

# Digitalizzazione

*«Notizia, dato o elemento che consente di avere conoscenza più o meno esatta di fatti, situazioni, modi di essere. In senso più generale, anche la trasmissione dei dati e l'insieme delle strutture che la consentono»*

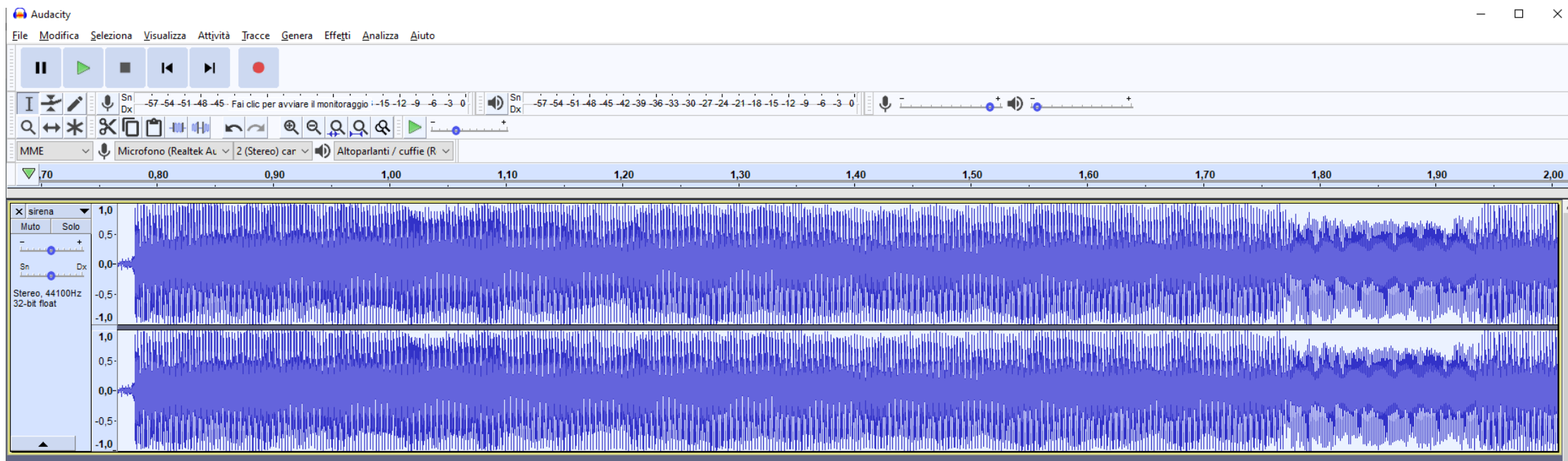
(da <https://www.treccani.it/enciclopedia/informazione>)

**Trasferimento di contenuti** (che si trasformeranno dentro di noi in concetti)

Per trasferire e memorizzare questi contenuti in un supporto serve una grandezza fisica che abbia almeno due stati. Es:

- Pressione
- Tensione
- Colore
- ...

# Esempio



**Digitalizzare un'informazione:** rappresentarla come una sequenza di numeri (dall'inglese *digit* – cifra – che deriva dal latino *digitus* – dito)

«**Digitalizzazione:** Nella tecnica, conversione di grandezze analogiche in informazioni digitali, effettuata mediante un dispositivo, detto digitalizzatore o convertitore analogico-digitale.»

