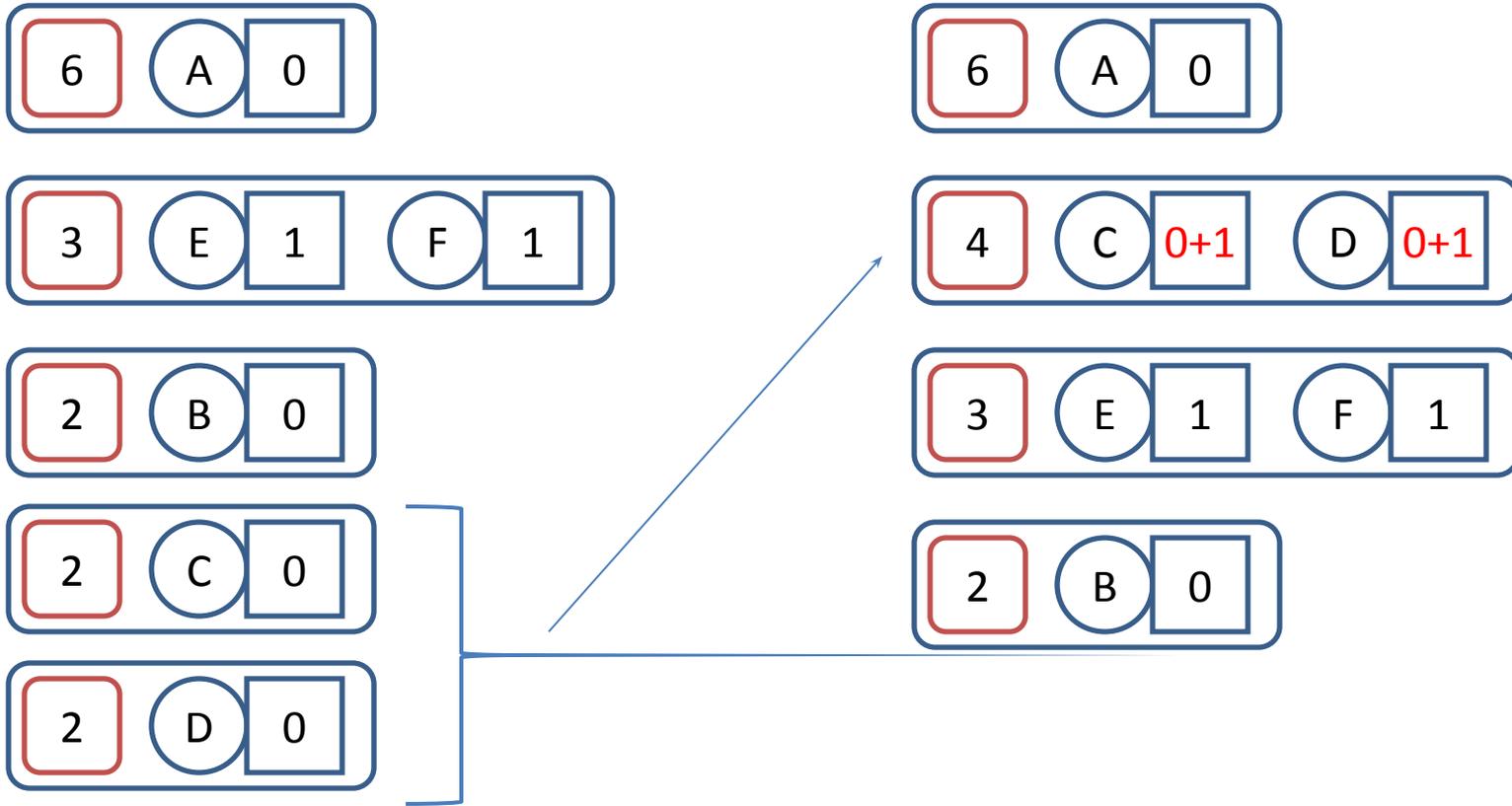
Объединение списков

Объединить самые «лёгкие» списки

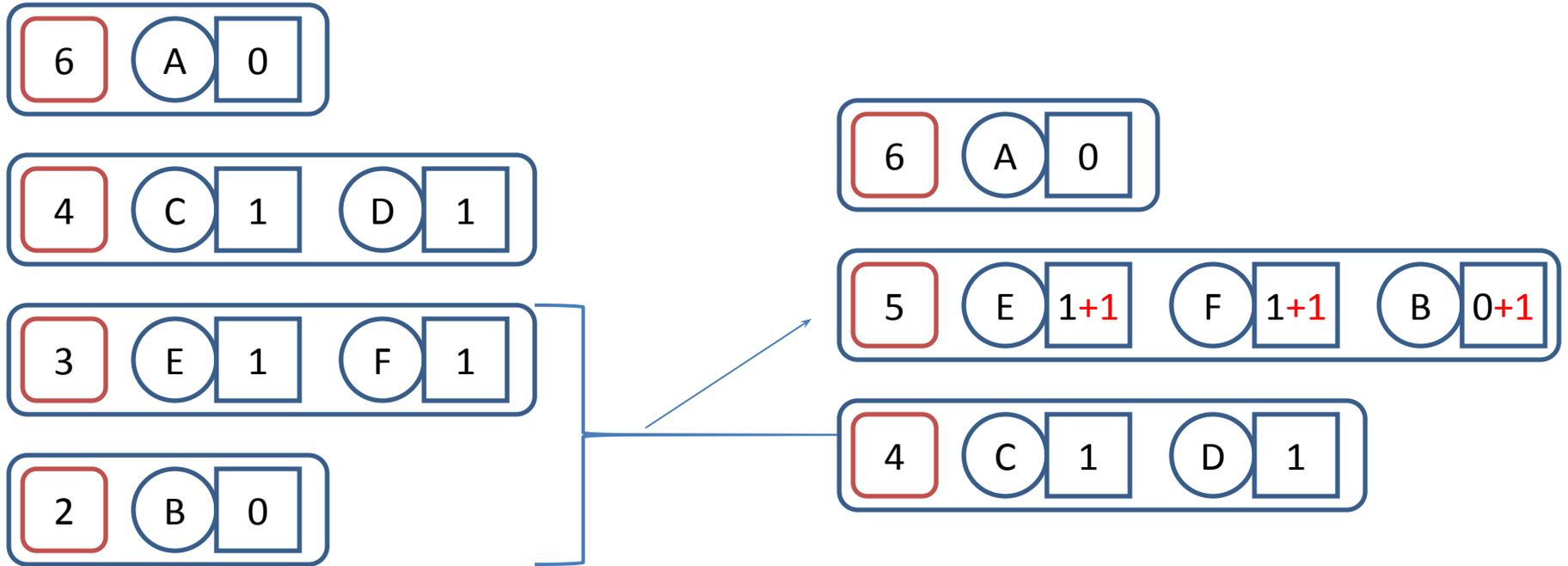
Упорядочить списки



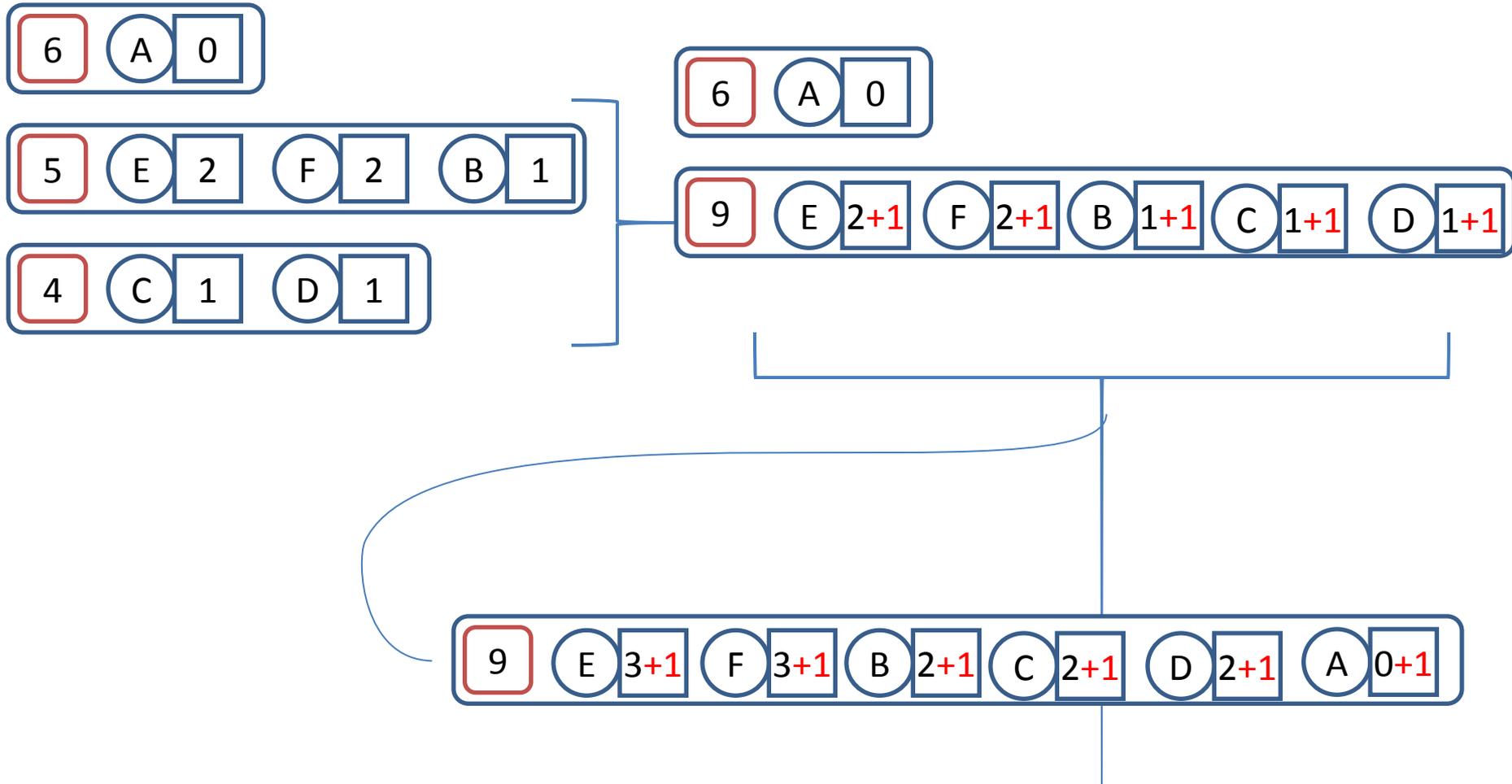
