

CAP 3. COMPRESSÃO DE DADOS MULTIMÍDIA

Aula 3: Técnicas de Codificação sem perdas - RLE e Codificação de Huffman

INE5431 Sistemas Multimídia
Prof. Roberto Willrich (INE/UFSC)
roberto.willrich@ufsc.br

Cap 3. Compressão de Dados Multimídia

Conteúdo:

- Necessidade de compressão
- Entropia: Teorema da codificação da fonte
- Princípios da compressão
- Classificação das técnicas de compressão
- Medição do desempenho de compressão
- Técnicas de compressão sem perdas
 - RLE, Huffman, (A)DPCM, LZW (GIF)
- Técnicas de compressão de áudio, vídeo e imagens
 - Técnicas de compressão de voz
 - Técnicas de compressão de som
- Padrões de compressão multimídia
 - JPEG, MPEG, MPEG-4, H.261, H.263

Técnicas de compressão sem perdas

Codificação RLE (Codificação por entropia)

- Técnica simples de compressão de dados: dados podem ser compactados através da supressão de sequências de mesmos símbolos
- Aplicação: BMP (RLE)
 - BMP RLE suporta 256 cores
- Um exemplo simples de codificação RLE
 - Codificação: <num.de repetições><valor><num.de repetições><valor>...
 - Número de repetições: 8 bits, até 255 repetições.
 - Original: 12 12 12 12 09 09 09 21 21 23 24 25 25 25 25 25 25 25 25
 - Compactado: 04 12 03 09 02 21 01 23 01 24 08 25

Técnicas de compressão sem perdas

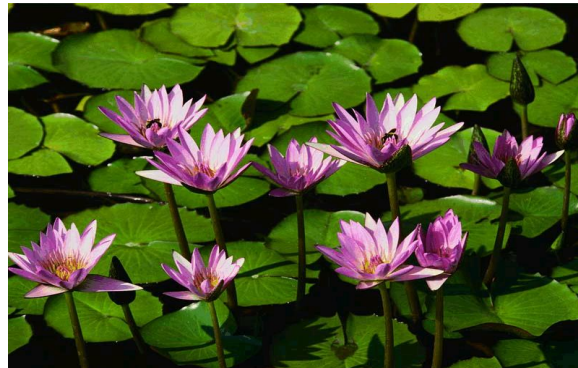
Codificação RLE (Codificação por entropia)







- Tem diversas variações
 - Sequências idênticas são substituídas por um símbolo especial, número de ocorrências e o símbolo repetido
 - Original: UHHHHHHIMMG1223 Compactado: U!6HIMMG1223
 - Se o símbolo especial ocorrer no dado de entrada, ele deve ser substituído por dois símbolos
 - entrada: U!HIIIIID saída: U!!H!5ID
 - Técnica não é utilizada para sequências menores que 4
 - exemplo: U!6HI!2MG1223 (não a compactação)
 - Algoritmo pode ser facilmente otimizado
 - pode-se substituir sequências maiores que um
 - requer que o tamanho da sequência seja codificado ou pode-se usar um caractere especial de fim
 - entrada: UFYUGDUFHUFHUFHUFHUFHBF
 - saída: UFYUGD!5UFH\$BFD

Técnicas de compressão sem perdas

Codificação RLE (Codificação por entropia)

- Fator de compressão depende do dado de entrada
 - Demonstração usando BMP RLE



Nome	Tamanho	Tipo
 Ninféias24bits.bmp	1.407 KB	Imagem de bitmap
 Ninféias8bits.bmp	470 KB	Imagem de bitmap
 Ninféias8bitsRLE.bmp	389 KB	Imagem de bitmap
 Jazz Man24bits.bmp	1.650 KB	Imagem de bitmap
 Jazz Man8bits.bmp	552 KB	Imagem de bitmap
 Jazz Man8bitsRLE.bmp	210 KB	Imagem de bitmap

Técnicas de compressão sem perdas

Codificação Run-Length

- Só traz ganhos relevantes se houver grandes agrupamentos de símbolos iguais
- As principais aplicações são imagens bitmap
 - em imagens com grandes espaços envolvendo uma só cor
 - em imagens geradas por computador
 - onde os dados estão agrupados de forma mais geometricamente definida



Nome	Tamanho	Tipo
Ninféias24bits.bmp	1.407 KB	Imagem de bitmap
Ninféias8bits.bmp	470 KB	Imagem de bitmap
Ninféias8bitsRLE.bmp	389 KB	Imagem de bitmap
Jazz Man24bits.bmp	1.650 KB	Imagem de bitmap
Jazz Man8bits.bmp	552 KB	Imagem de bitmap
Jazz Man8bitsRLE.bmp	210 KB	Imagem de bitmap

Técnicas de compressão sem perdas

Codificação de Huffman (Codificação Estatística)

- Método que atribui menos bits a símbolos que aparecem mais frequentemente e mais bits para símbolos que aparecem menos
- Ideia usada no código de Morse

A	.-	M	--	Y	-.--	6	-....
B	-...	N	-. .	Z	--..	7	--...
C	-.-.	O	---	Ä	.-.-	8	---..
D	-..	P	.-.	Ö	---.	9	----.
E	.	Q	--.-	Ü	..--	.	.-.-.-
F	..-	R	.-.	Ch	----	,	--..--
G	--.	S	...	0	-----	?	..--..
H	T	-	1	.----	!	..---
I	..	U	..-	2	..----	:	---...
J	.---	V	...-	3	...--	"	.-.-.-
K	-. -	W	.- -	4-	'	.----.
L	.-..	X	-.- -	5	=	-...-

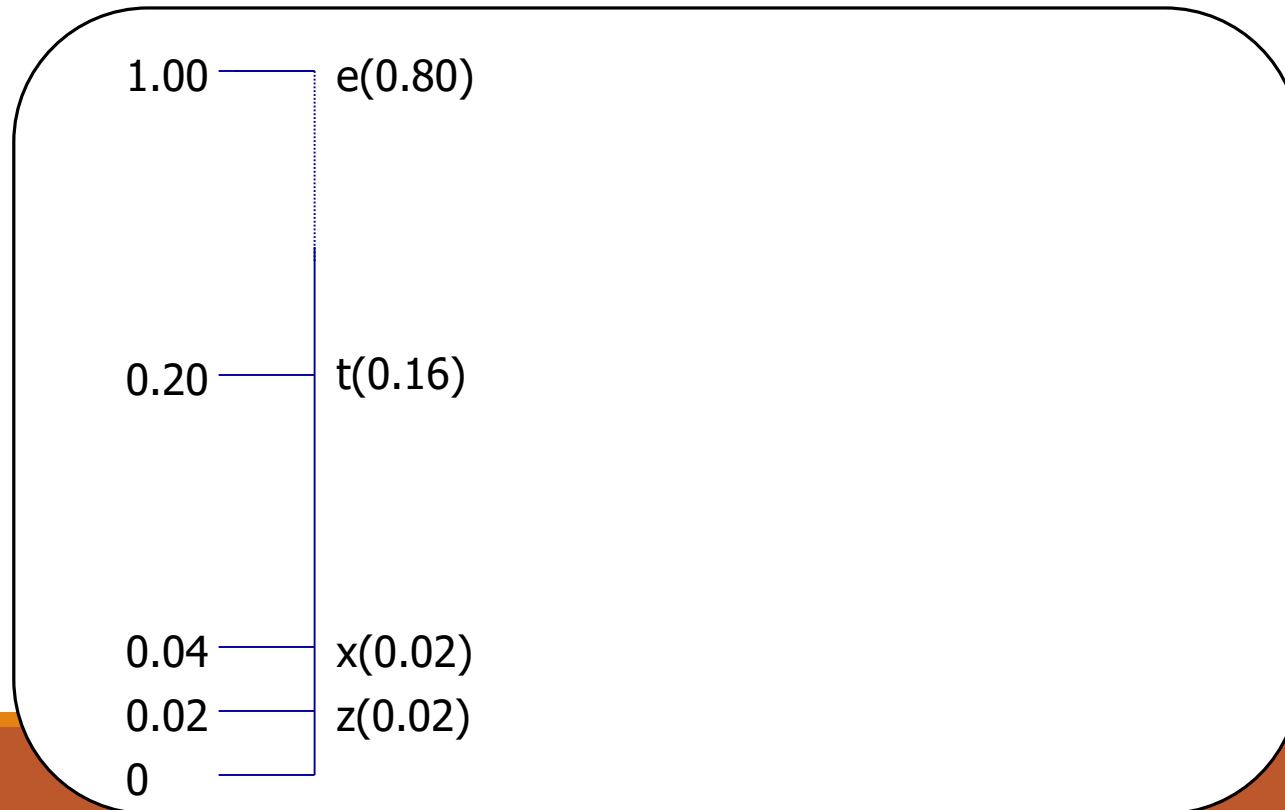
Codificação de Huffman (Codificação Estatística)

- INE5431 SISTEMAS MULTIMÍDIA - PROF. ROBERTO WILLRICH

Técnicas de compressão sem perdas

Codificação de Huffman (Original)

- Geração dos códigos Huffman
 - a) colocação dos símbolos ao longo de uma linha de probabilidade acumulada (probabilidade aumenta de baixo para cima)
 - símbolos de mesma frequência: colocar em qualquer ordem