(Da <https://www.treccani.it/enciclopedia/ricerca/digitalizzare/>)

# Digitalizzazione

Un elaboratore digitale rappresenta i numeri con solo due cifre, 0 e 1  
(codifica binaria)

Possiamo digitalizzare

- Testi
- Suoni
- Immagini
- Filmati
- E ovviamente qualsiasi tipo di numero

*«Scienza che studia l’elaborazione delle informazioni e le sue applicazioni; più precisamente l’informatica si occupa della rappresentazione, dell’organizzazione e del trattamento automatico della informazione.»*

(Da <https://www.treccani.it/enciclopedia/ricerca/digitalizzare/>)

Codificare informazione multimediale in forma digitale

Concepire algoritmi per elaborarla al fine di risolvere problemi dati



# Algoritmo

Insieme di regole volte a risolvere un determinato problema in un numero finito di passi (operazioni, istruzioni)



# Codifica ASCII

## American Standard Code for Information Interchange

- Sviluppato da una commissione dell'American Standard Association
- Prima versione 1963, revisionato nel 1967
- Caratteri su **7 bit**
- Cifre e simboli di punteggiatura prima delle lettere
- Lettere maiuscole prima delle minuscole

# Codifica ASCII

La tabella si legge prendendo prima i 3 bit dell'intestazione di colonna, poi i 4 bit dell'intestazione di riga.

Ad esempio il codice ASCII della lettera 'A' è 100 0001 (equivalente a 65 in base 10)

Il carattere composto solamente da 0 (000 0000) corrisponde al concetto di «niente», «nessun carattere».

000	001	010	011	100	101	110	111	
NUL \0	DLE	SP	0	@	P	`	p	0000
SOH	XON	!	1	A	Q	a	q	0001
STX	DC2	"	2	B	R	b	r	0010
ETX	XOFF	#	3	C	S	c	s	0011
EQT	DC4	\$	4	D	T	d	t	0100
ENQ	NAK	%	5	E	U	e	u	0101
ACK	SYN	&	6	F	V	f	v	0110
BEL \a	ETB	'	7	G	W	g	w	0111
BS \b	CAN	(	8	H	X	h	x	1000
HT \t	EM	)	9	I	Y	i	y	1001
LF \n	SUB	*	:	J	Z	j	z	1010
VF \v	ESC	+	;	K	(	k	{	1011
FF \f	FS	,	<	L	\	l		1100
CR \r	GS	-	=	M	)	m	}	1101
SO	RS	.	>	N	^	n	~	1110
SI	US	/	?	O	_	o	DEL	1111

# Codifica ASCII

L'ASCII nasce nel mondo anglosassone. E' sufficiente per lo scambio di caratteri in lingua inglese.

E i caratteri accentati?

E simboli monetari diversi dal dollaro?

# Codifiche ISO 8859

**ISO 8859** – Serie di standard promossi dall'International Standard Organization (ISO) e dall'International Electrotechnical Commission (IEC) per la codifica di caratteri su **8 bit**.

- Nasce per rimediare al problema della necessità di caratteri aggiuntionali
- Aggiunge un ulteriore bit a sinistra, passando dai 7 bit dell'ASCII a 8 bit e raddoppiando i posti disponibili, da 128 ( $2^7$ ) a 256 ( $2^8$ )

# Codifiche ISO 8859

- ISO 8859-1 west European languages (Latin-1)
- ISO 8859-2 east European languages (Latin-2)
- ISO 8859-3 southeast European languages (Latin-3)
- ISO 8859-4 Scandinavian/Baltic languages (Latin-4)
- ISO 8859-5 Latin/Cyrillic
- ISO 8859-6 Latin/Arabic
- ISO 8859-7 Latin/Greek
- ISO 8859-8 Latin/Hebrew
- ISO 8859-9 Latin-1 modification for Turkish (Latin-5)
- ISO 8859-10 Lappish/Nordic/Eskimo languages (Latin-6)
- ISO 8859-11 Latin/Thai
- ~~ISO 8859-12 Latin/Devanagari~~
- ISO 8859-13 Baltic Rim languages (Latin-7)
- ISO 8859-14 Celtic (Latin-8)
- ISO 8859-15 west European languages (Latin-9)