Объединение списков



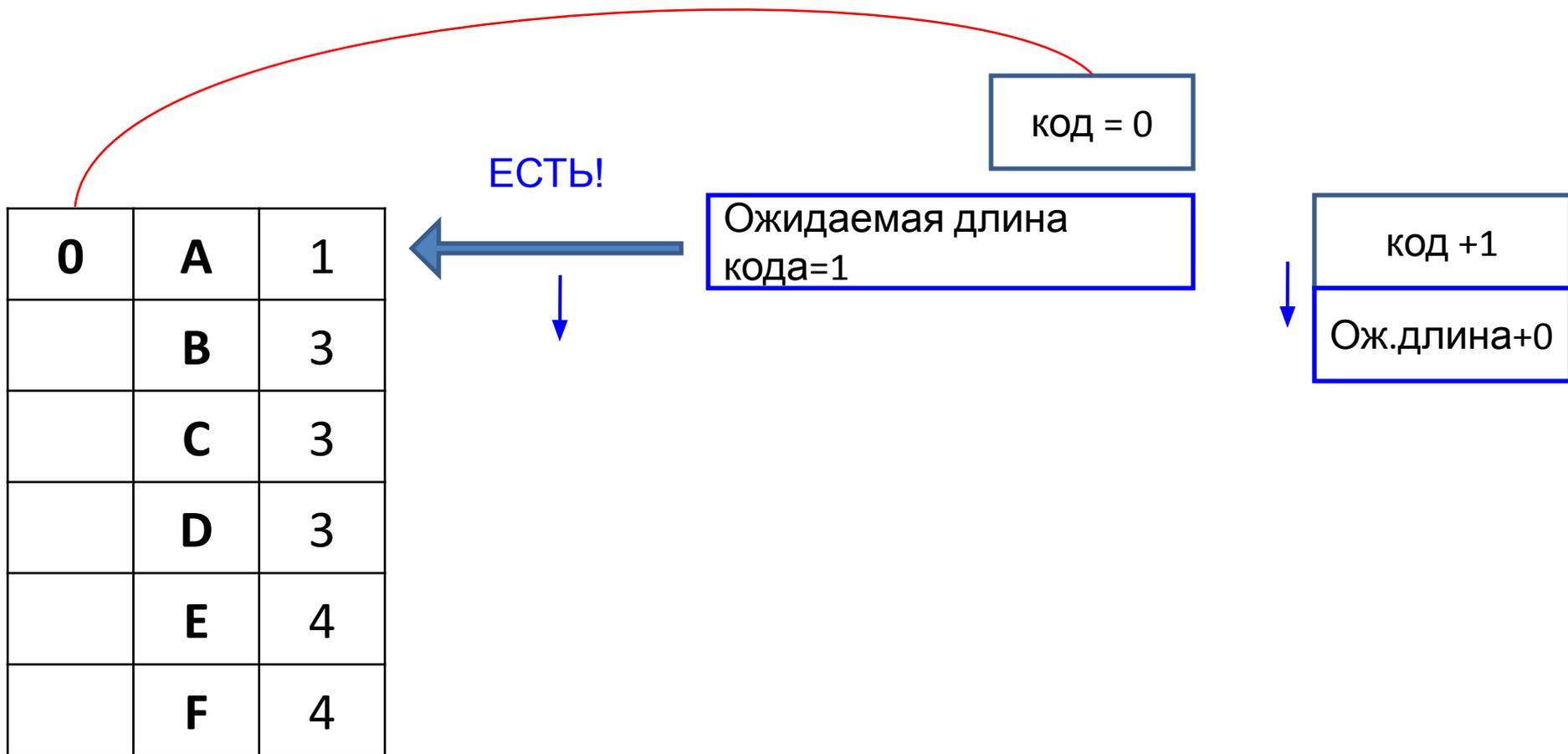
Объединение списков



Объединение списков



Восстановление кодов Хаффмана



Восстановление кодов Хаффмана