Codificação de Huffman

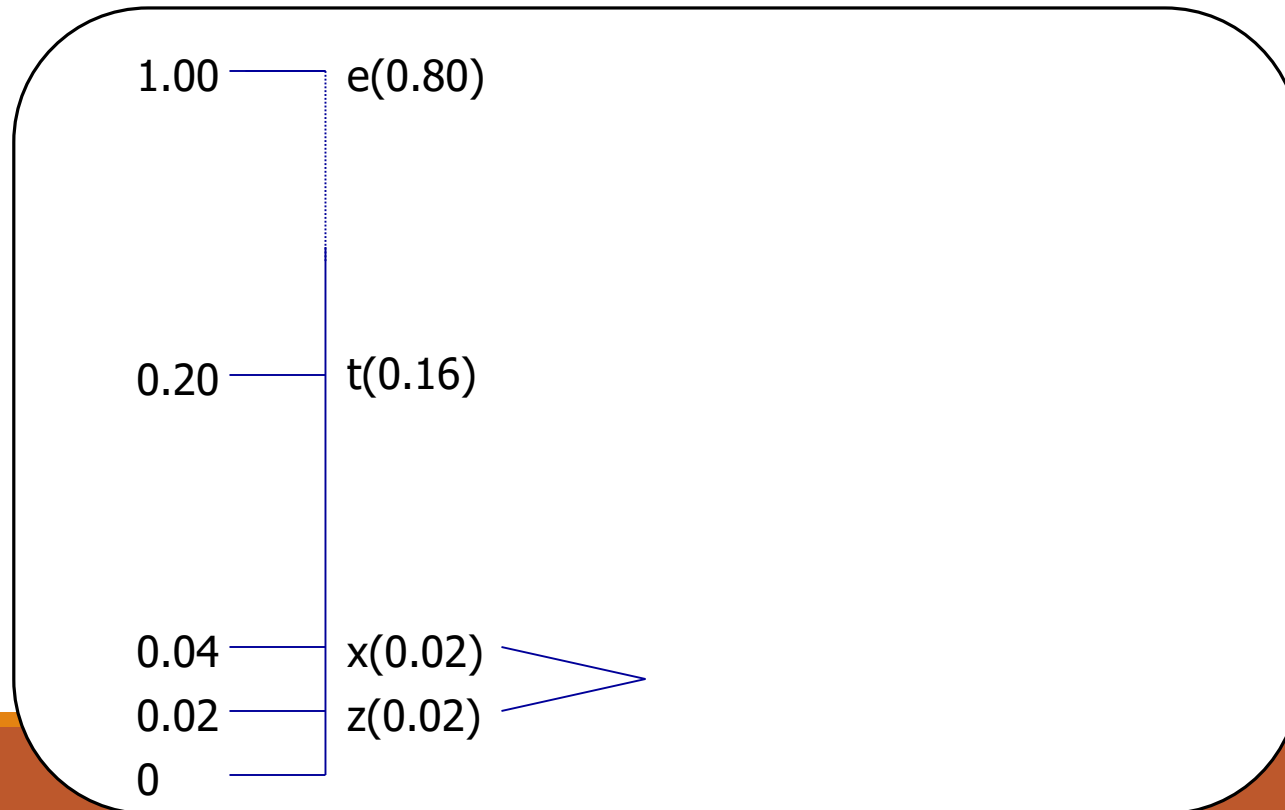
- o b) Junta-se os dois símbolos de menor probabilidade a um nó para formar dois ramos na árvore



Técnicas de compressão sem perdas

Codificação de Huffman

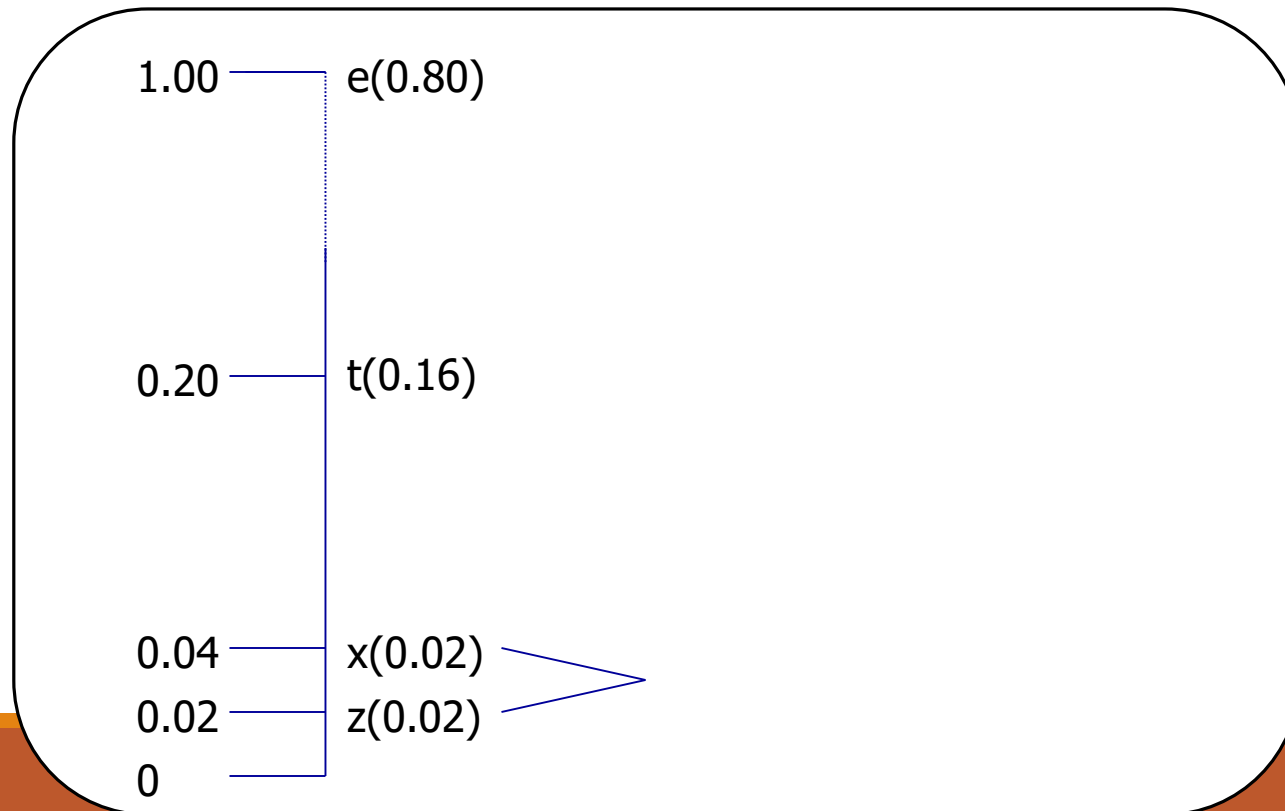
- Geração dos códigos Huffman
 - b) Junta-se os dois símbolos de menor probabilidade a um nó para formar dois ramos na árvore



Técnicas de compressão sem perdas

Codificação de Huffman

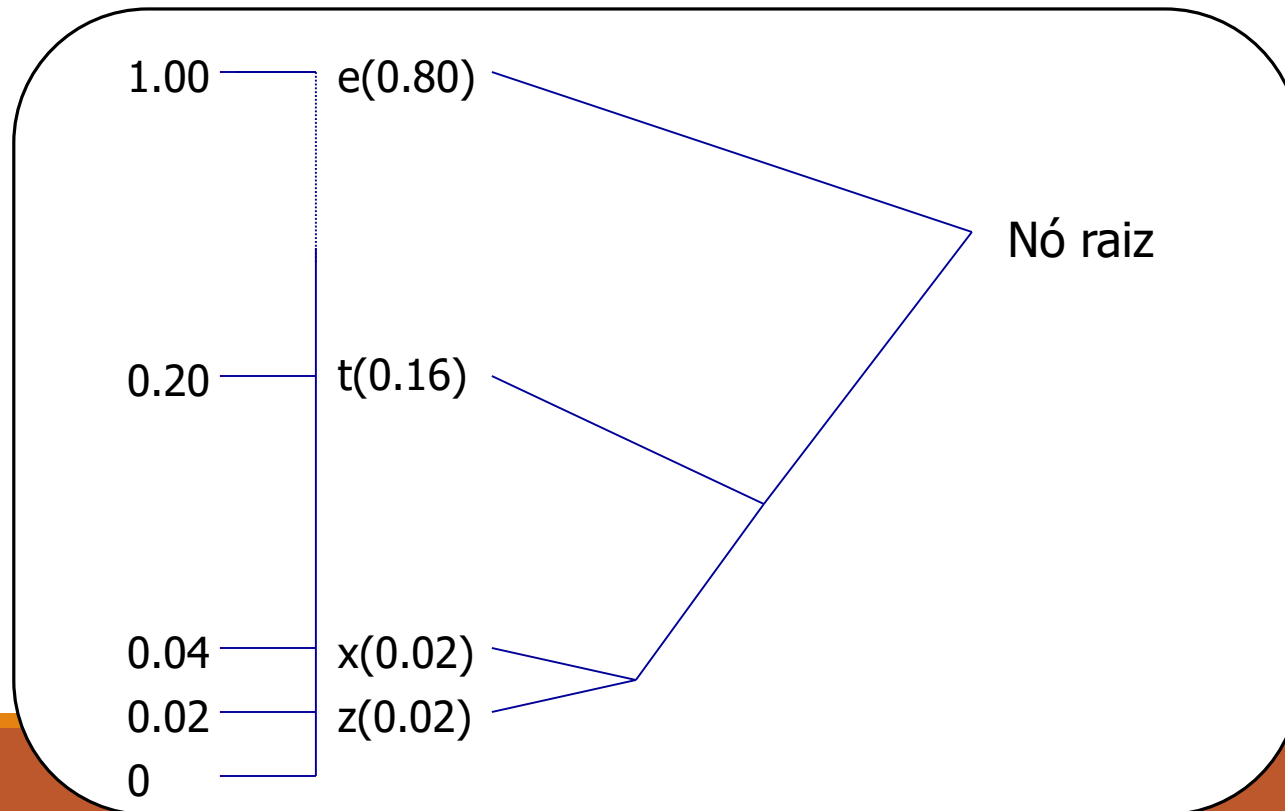
- Geração dos códigos Huffman
 - c) Nova árvore formada é tratada como um símbolo único com a probabilidade igual a soma dos símbolos ramos