# ISO 8859-1



unIMC

0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111	
NUL \0	DLE	SP	0	@	P	`	p			NBS	°	À	Ð	à	ð	0000
SOH	XON	!	1	A	Q	a	q			¡	±	Á	Ñ	á	ñ	0001
STX	DC2	"	2	B	R	b	r			¢	²	Â	Ò	â	ò	0010
ETX	XOFF	#	3	C	S	c	s			£	³	Ã	Ó	ã	ó	0011
EQT	DC4	\$	4	D	T	d	t			¤	´	Ä	Ô	ä	ô	0100
ENQ	NAK	%	5	E	U	e	u			¥	µ	Å	Õ	å	õ	0101
ACK	SYN	&	6	F	V	f	v			¦	¶	Æ	Ö	æ	ö	0110
BEL \a	ETB	'	7	G	W	g	w			§	*	Ç	×	ç	÷	0111
BS \b	CAN	(	8	H	X	h	x			"	,	È	Ø	è	ø	1000
HT \t	EM	)	9	I	Y	i	y			©	ı	É	Ù	é	ù	1001
LF \n	SUB	*	:	J	Z	j	z			ª	º	Ê	Ú	ê	ú	1010
VF \v	ESC	+	;	K	(	k	{			«	»	Ë	Û	ë	û	1011
FF \f	FS	,	<	L	\	l				¬	¼	Ì	Ü	ì	ü	1100
CR \r	GS	-	=	M	)	m	}			SH	½	Í	Ý	í	ý	1101
SO	RS	.	>	N	^	n	~			®	¾	Î	Þ	î	þ	1110
SI	US	/	?	O	_	o	DEL			¯	¿	Ï	ß	ï	ÿ	1111

# ISO 8859-15

L'ISO 8859-15 è un aggiornamento dell'ISO 8859-1 per aggiungere il simbolo dell'Euro e completare la copertura di francese, finnico e estone.

0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111	
NUL \0	DLE	SP	0	@	P	`	p			NBS	°	À	Đ	à	đ	0000
SOH	XON	!	1	A	Q	a	q			ı	±	Á	Ñ	á	ñ	0001
STX	DC2	"	2	B	R	b	r			ç	²	Â	Ò	â	ò	0010
ETX	XOFF	#	3	C	S	c	s			£	³	Ã	Ó	ã	ó	0011
EQT	DC4	\$	4	D	T	d	t			€	Ž	Ä	Ô	ä	ö	0100
ENQ	NAK	%	5	E	U	e	u			¥	μ	Å	Õ	å	õ	0101
ACK	SYN	&	6	F	V	f	v			Š	¶	Æ	Ö	æ	ö	0110
BEL \a	ETB	'	7	G	W	g	w			§	*	Ç	×	ç	÷	0111
BS \b	CAN	(	8	H	X	h	x			š	ž	È	Ø	è	ø	1000
HT \t	EM	)	9	I	Y	i	y			©	ı	É	Ù	é	ù	1001
LF \n	SUB	*	:	J	Z	j	z			ª	º	Ê	Ú	ê	ú	1010
VF \v	ESC	+	;	K	(	k	{			«	»	Ë	Û	ë	û	1011
FF \f	FS	,	<	L	\	l				¬	œ	Ì	Ü	ì	ü	1100
CR \r	GS	-	=	M	)	m	}			SH	œ	Í	Ý	í	ý	1101
SO	RS	.	>	N	^	n	~			®	ÿ	Î	Þ	î	þ	1110
SI	US	/	?	O	_	o	DEL			-	ı	İ	ß	ï	ÿ	1111



# Codifiche ISO 8859

E se un documento dovesse contenere sia caratteri in greco che parole in italiano?

E cinese, giapponese, coreano...?

# Standard Unicode (ISO 10646)

**Obiettivo:** sistema di codifica che assegni un numero univoco a ciascun carattere in modo da gestire testi scritti nella maggior parte delle sistemi di scrittura del mondo.

- Caratteri su 31 bit (quindi ci sono 2 147 483 648 slot disponibili)
- I primi 16 bit (65 536 caratteri) costituiscono il Basic Multilingual Plane che contiene la maggior parte dei caratteri usati nelle lingue moderne