код = 1

0	A	1
	B	3
	C	3
	D	3
	E	4
	F	4

НЕТ



Ожидаемая длина
кода=1

код << 1

Ож.длина+1

Восстановление кодов Хаффмана

код = 10

0	A	1
	B	3
	C	3
	D	3
	E	4
	F	4

НЕТ

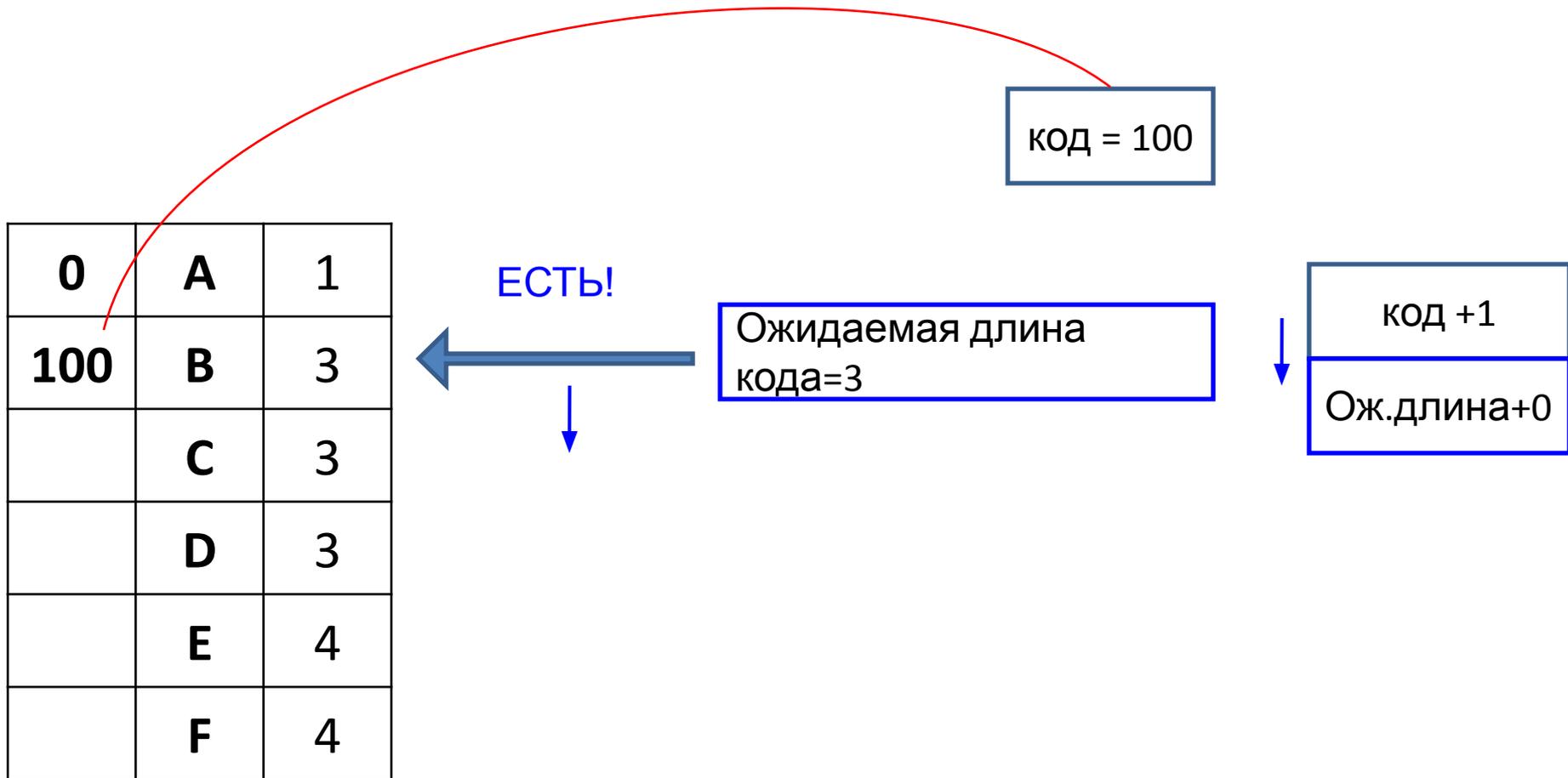


Ожидаемая длина
кода=2

код << 1

Ож.длина+1

Восстановление кодов Хаффмана



Восстановление кодов Хаффмана

код = 101

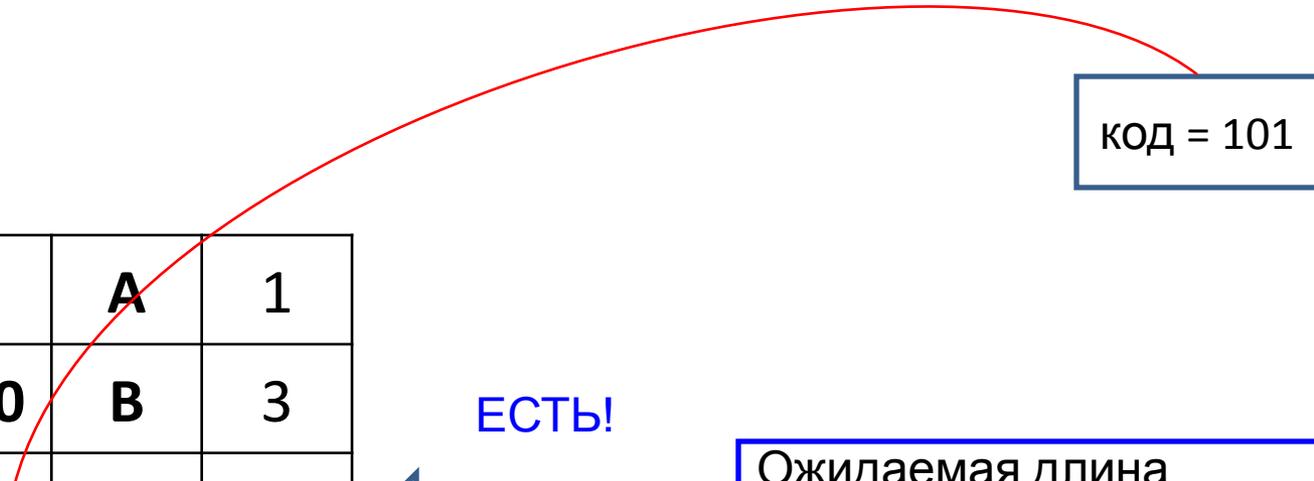
0	A	1
100	B	3
101	C	3
	D	3
	E	4
	F	4

ЕСТЬ!

Ожидаемая длина
кода=3

код +1

Ож.длина+0



Восстановление кодов Хаффмана

0	A	1
100	B	3
101	C	3
110	D	3
	E	4
	F	4

код = 110

ЕСТЬ!