Técnicas de compressão sem perdas

Codificação de Huffman

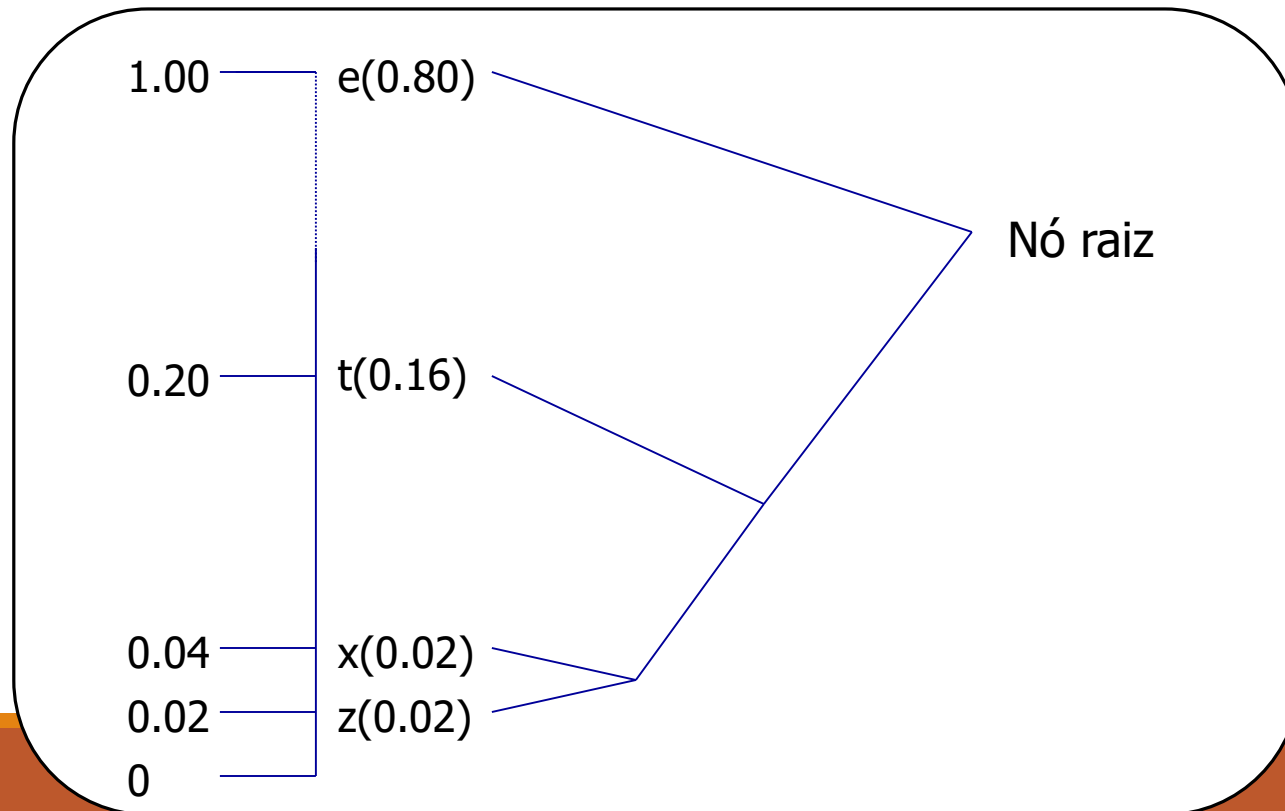
- Geração dos códigos Huffman
 - d) Repita b) e c) até que todos os símbolos sejam inseridos na árvore
 - último nó é chamado de raiz



Técnicas de compressão sem perdas

Codificação de Huffman

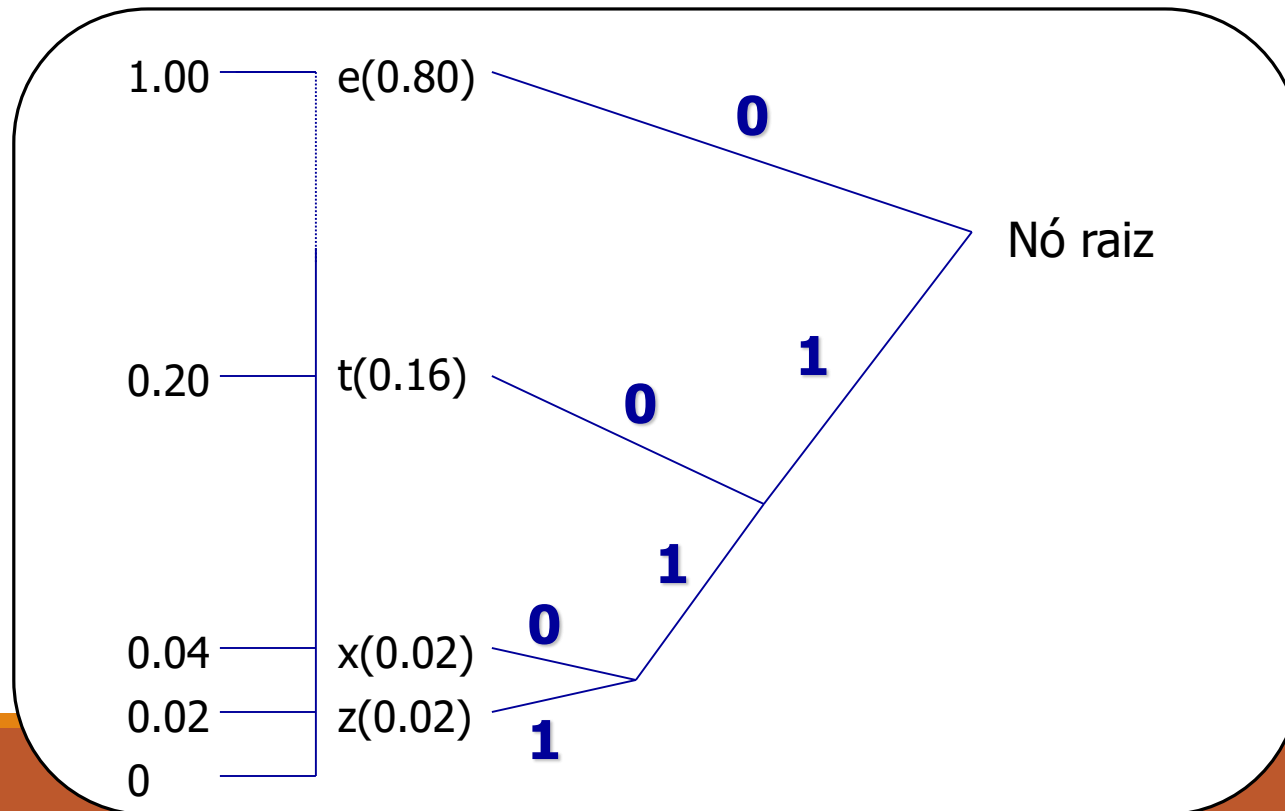
- Geração dos códigos Huffman
 - d) Repita b) e c) até que todos os símbolos sejam inseridos na árvore
 - último nó é chamado de raiz



Técnicas de compressão sem perdas

Codificação de Huffman

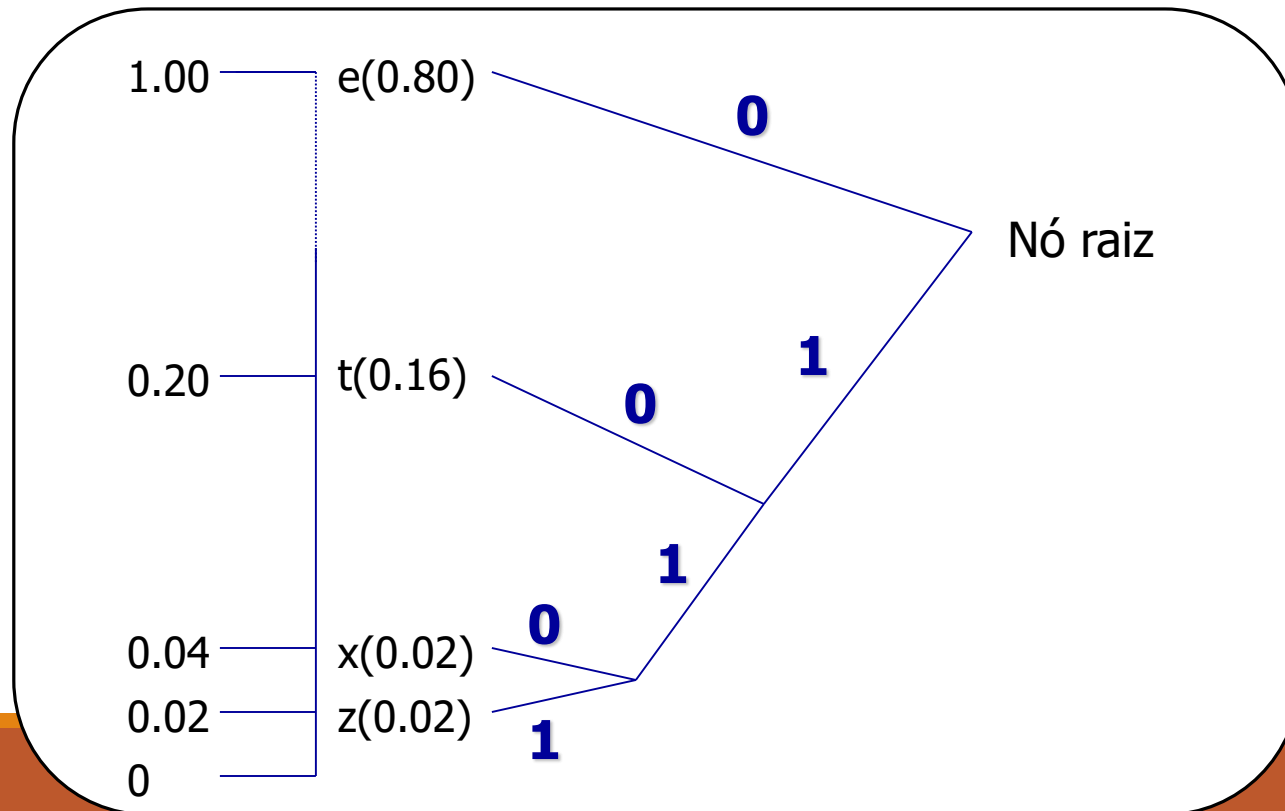
- Geração dos códigos Huffman
 - e) Partindo do nó raiz, atribua bit 0 ao ramo de maior prioridade e bit 1 ao ramo de menor prioridade de cada nó



