

# Codifica di Huffman

Claudio Mirolo

Dipartimento di Scienze Matematiche, Informatiche e Fisiche  
Università di Udine, via delle Scienze 206 – Udine

[claudio.mirolo@uniud.it](mailto:claudio.mirolo@uniud.it)



Campus Comprensione e Comprensione

14–18 Giugno 2021

*La definizione di informatica che prediligo è che si tratta dello studio degli algoritmi [...].*

*Un programma è una particolare rappresentazione di un algoritmo, esattamente nello stesso senso in cui il termine "dato" si riferisce a una particolare rappresentazione di "informazione".*

*Forse la scoperta più significativa determinata dall'avvento del computer sarà che gli algoritmi, come oggetti di studio, sono straordinariamente ricchi di proprietà interessanti; inoltre, un punto di vista algoritmico costituisce un modo utile di organizzare la conoscenza in generale.*



*La definizione di informatica che prediligo è che si tratta dello studio degli algoritmi [...].*

*Un programma è una particolare rappresentazione di un algoritmo, esattamente nello stesso senso in cui il termine “dato” si riferisce a una particolare rappresentazione di “informazione”.*

*Forse la scoperta più significativa determinata dall'avvento del computer sarà che gli algoritmi, come oggetti di studio, sono straordinariamente ricchi di proprietà interessanti; inoltre, un punto di vista algoritmico costituisce un modo utile di organizzare la conoscenza in generale.*

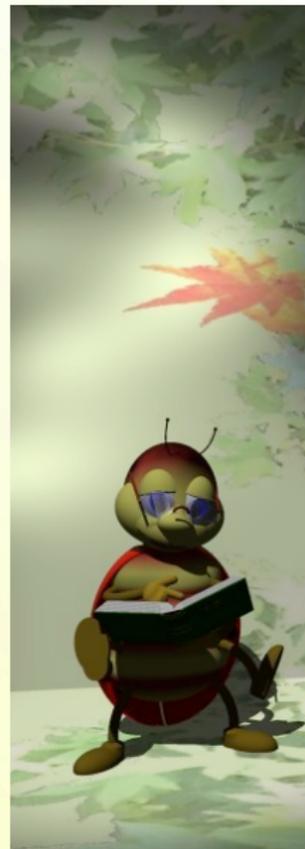


*La definizione di informatica che prediligo è che si tratta dello studio degli algoritmi [...].*

*Un programma è una particolare rappresentazione di un algoritmo, esattamente nello stesso senso in cui il termine “dato” si riferisce a una particolare rappresentazione di “informazione”.*

*Forse la scoperta più significativa determinata dall'avvento del computer sarà che gli algoritmi, come oggetti di studio, sono straordinariamente ricchi di proprietà interessanti;*

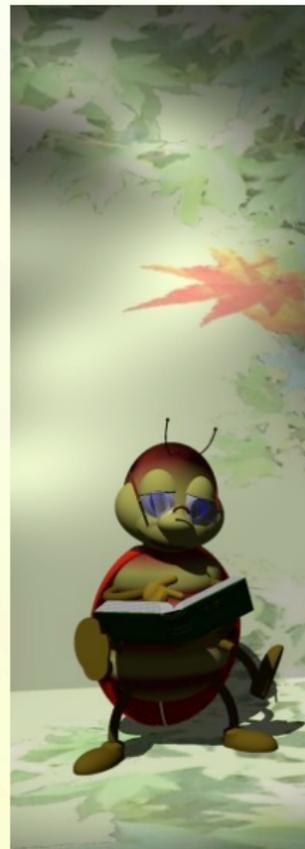
*inoltre, un punto di vista algoritmico costituisce un modo utile di organizzare la conoscenza in generale.*



*La definizione di informatica che prediligo è che si tratta dello studio degli algoritmi [...].*

*Un programma è una particolare rappresentazione di un algoritmo, esattamente nello stesso senso in cui il termine “dato” si riferisce a una particolare rappresentazione di “informazione”.*

*Forse la scoperta più significativa determinata dall'avvento del computer sarà che gli algoritmi, come oggetti di studio, sono straordinariamente ricchi di proprietà interessanti; inoltre, un punto di vista algoritmico costituisce un modo utile di organizzare la conoscenza in generale.*



# Codifica di Huffman: Inventarsi nuovo spazio...



# Codifica di Huffman: Esempio

- Sequenza di 12 basi nucleiche (pretesto):

**A T T C T A C C T T G T**

- Codifica binaria standard (stile ASCII)  
— due bit per distinguere fra quattro simboli:

A	00
T	01
C	10
G	11

- Rappresentazione della sequenza (24 bit):

0 0 0 1 0 1 1 0 0 1 0 0 1 0 1 0 0 1 0 1 1 1 0 1

# Codifica di Huffman: Esempio

- Sequenza di 12 basi nucleiche (pretesto):

A T T C T A C C T T G T

- Codifica binaria standard (stile ASCII)  
— due bit per distinguere fra quattro simboli:

A	00
T	01
C	10
G	11

- Rappresentazione della sequenza (24 bit):

000101100100101001011101

# Codifica di Huffman: Esempio

- Sequenza di 12 basi nucleiche (pretesto):

**A T T C T A C C T T G T**

- Codifica binaria standard (stile ASCII)  
— due bit per distinguere fra quattro simboli:

<b>A</b>	00
<b>T</b>	01
<b>C</b>	10
<b>G</b>	11

- Rappresentazione della sequenza (24 bit):

0 0 0 1 0 1 1 0 0 1 0 0 1 0 1 0 0 1 0 1 1 1 0 1

# Codifica di Huffman: Esempio

- Sequenza di 12 basi nucleiche (pretesto):

**A T T C T A C C T T G T**

- Codifica binaria standard (stile ASCII)  
— due bit per distinguere fra quattro simboli:

<b>A</b>	00
<b>T</b>	01
<b>C</b>	10
<b>G</b>	11

- Rappresentazione della sequenza (24 bit):

0 0 0 1 0 1 1 0 0 1 0 0 1 0 1 0 0 1 0 1 1 1 0 1

# Codifica di Huffman: Esempio

- *Peso* dei simboli = numero di occorrenze (o frequenza)  
coppie simbolo/peso ordinate per peso crescente:

< **G** : 1, **A** : 2, **C** : 3, **T** : 6 >

- Costruzione dell'albero di Huffman — aggregazione dei 2 elementi di minor peso e riordinamento:

 <  $\left\langle \begin{array}{l} \swarrow \searrow \\ \text{G} \quad \text{A} \end{array} : 1+2 = 3, \text{C} : 3, \text{T} : 6 \right\rangle$

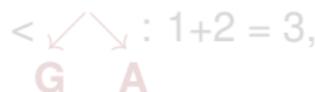
 <  $\left\langle \begin{array}{l} \swarrow \searrow \\ \swarrow \searrow \\ \text{G} \quad \text{A} \quad \text{C} \end{array} : 3+3 = 6, \text{T} : 6 \right\rangle$

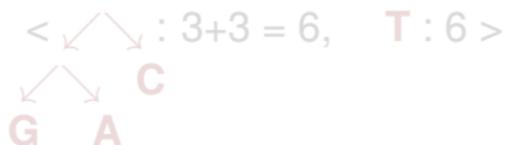
# Codifica di Huffman: Esempio

- *Peso* dei simboli = numero di occorrenze (o frequenza)  
coppie simbolo/peso ordinate per peso crescente:

< **G** : 1, **A** : 2, **C** : 3, **T** : 6 >

- Costruzione dell'albero di Huffman — aggregazione dei 2 elementi di minor peso e riordinamento:

<  : 1+2 = 3, **C** : 3, **T** : 6 >

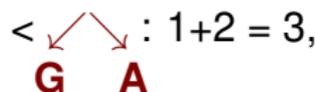
<  : 3+3 = 6, **T** : 6 >

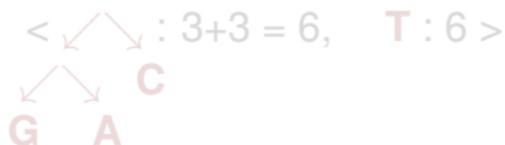
# Codifica di Huffman: Esempio

- *Peso* dei simboli = numero di occorrenze (o frequenza)  
coppie simbolo/peso ordinate per peso crescente:

< **G** : 1, **A** : 2, **C** : 3, **T** : 6 >

- Costruzione dell'albero di Huffman — aggregazione dei 2 elementi di minor peso e riordinamento:

<  : 1+2 = 3, **C** : 3, **T** : 6 >

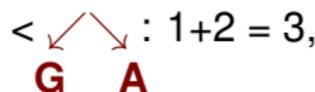
<  : 3+3 = 6, **T** : 6 >

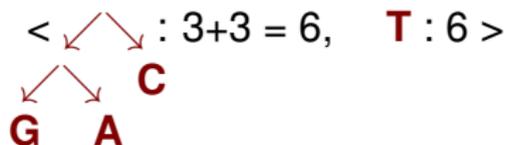
# Codifica di Huffman: Esempio

- *Peso* dei simboli = numero di occorrenze (o frequenza)  
coppie simbolo/peso ordinate per peso crescente:

< **G** : 1, **A** : 2, **C** : 3, **T** : 6 >

- Costruzione dell'albero di Huffman — aggregazione dei 2 elementi di minor peso e riordinamento:

<  : 1+2 = 3, **C** : 3, **T** : 6 >

<  : 3+3 = 6, **T** : 6 >

# Codifica di Huffman: Esempio

- Albero di Huffman:

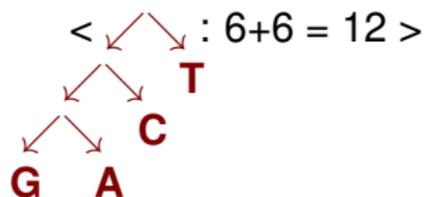


- Codice definito dal percorso attraverso l'albero per raggiungere il simbolo (0 = sinistra, 1 = destra):

A	001
T	1
C	01
G	000

# Codifica di Huffman: Esempio

- Albero di Huffman:

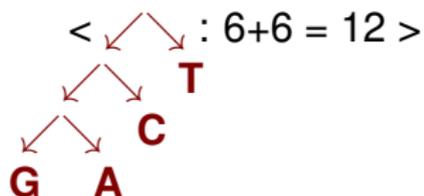


- Codice definito dal percorso attraverso l'albero per raggiungere il simbolo (0 = sinistra, 1 = destra):

A	001
T	1
C	01
G	000

# Codifica di Huffman: Esempio

- Albero di Huffman:

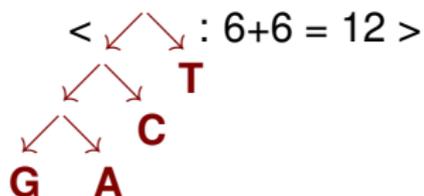


- Codice definito dal percorso attraverso l'albero per raggiungere il simbolo (0 = sinistra, 1 = destra):

A	001
T	1
C	01
G	000

# Codifica di Huffman: Esempio

- Albero di Huffman:

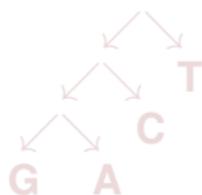


- Codice definito dal percorso attraverso l'albero per raggiungere il simbolo (0 = sinistra, 1 = destra):

<b>A</b>	001
<b>T</b>	1
<b>C</b>	01
<b>G</b>	000

# Codifica di Huffman: Esempio

- Albero di Huffman:



- Rappresentazione “compressa” della sequenza (21 bit):

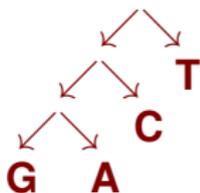
A T T C T A C C T T G T

0 0 1 1 1 0 1 1 0 0 1 0 1 0 1 1 1 0 0 0 1

- Decodifica: interpretazione della sequenza binaria come istruzioni per scendere attraverso l'albero...

# Codifica di Huffman: Esempio

- Albero di Huffman:



- Rappresentazione “compressa” della sequenza (21 bit):

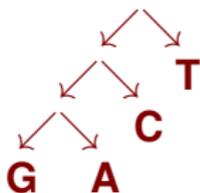
A T T C T A C C T T G T

0 0 1 1 1 0 1 1 0 0 1 0 1 0 1 1 1 0 0 0 1

- Decodifica: interpretazione della sequenza binaria come istruzioni per scendere attraverso l'albero...

# Codifica di Huffman: Esempio

- Albero di Huffman:



- Rappresentazione “compressa” della sequenza (21 bit):

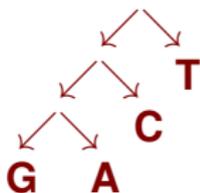
**A T T C T A C C T T G T**

**0 0 1 1 1 0 1 1 0 0 1 0 1 0 1 1 1 0 0 0 1**

- Decodifica: interpretazione della sequenza binaria come istruzioni per scendere attraverso l'albero...

# Codifica di Huffman: Esempio

- Albero di Huffman:



- Rappresentazione “compressa” della sequenza (21 bit):

**A T T C T A C C T T G T**

0 0 1 1 1 0 1 1 0 0 1 0 1 0 1 1 1 0 0 0 1

- Decodifica: interpretazione della sequenza binaria come istruzioni per scendere attraverso l'albero...