Ожидаемая длина
кода=3

код +1

Ож.длина+0

Восстановление кодов Хаффмана

КОД = 111

0	A	1
100	B	3
101	C	3
110	D	3
	E	4
	F	4

НЕТ

Ожидаемая длина
кода=3

код $\ll 1$

Ож.длина+1

Восстановление кодов Хаффмана

0	A	1
100	B	3
101	C	3
110	D	3
1110	E	4
	F	4

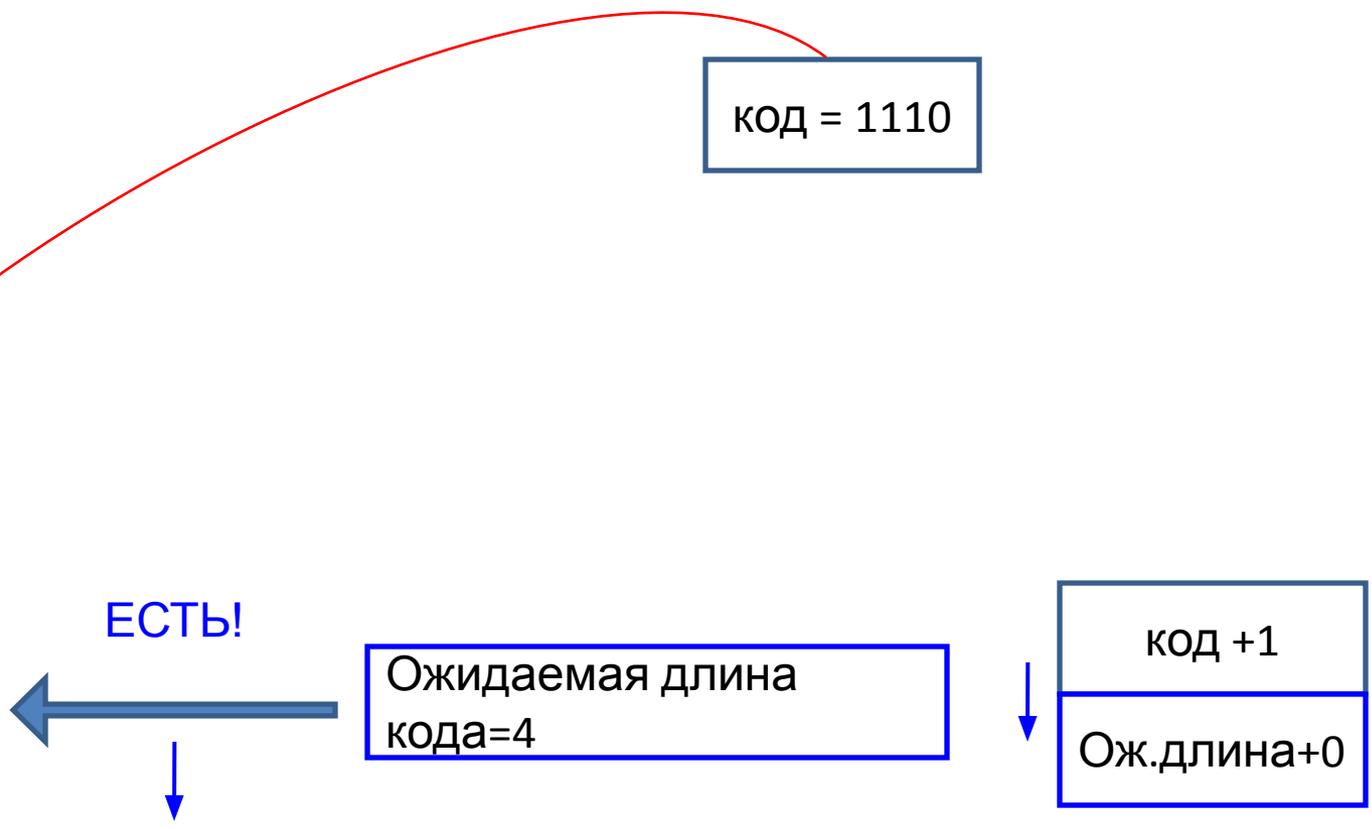
КОД = 1110

ЕСТЬ!

Ожидаемая длина
кода=4

код +1

Ож.длина+0



Восстановление кодов Хаффмана

0	A	1
100	B	3
101	C	3
110	D	3
1110	E	4
1111	F	4

КОД = 1111

ЕСТЬ!

Ожидаемая длина
кода=4

код +1

Ож.длина+0