Técnicas de compressão sem perdas

Codificação de Huffman

- Geração dos códigos Huffman
 - f) Código para cada símbolo é obtido montando códigos ao longo do caminho entre nó raiz ao símbolo



Codebook

Símbolo	Código
e	0
t	10
x	110
z	111

Técnicas de compressão sem perdas

Outro exemplo de codificação de Huffman

- Frequências dos caracteres
- Pode explorar considerando a entropia

$$H(X) = \sum_{i=1}^n p_i \times \log_2 \left(\frac{1}{p_i} \right)$$

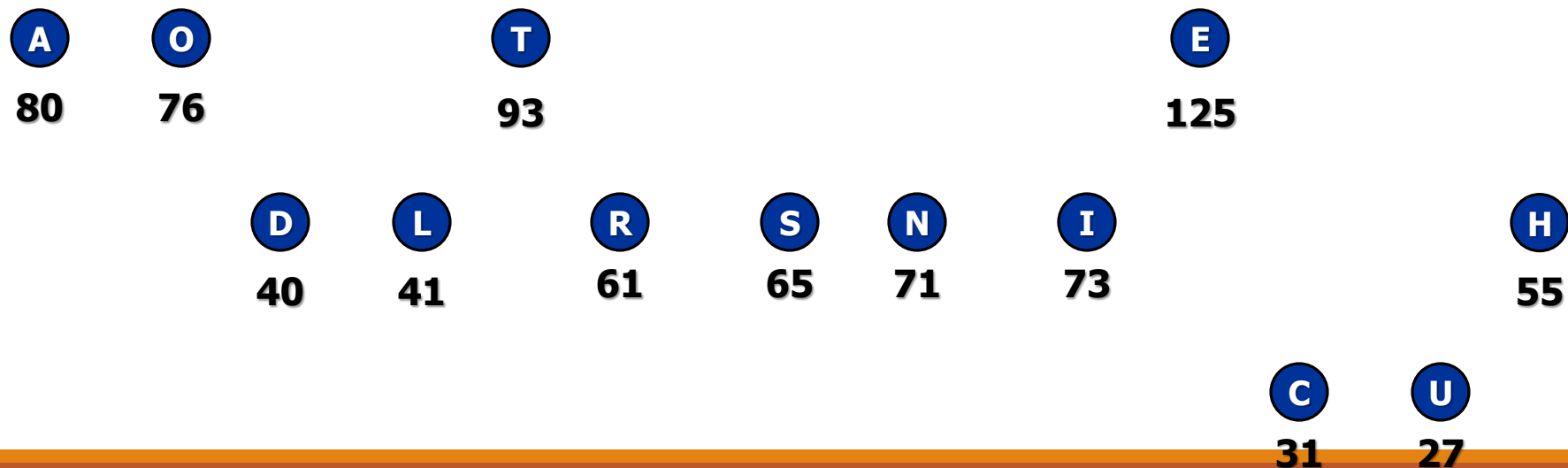
- $H(x) = 3,58$

	f_x	Σ	=	=B1*LOG(1/B1;2)
A	B	C		
125	0,14916468	0,40946034		
93	0,11097852	0,35198476		
80	0,09546539	0,32352061		
76	0,09069212	0,31405584		
73	0,08711217	0,30672038		
71	0,08472554	0,30171266		
65	0,07756563	0,28609607		
61	0,07279236	0,27516016		
55	0,06563246	0,25789925		
41	0,04892601	0,21298739		
40	0,0477327	0,209493		
31	0,03699284	0,17596052		
27	0,03221957	0,15967758		
838	1	3,58472856		

Char	Freq	Fixo
E	125	0000
T	93	0001
A	80	0010
O	76	0011
I	73	0100
N	71	0101
S	65	0110
R	61	0111
H	55	1000
L	41	1001
D	40	1010
C	31	1011
U	27	1100
Total	838	4.00

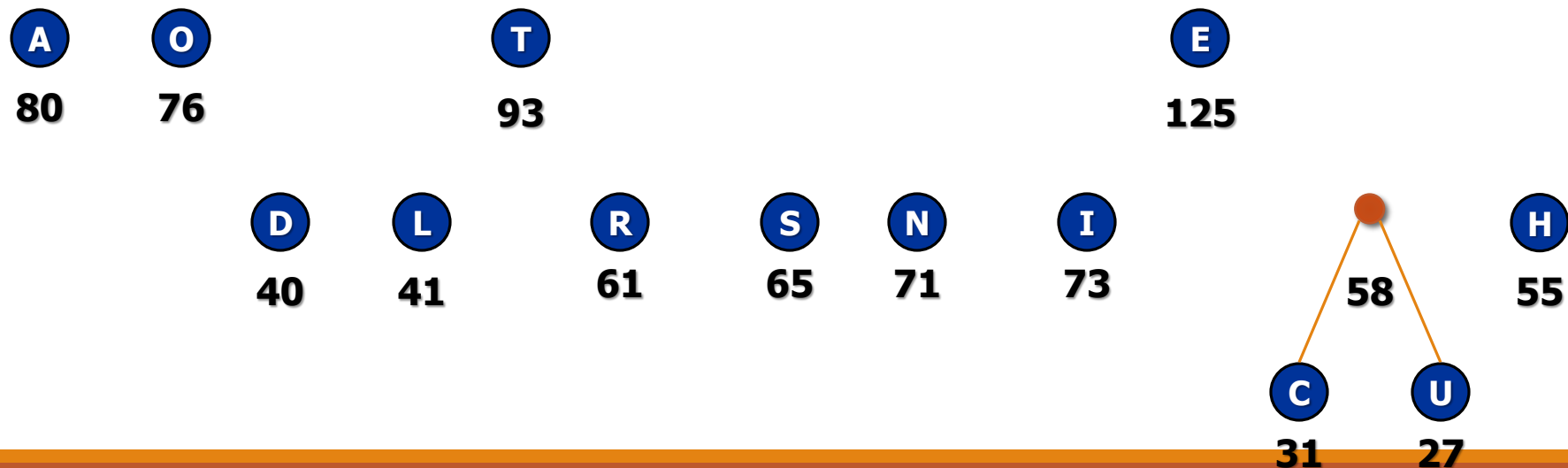
Técnicas de compressão sem perdas

Codificação de Huffman (Geração da tabela)



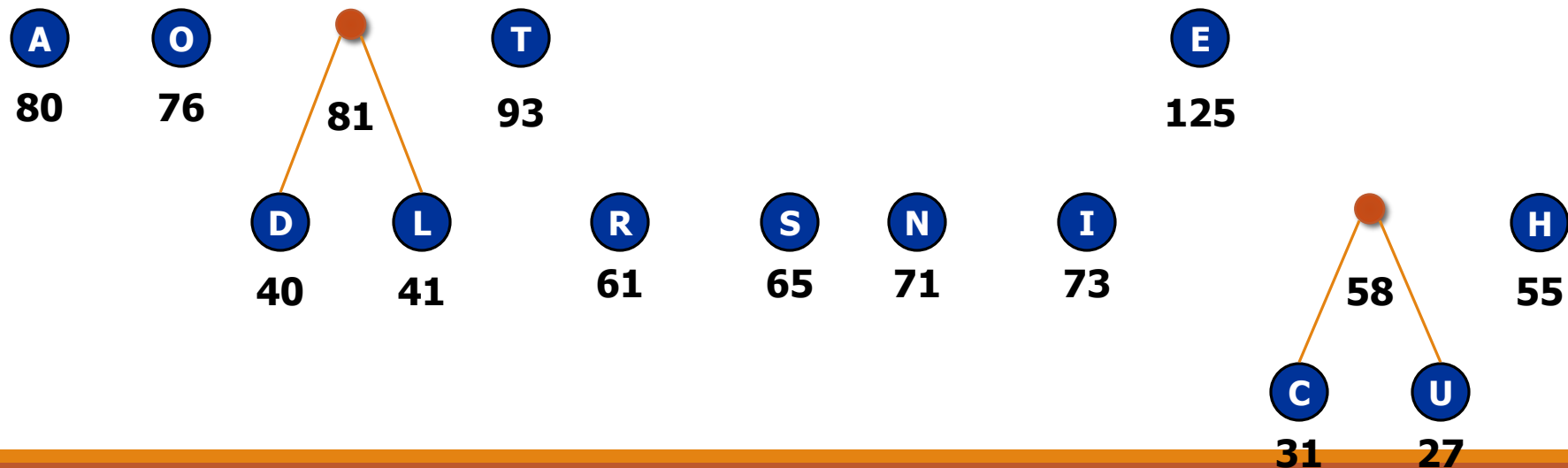
Técnicas de compressão sem perdas

Codificação de Huffman (Geração da tabela)



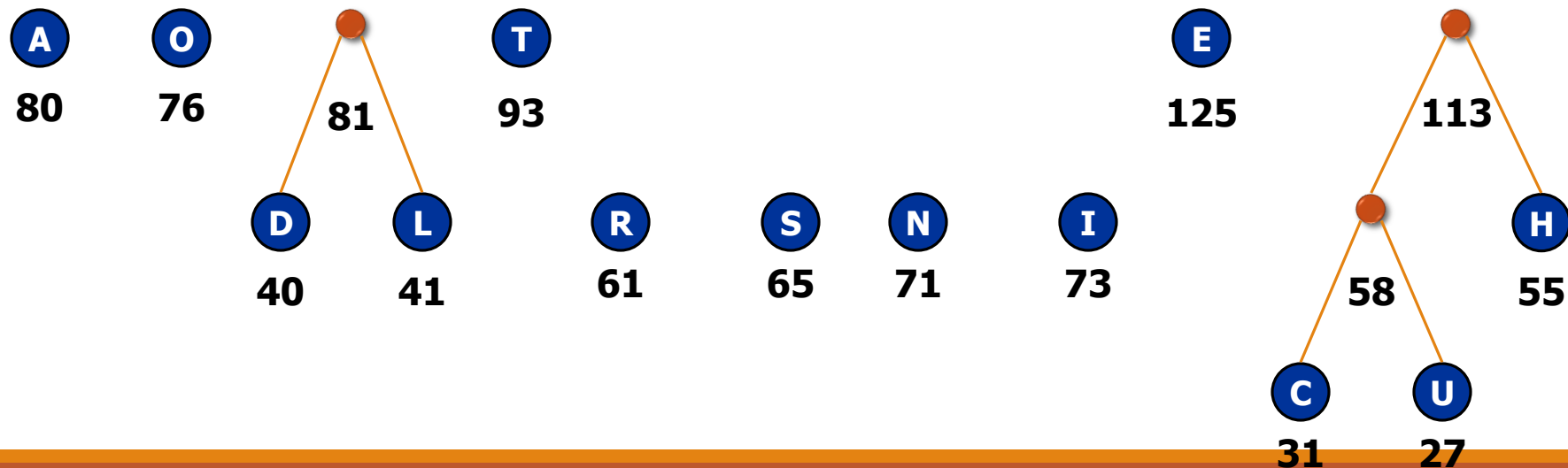
Técnicas de compressão sem perdas

Codificação de Huffman (Geração da tabela)