# Codifica di Huffman: Ancora un esempio?



# Codifica di Huffman: Ancora un esempio

- Testo con 4 vocali e 4 consonanti:

**ILTEATROALLIETALASERATA**

- Codifica binaria standard (stile ASCII)  
— tre bit per distinguere fra otto lettere:

A	000
E	001
I	010
O	011

L	100
R	101
S	110
T	111

- Rappresentazione della sequenza (69 bit):

```
010100111001000111101011000100100010  
001111000100000110001101000111000
```

# Codifica di Huffman: Ancora un esempio

- Testo con 4 vocali e 4 consonanti:

**ILTEATROALLIETALASERATA**

- Codifica binaria standard (stile ASCII)  
— tre bit per distinguere fra otto lettere:

A	000
E	001
I	010
O	011

L	100
R	101
S	110
T	111

- Rappresentazione della sequenza (69 bit):

010100111001000111101011000100100010  
001111000100000110001101000111000

# Codifica di Huffman: Ancora un esempio

- Testo con 4 vocali e 4 consonanti:

**I L T E A T R O A L L I E T A L A S E R A T A**

- Codifica binaria standard (stile ASCII)  
— tre bit per distinguere fra otto lettere:

<b>A</b>	000
<b>E</b>	001
<b>I</b>	010
<b>O</b>	011

<b>L</b>	100
<b>R</b>	101
<b>S</b>	110
<b>T</b>	111

- Rappresentazione della sequenza (69 bit):

010100111001000111101011000100100010  
001111000100000110001101000111000

# Codifica di Huffman: Ancora un esempio

- Testo con 4 vocali e 4 consonanti:

**I L T E A T R O A L L I E T A L A S E R A T A**

- Codifica binaria standard (stile ASCII)  
— tre bit per distinguere fra otto lettere:

<b>A</b>	000
<b>E</b>	001
<b>I</b>	010
<b>O</b>	011

<b>L</b>	100
<b>R</b>	101
<b>S</b>	110
<b>T</b>	111

- Rappresentazione della sequenza (69 bit):

010100111001000111101011000100100010  
001111000100000110001101000111000

# Codifica di Huffman: Ancora un esempio

- Lettere e pesi = numeri di occorrenze:

< **O** : ?, **S** : ?, **I** : ?, **R** : ?, **E** : ?, **L** : ?, **T** : ?, **A** : ? >

- Costruzione dell'albero di Huffman:

<  : 2, I : 2, R : 2, E : 3, L : 4, T : 4, A : 6 >

< R : 2, E : 3,  : 4, L : 4, T : 4, A : 6 >

# Codifica di Huffman: Ancora un esempio

- Lettere e pesi = numeri di occorrenze:

< **O** : 1, **S** : 1, **I** : 2, **R** : 2, **E** : 3, **L** : 4, **T** : 4, **A** : 6 >

- Costruzione dell'albero di Huffman:



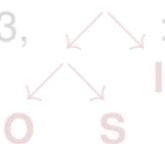
# Codifica di Huffman: Ancora un esempio

- Lettere e pesi = numeri di occorrenze:

< **O** : 1, **S** : 1, **I** : 2, **R** : 2, **E** : 3, **L** : 4, **T** : 4, **A** : 6 >

- Costruzione dell'albero di Huffman:

<  : 2, **I** : 2, **R** : 2, **E** : 3, **L** : 4, **T** : 4, **A** : 6 >

< **R** : 2, **E** : 3,  : 4, **L** : 4, **T** : 4, **A** : 6 >

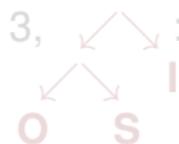
# Codifica di Huffman: Ancora un esempio

- Lettere e pesi = numeri di occorrenze:

< **O** : 1, **S** : 1, **I** : 2, **R** : 2, **E** : 3, **L** : 4, **T** : 4, **A** : 6 >

- Costruzione dell'albero di Huffman:

<  : 2, **I** : 2, **R** : 2, **E** : 3, **L** : 4, **T** : 4, **A** : 6 >

< **R** : 2, **E** : 3,  : 4, **L** : 4, **T** : 4, **A** : 6 >

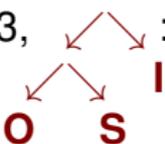
# Codifica di Huffman: Ancora un esempio

- Lettere e pesi = numeri di occorrenze:

< **O** : 1, **S** : 1, **I** : 2, **R** : 2, **E** : 3, **L** : 4, **T** : 4, **A** : 6 >

- Costruzione dell'albero di Huffman:

<  : 2, **I** : 2, **R** : 2, **E** : 3, **L** : 4, **T** : 4, **A** : 6 >

< **R** : 2, **E** : 3,  : 4, **L** : 4, **T** : 4, **A** : 6 >

# Codifica di Huffman: Ancora un esempio

- Costruzione dell'albero di Huffman:

< **R**:2, **E**:3, **I**:4, **L**:4, **T**:4, **A**:6 >



< **I**:4, **L**:4, **T**:4, **R**:5, **E**:5, **A**:6 >

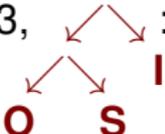


< **T**:4, **R**:5, **E**:5, **A**:6, **L**:8 >

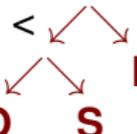


# Codifica di Huffman: Ancora un esempio

- Costruzione dell'albero di Huffman:

< **R**:2, **E**:3, :4, **L**:4, **T**:4, **A**:6 >



< :4, **L**:4, **T**:4, :5, **A**:6 >

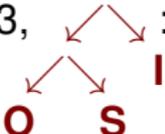


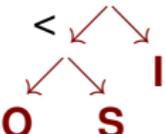
< **T**:4, :5, **A**:6, :8 >

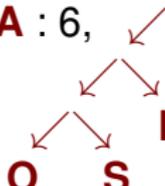


# Codifica di Huffman: Ancora un esempio

- Costruzione dell'albero di Huffman:

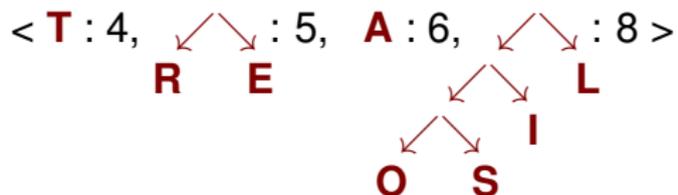
< **R** : 2, **E** : 3,  : 4, **L** : 4, **T** : 4, **A** : 6 >

<  : 4, **L** : 4, **T** : 4,  : 5, **A** : 6 >

< **T** : 4,  : 5, **A** : 6,  : 8 >

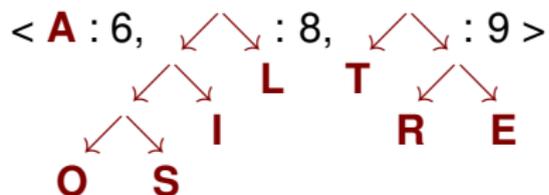
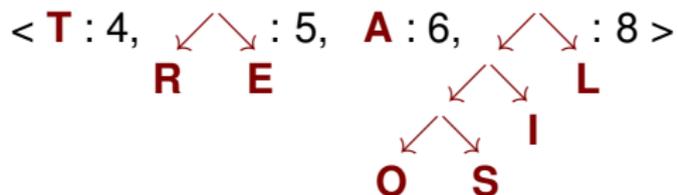
# Codifica di Huffman: Ancora un esempio

- Costruzione dell'albero di Huffman:



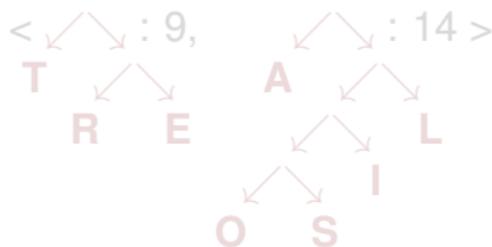
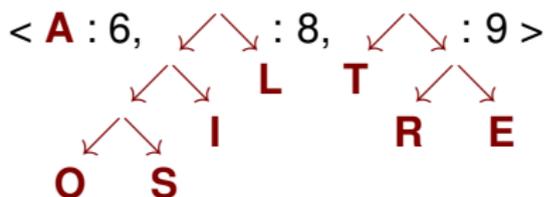
# Codifica di Huffman: Ancora un esempio

- Costruzione dell'albero di Huffman:



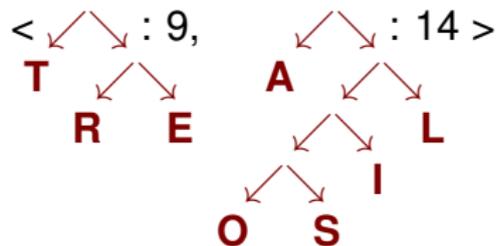
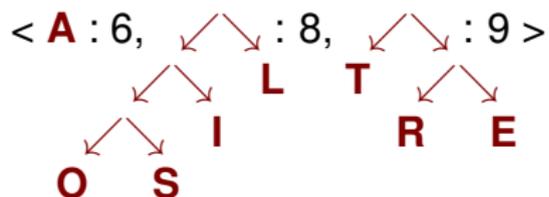
# Codifica di Huffman: Ancora un esempio

- Costruzione dell'albero di Huffman:



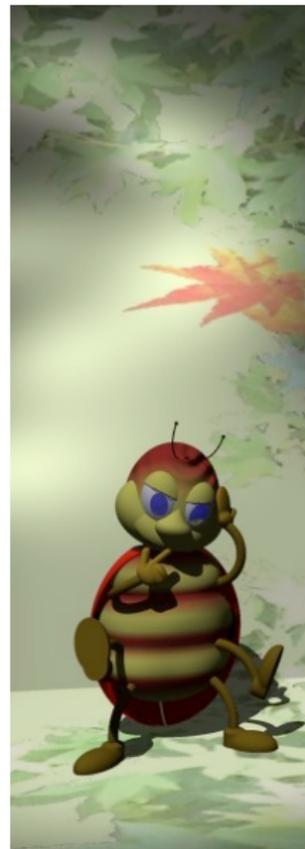
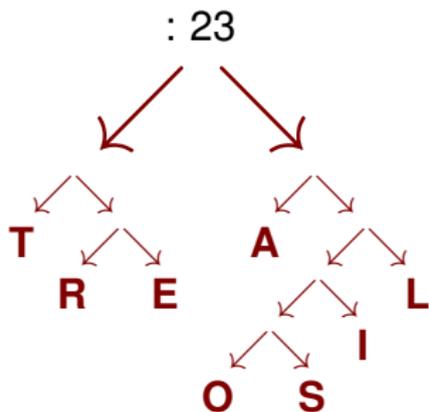
# Codifica di Huffman: Ancora un esempio

- Costruzione dell'albero di Huffman:



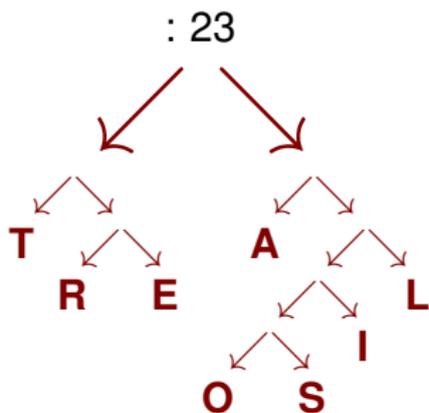
# Codifica di Huffman: Ancora un esempio

- Albero di Huffman:



# Codifica di Huffman: Ancora un esempio

- Albero di Huffman:

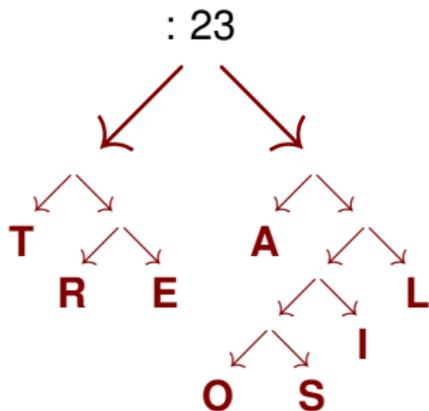


<b>A</b>	
<b>E</b>	
<b>I</b>	
<b>O</b>	

<b>L</b>	
<b>R</b>	
<b>S</b>	
<b>T</b>	

# Codifica di Huffman: Ancora un esempio

- Albero di Huffman:

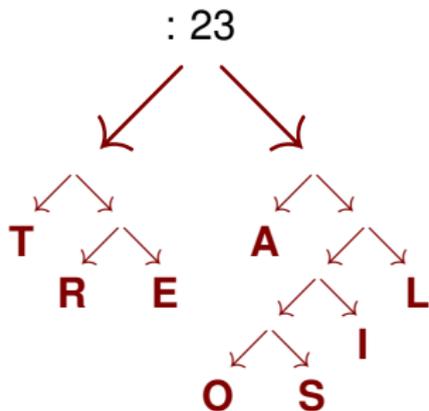


<b>A</b>	10
<b>E</b>	
<b>I</b>	
<b>O</b>	

<b>L</b>	
<b>R</b>	
<b>S</b>	
<b>T</b>	

# Codifica di Huffman: Ancora un esempio

- Albero di Huffman:

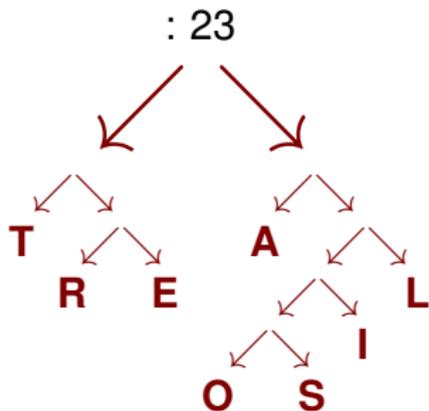


<b>A</b>	10
<b>E</b>	011
<b>I</b>	
<b>O</b>	

<b>L</b>	
<b>R</b>	
<b>S</b>	
<b>T</b>	

# Codifica di Huffman: Ancora un esempio

- Albero di Huffman:

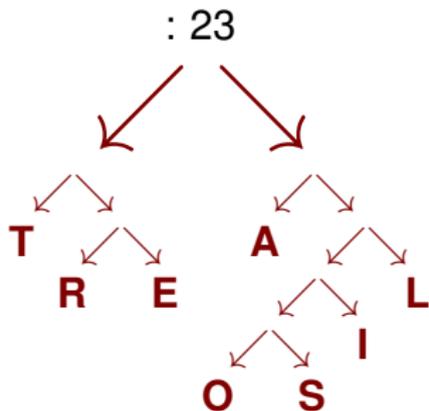


<b>A</b>	10
<b>E</b>	011
<b>I</b>	1101
<b>O</b>	

<b>L</b>	
<b>R</b>	
<b>S</b>	
<b>T</b>	

# Codifica di Huffman: Ancora un esempio

- Albero di Huffman:

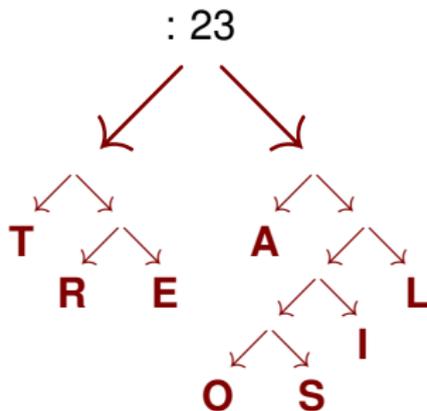


<b>A</b>	10
<b>E</b>	011
<b>I</b>	1101
<b>O</b>	11000

<b>L</b>	
<b>R</b>	
<b>S</b>	
<b>T</b>	

# Codifica di Huffman: Ancora un esempio

- Albero di Huffman:

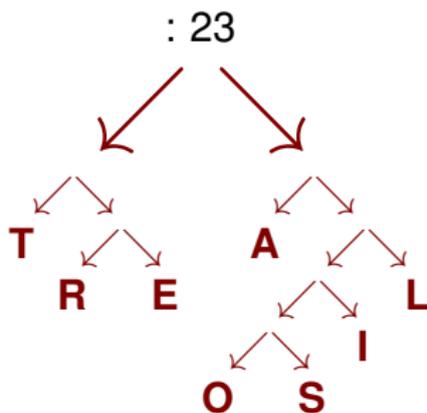


<b>A</b>	10
<b>E</b>	011
<b>I</b>	1101
<b>O</b>	11000

<b>L</b>	111
<b>R</b>	
<b>S</b>	
<b>T</b>	

# Codifica di Huffman: Ancora un esempio

- Albero di Huffman:

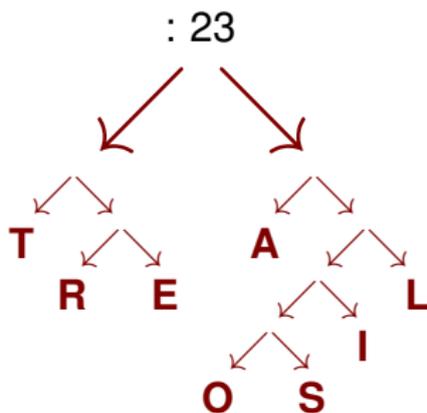


<b>A</b>	10
<b>E</b>	011
<b>I</b>	1101
<b>O</b>	11000

<b>L</b>	111
<b>R</b>	010
<b>S</b>	
<b>T</b>	

# Codifica di Huffman: Ancora un esempio

- Albero di Huffman:

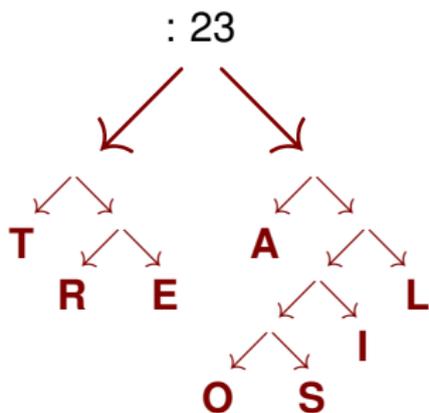


<b>A</b>	10
<b>E</b>	011
<b>I</b>	1101
<b>O</b>	11000

<b>L</b>	111
<b>R</b>	010
<b>S</b>	11001
<b>T</b>	

# Codifica di Huffman: Ancora un esempio

- Albero di Huffman:



<b>A</b>	10
<b>E</b>	011
<b>I</b>	1101
<b>O</b>	11000

<b>L</b>	111
<b>R</b>	010
<b>S</b>	11001
<b>T</b>	00

# Codifica di Huffman: Ancora un esempio

- Codifica del testo:

**ILTEATROALLIETALASERATA**

- Codifica di Huffman:

A	10
E	011
I	1101
O	11000

L	111
R	010
S	11001
T	00

- Rappresentazione della sequenza (65 bit):

110111100011100001011000101111111101  
01100101111011001011010100010

# Codifica di Huffman: Ancora un esempio

- Codifica del testo:

**ILTEATROALLIETALASERATA**

- Codifica di Huffman:

A	10
E	011
I	1101
O	11000

L	111
R	010
S	11001
T	00

- Rappresentazione della sequenza (65 bit):

110111100011100001011000101111111101  
01100101111011001011010100010

# Codifica di Huffman: Ancora un esempio

- Codifica del testo:

**ILTEATROALLIETALASERATA**

- Codifica di Huffman:

<b>A</b>	10
<b>E</b>	011
<b>I</b>	1101
<b>O</b>	11000

<b>L</b>	111
<b>R</b>	010
<b>S</b>	11001
<b>T</b>	00

- Rappresentazione della sequenza (65 bit):

110111100011100001011000101111111101  
01100101111011001011010100010

# Codifica di Huffman: Ancora un esempio

- Codifica del testo:

**ILTEATROALLIETALASERATA**

- Codifica di Huffman:

<b>A</b>	10
<b>E</b>	011
<b>I</b>	1101
<b>O</b>	11000

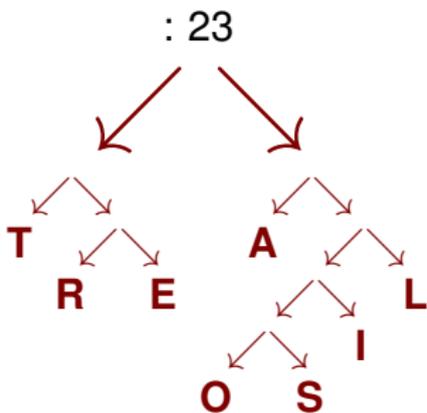
<b>L</b>	111
<b>R</b>	010
<b>S</b>	11001
<b>T</b>	00

- Rappresentazione della sequenza (65 bit):

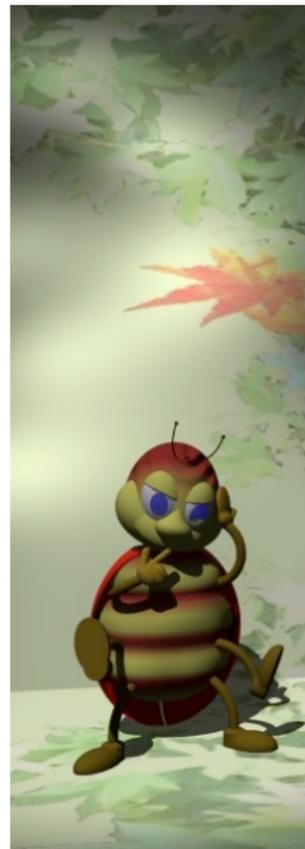
110111100011100001011000101111111101  
01100101111011001011010100010

# Codifica di Huffman: Ancora un esempio

- Decodifica?



110111100011100001011000101111111101  
01100101111011001011010100010



**Domande?**