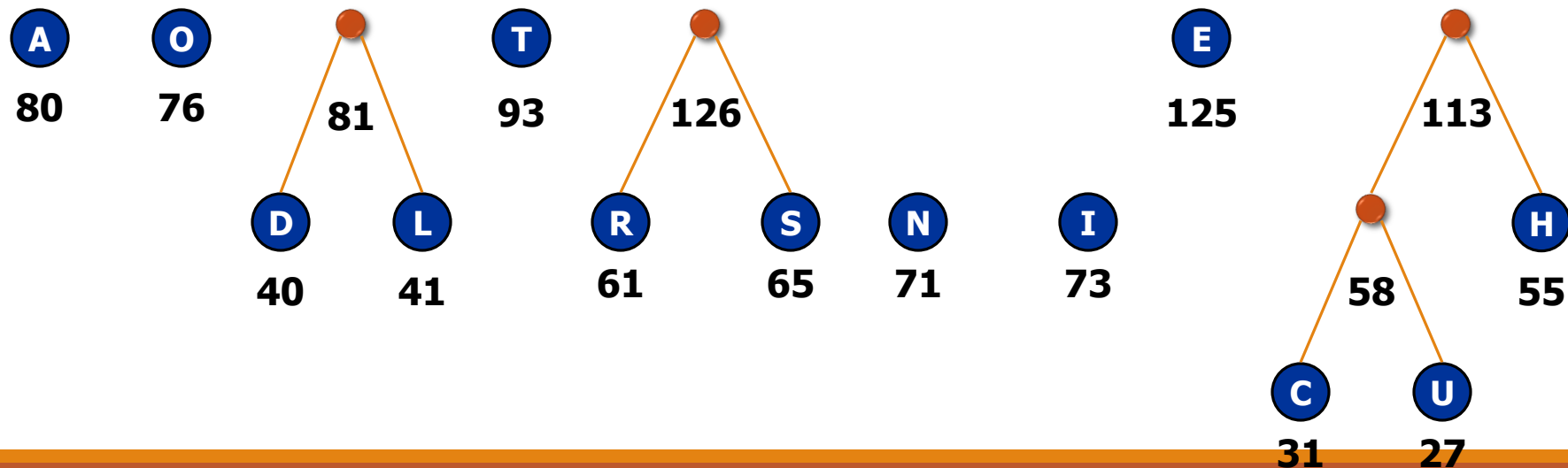
Técnicas de compressão sem perdas

Codificação de Huffman (Geração da tabela)



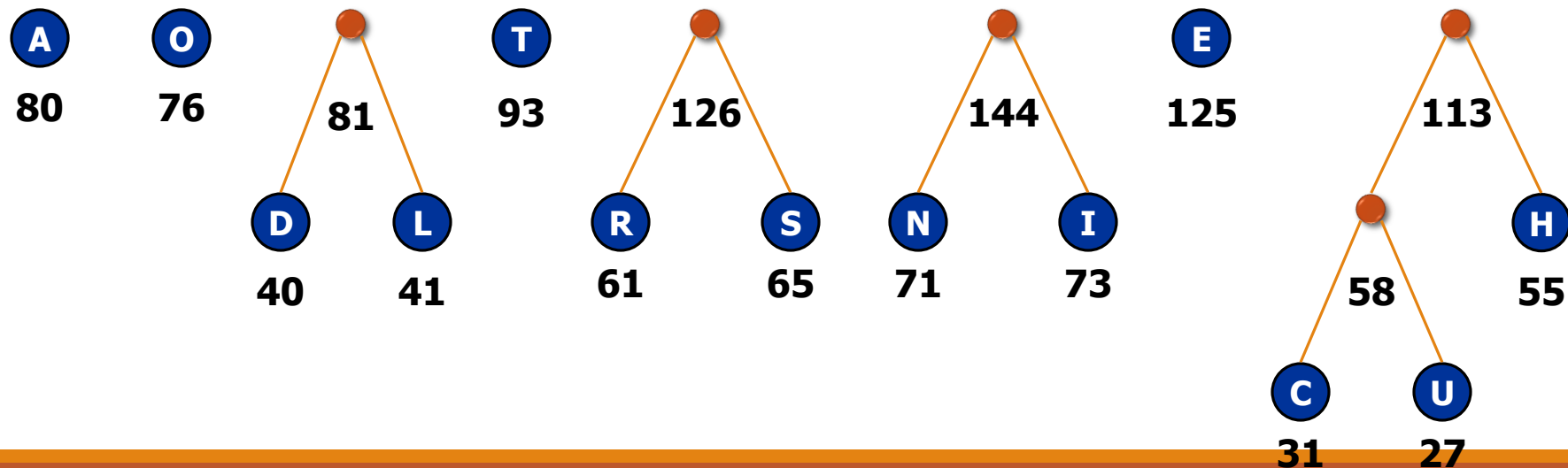
Técnicas de compressão sem perdas

Codificação de Huffman (Geração da tabela)



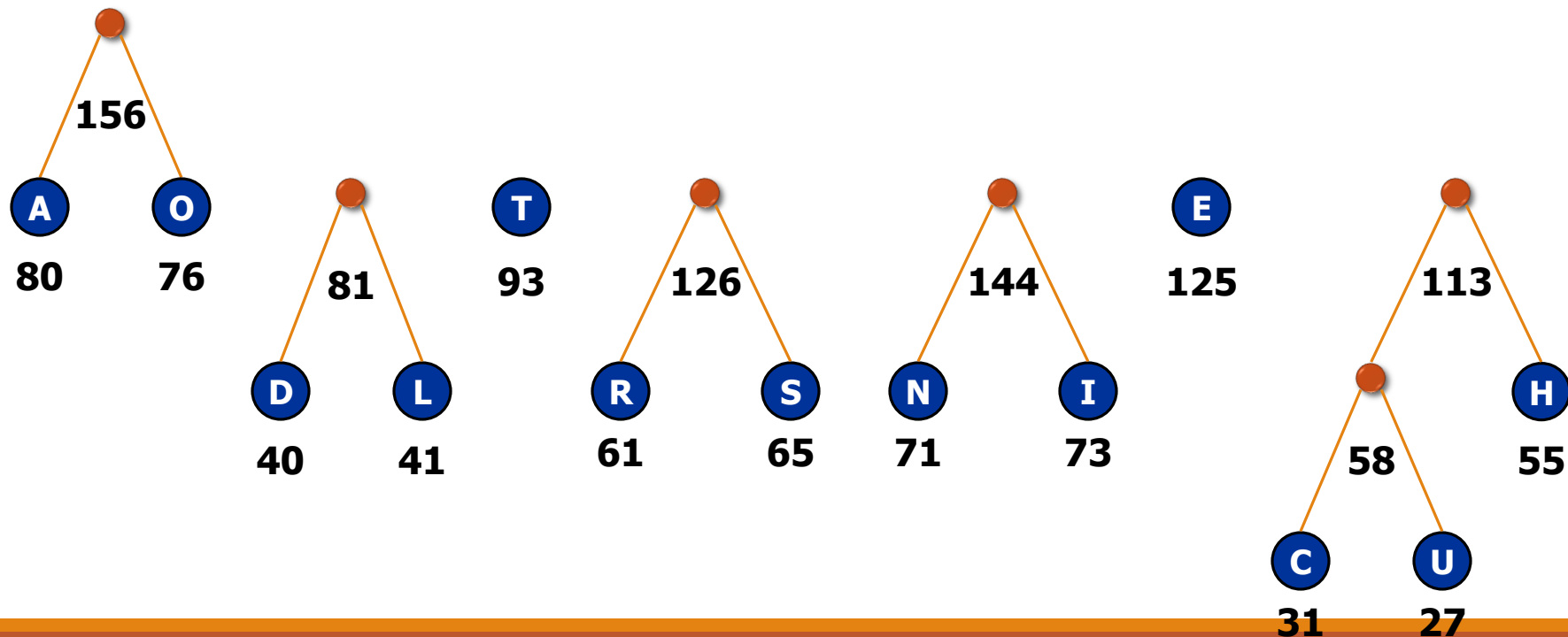
Técnicas de compressão sem perdas

Codificação de Huffman (Geração da tabela)



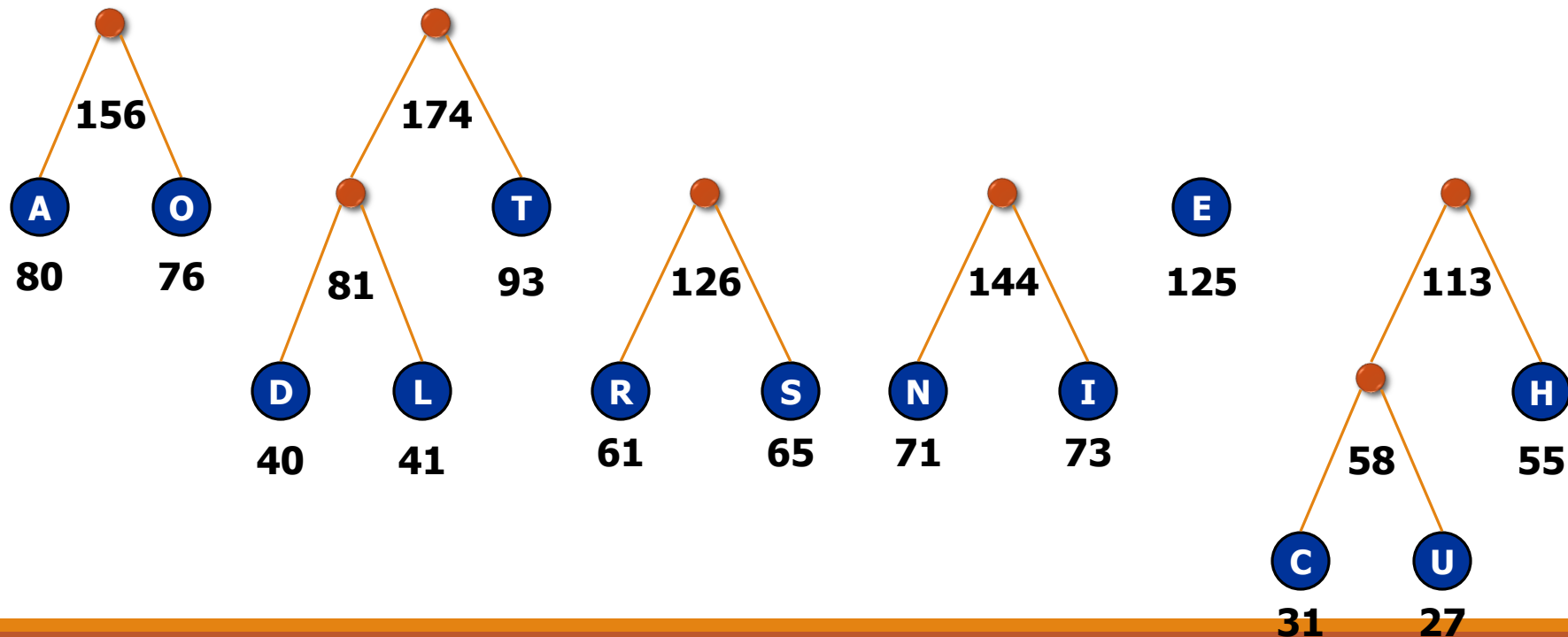
Técnicas de compressão sem perdas

Codificação de Huffman (Geração da tabela)



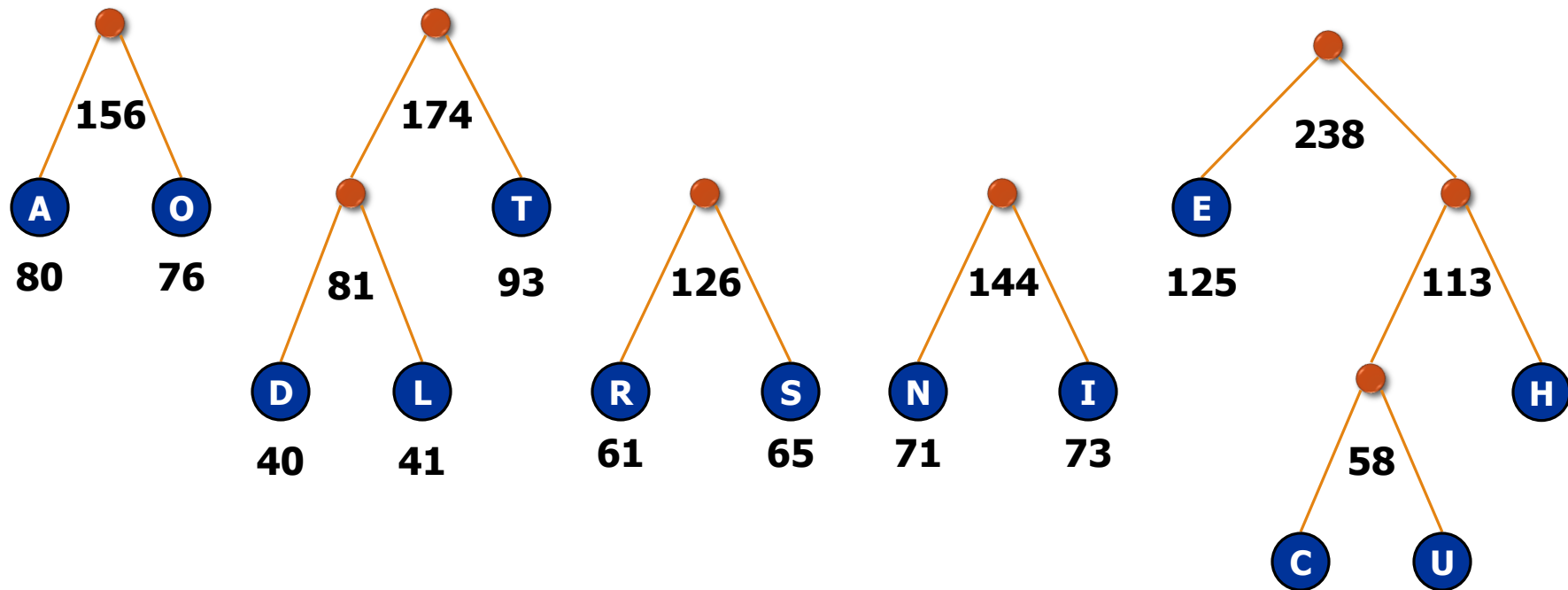
Técnicas de compressão sem perdas

Codificação de Huffman (Geração da tabela)



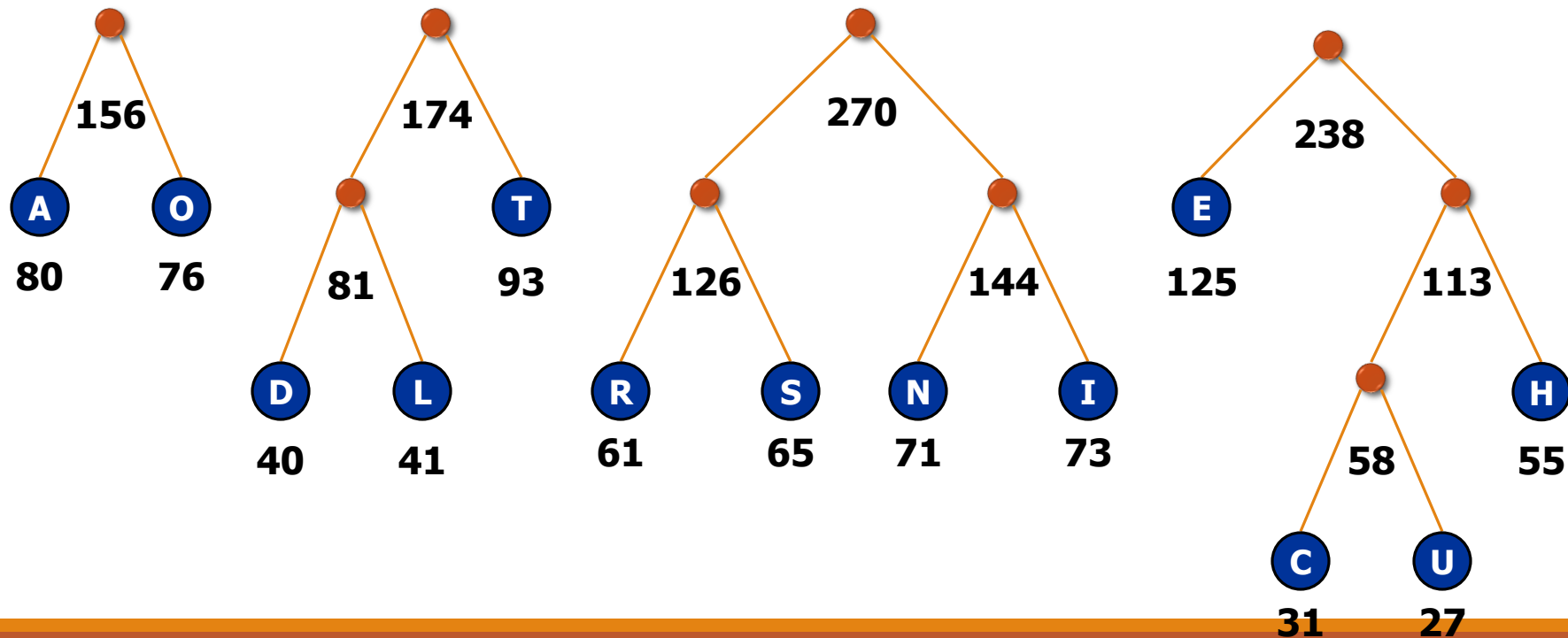
Técnicas de compressão sem perdas

Codificação de Huffman (Geração da tabela)



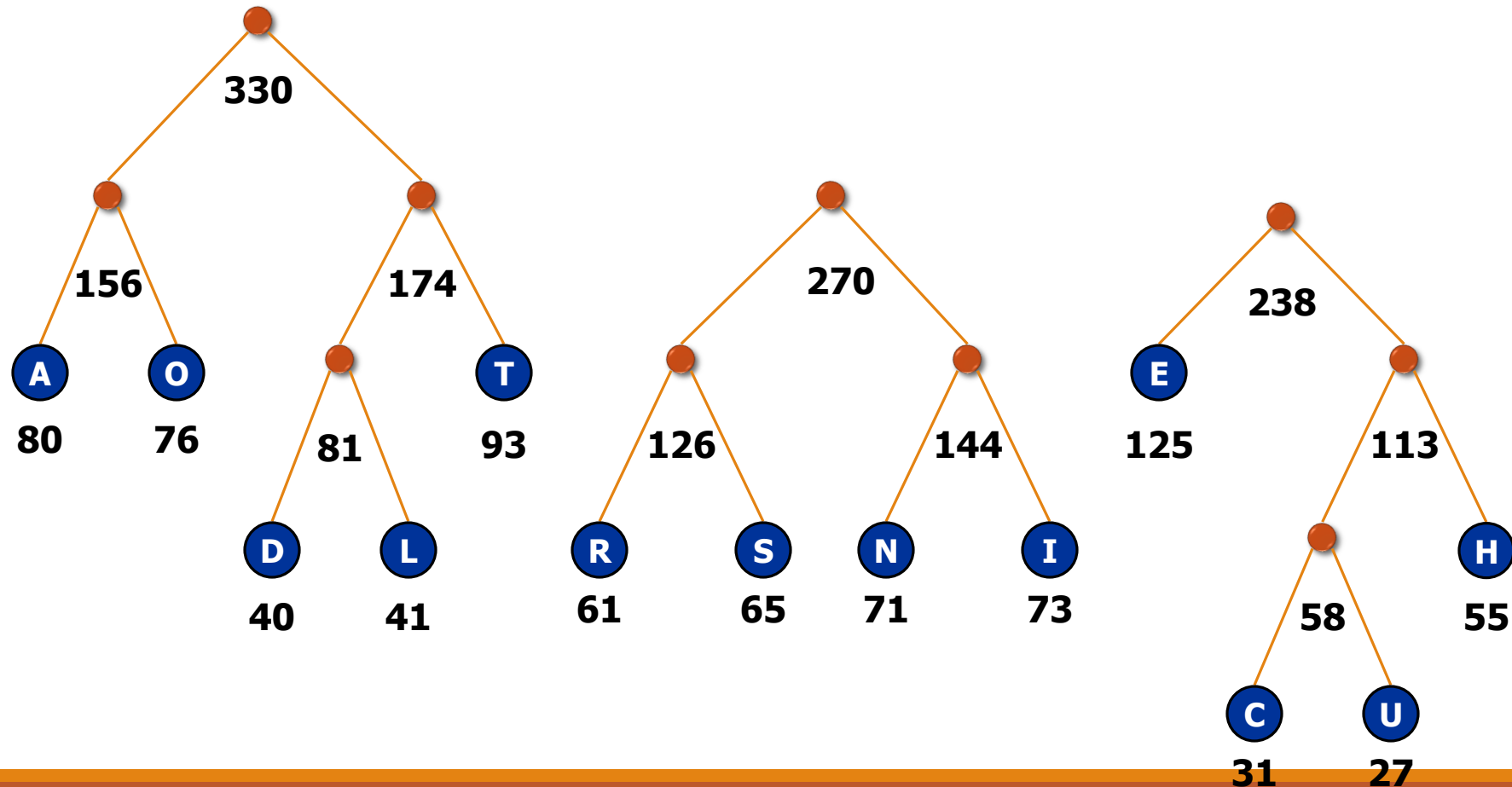
Técnicas de compressão sem perdas

Codificação de Huffman (Geração da tabela)



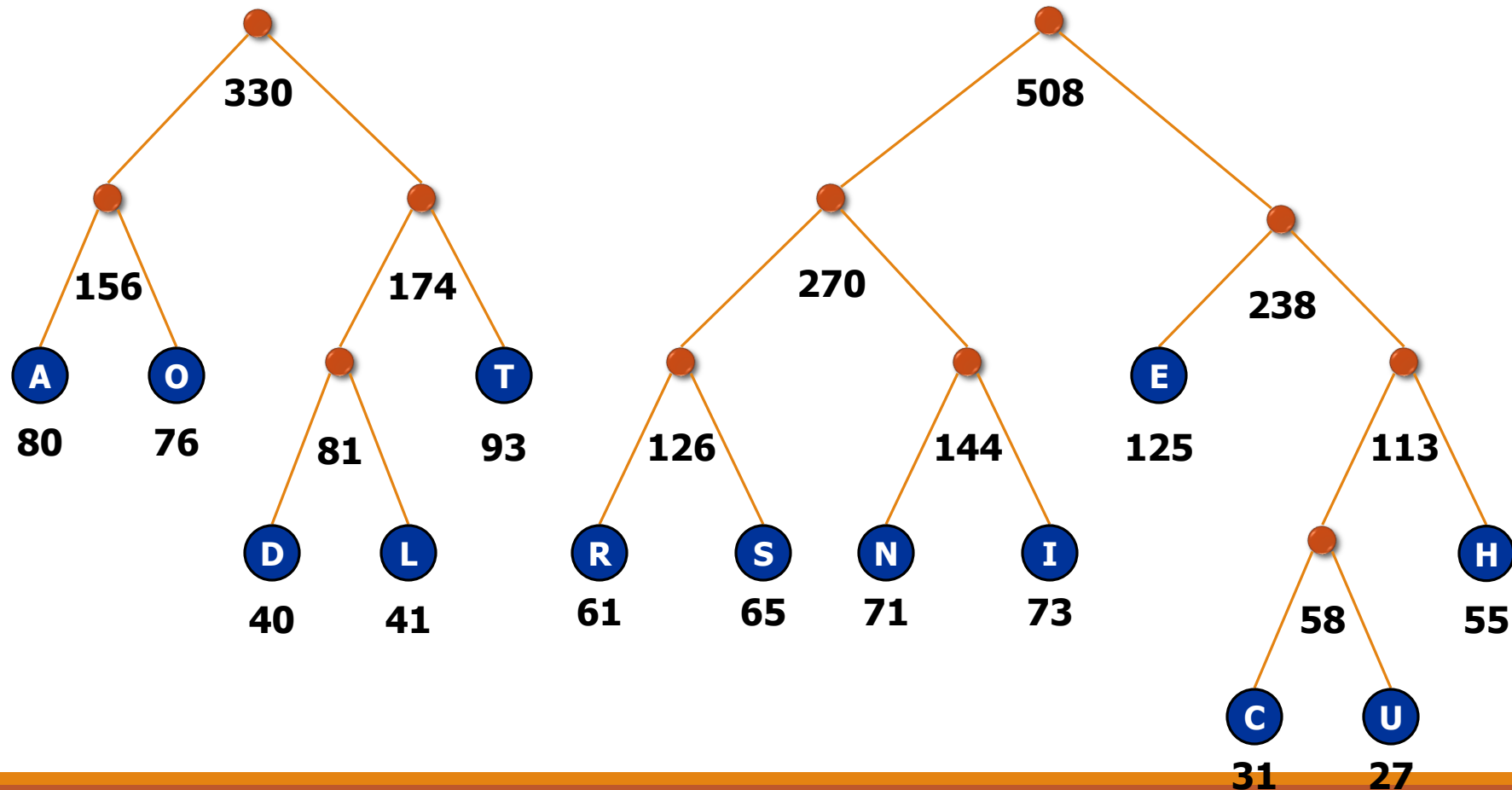
Técnicas de compressão sem perdas

Codificação de Huffman (Geração da tabela)

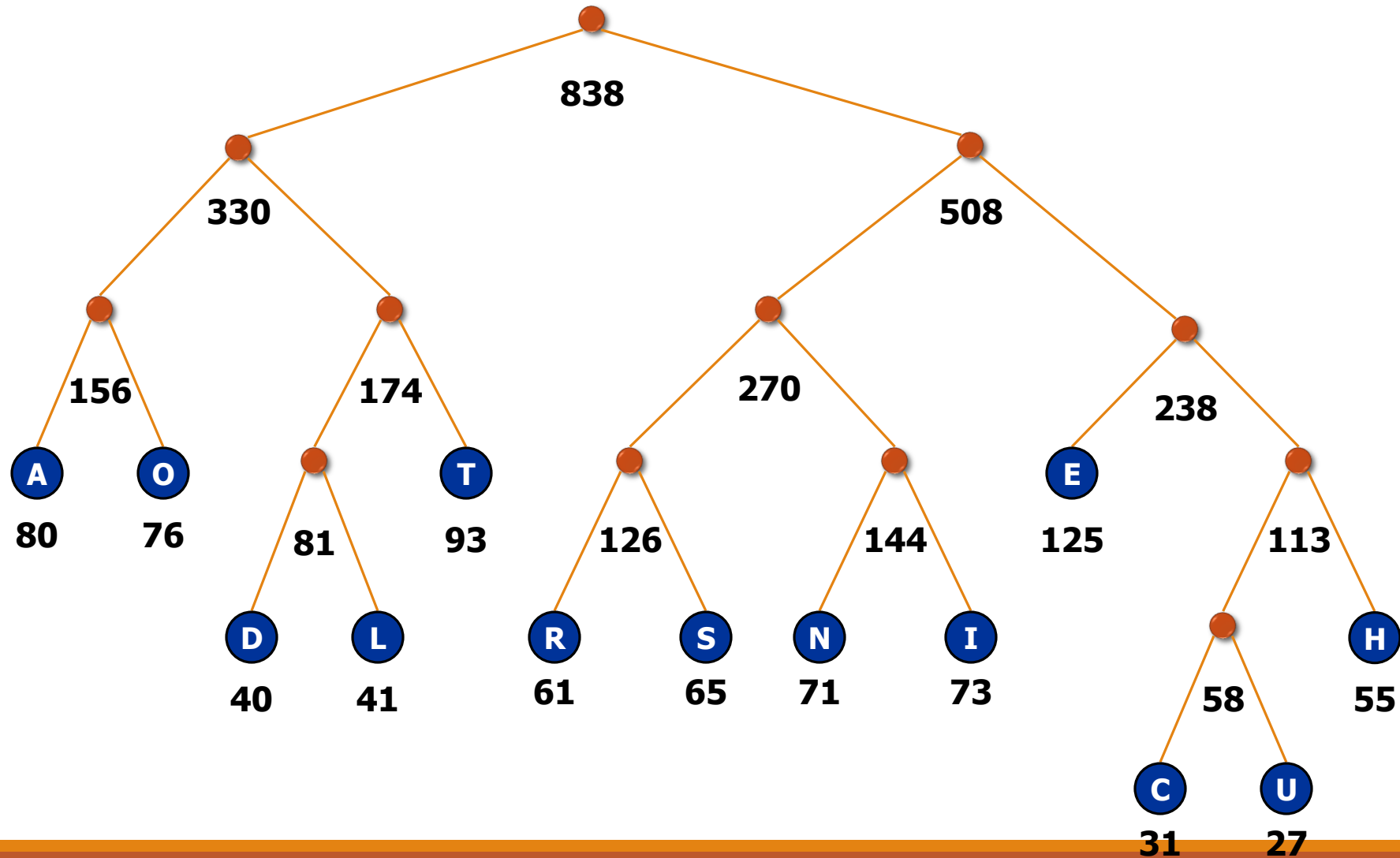


Técnicas de compressão sem perdas

Codificação de Huffman (Geração da tabela)

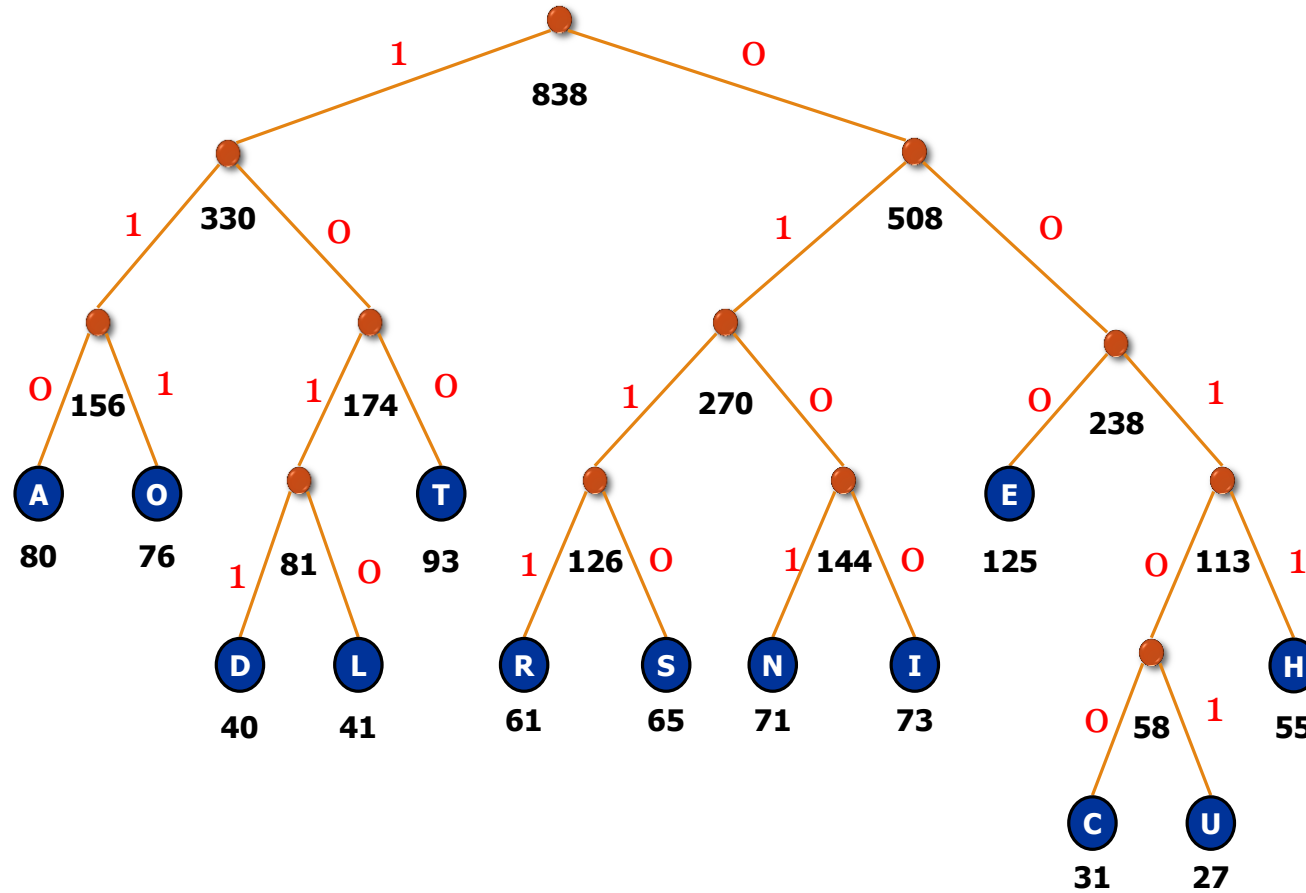


Técnicas de compressão sem perdas



Técnicas de compressão sem perdas

Codificação de Huffman (Geração da tabela)

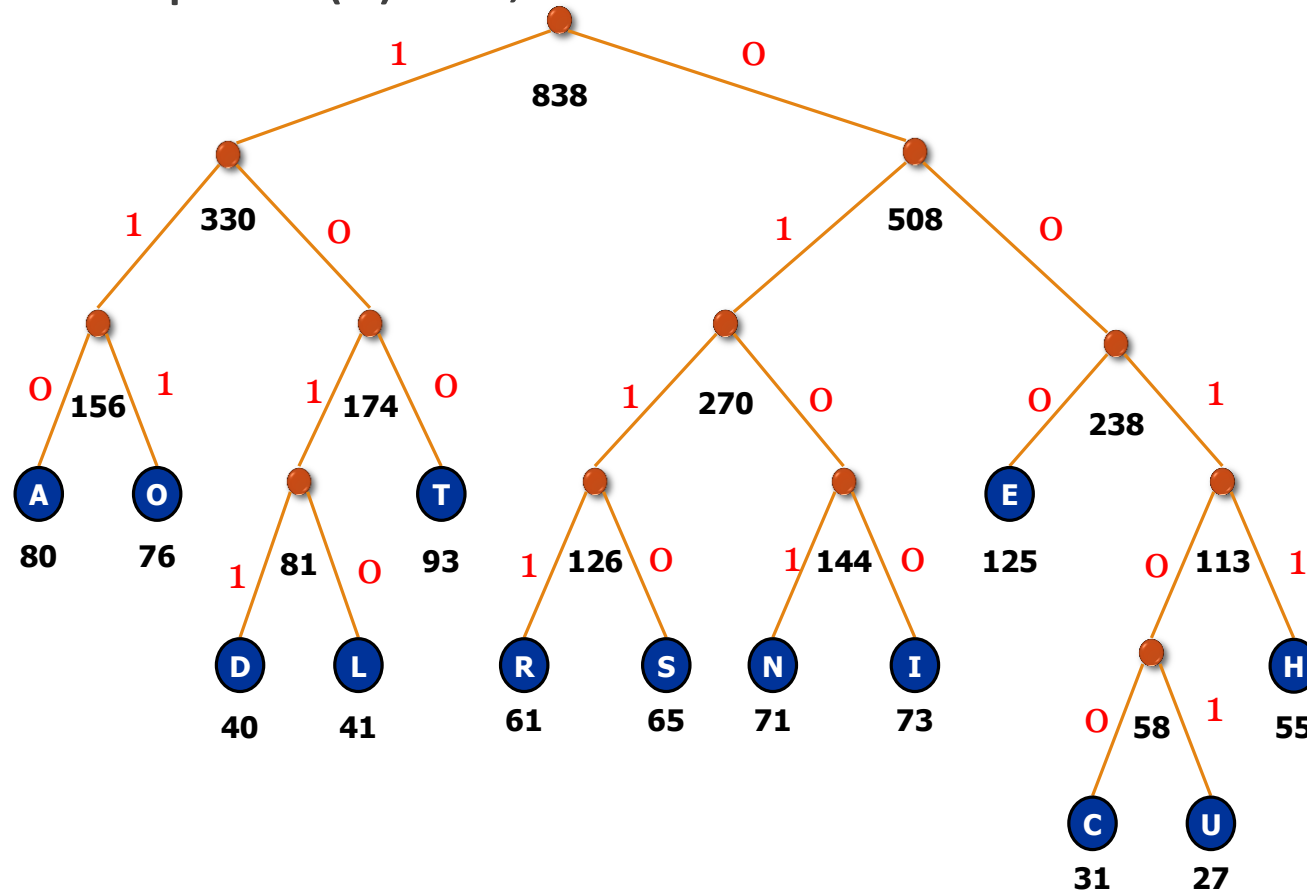


Char	Freq	Fixo
E	125	0000
T	93	0001
A	80	0010
O	76	0011
I	73	0100
N	71	0101
S	65	0110
R	61	0111
H	55	1000
L	41	1001
D	40	1010
C	31	1011
U	27	1100
Total	838	4.00

Técnicas de compressão sem perdas

Codificação de Huffman (Geração da tabela)

- Entropia $H(x) = 3,58$



Char	Freq	Fixo	Huff
E	125	0000	000
T	93	0001	100
A	80	0010	110
O	76	0011	111
I	73	0100	0100
N	71	0101	0101
S	65	0110	0110
R	61	0111	0111
H	55	1000	0011
L	41	1001	1010
D	40	1010	1011
C	31	1011	00100
U	27	1100	00101
Média		4.00	3.62
Total	383	3352	3036

Técnicas de compressão sem perdas

Codificação de Huffman

- Operação computacional mais custosa na codificação
- No decodificador
 - realiza uma simples verificação na tabela de Huffman
 - tabela de Huffman é parte do fluxo de dados ou é conhecida pelo decodificador
- Tabelas de Huffman padrões são muito usadas
 - usada para vídeo em tempo-real
 - tabelas são conhecidas pelo codificador e decodificador
 - codificação e decodificação são mais rápidas
 - desvantagem: tabelas padrões obtém fator de compressão um pouco menores
 - elas não são necessariamente ótimas

Técnicas de compressão sem perdas


Huffman - otimalidade

- Huffman é ótimo para codificação símbolo-a-símbolo com uma distribuição de probabilidade conhecida, porém como trabalha com números binário inteiros há algumas redundâncias.
- Ainda assim, é garantido que:
 - $H(X) \leq Huffman(X) \leq H(X) + 1$

Entropia



Média de bits por símbolo após a
codificação por Huffman



Técnicas de compressão sem perdas

Outros métodos

- É possível melhorar ainda mais a codificação de Huffman
 - Huffman adaptativo:
 - Constrói a árvore dinamicamente
 - Cálculo das probabilidades são dinâmicas com base nas frequências recentes na sequência de símbolos, e altera a estrutura da árvore para atualizar probabilidades estimadas.
- Estado-da-arte: Codificação aritmética!

Pontos Importantes

RLE e Codif. de Huffman

- Entender o princípio geral, vantagens e desvantagens