

CPSC 100

Computational Thinking

Data Representation

Instructor: Firas Moosvi
Department of Computer Science
University of British Columbia



Course Admin



Learning Goals

After this **today's lecture**, you should be able to:

- Explain what ASCII and Unicode are, including their historical context, purpose, and significance in computing.
- Decode an ASCII representation of a short text document
 - (with a list of ASCII codes provided)
- Articulate why Unicode was created and how it solved the problems of earlier encoding systems like ASCII
- Compare and contrast Unicode with ASCII in terms of character range, encoding length, and use cases



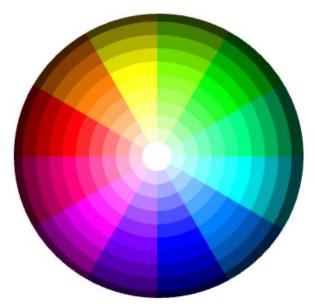
Learning Goals



Q: Which colour best describes the one represented by the hexadecimal colour code: #32CD32?



- A. Shade of red
- B. Shade of blue
- C. Shade of green
- D. Shade of purple
- E. Shade of yellow

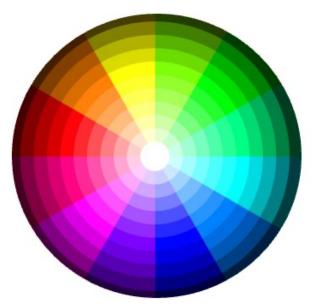




Q: Which colour best describes the one represented by the hexadecimal colour code: #800B80?



- A. Shade of red
- B. Shade of blue
- C. Shade of green
- D. Shade of purple
- E. Shade of yellow





Activity



Color Mixing: Match the Colour to its Hex Rep.

Hex. Rep.

#FFA933

#FF99FF

#EAE51D

#A1A2A3

#1234F8

shade of ...

Yellow

Pink or

Magenta

Blue

Orange

Grey







Optional Video about Colours



Colour Theory

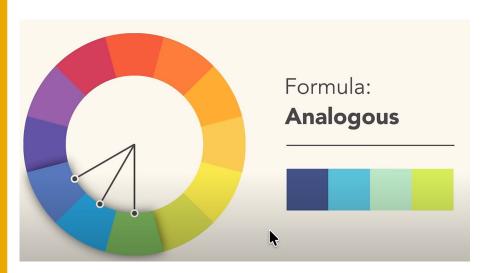
https://youtu.be/ 2LLXnUdUlc?si=ZC0gCVCkhlmnc3KT





Colour Theory

https://youtu.be/ 2LLXnUdUlc?si=ZC0gCVCkhlmnc3KT







Data Representation in characters!



How do computer store etters and characters?



ASCII: Overview

American Standard Code for Information Interchange (ASCII)

- Character encoding standard
- Allows computers and electronic devices to represent text
- Developed in 1960s
 - Standardize how computers represent characters



ASCII Table

Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char
0	0	[NULL]	32	20	[SPACE]	64	40	@	96	60	`
1	1	[START OF HEADING]	33	21	1	65	41	A	97	61	a
2	2	[START OF TEXT]	34	22		66	42	В	98	62	b
3	3	[END OF TEXT]	35	23	#	67	43	С	99	63	c
4	4	[END OF TRANSMISSION]	36	24	\$	68	44	D	100	64	d
5	5	[ENQUIRY]	37	25	%	69	45	E	101	65	e
6	6	[ACKNOWLEDGE]	38	26	&	70	46	F	102	66	f
7	7	(BELL)	39	27		71	47	G	103	67	g
8	8	[BACKSPACE]	40	28	(72	48	н	104	68	h
9	9	[HORIZONTAL TAB]	41	29)	73	49	1	105	69	i
10	Α	[LINE FEED]	42	2A	*	74	4A	J	106	6A	j
11	В	[VERTICAL TAB]	43	2B	+	75	4B	K	107	6B	k
12	С	[FORM FEED]	44	2C	,	76	4C	L	108	6C	1
13	D	[CARRIAGE RETURN]	45	2D	-	77	4D	M	109	6D	m
14	E	[SHIFT OUT]	46	2E		78	4E	N	110	6E	n
15	F	[SHIFT IN]	47	2F	/	79	4F	0	111	6F	0
16	10	[DATA LINK ESCAPE]	48	30	0	80	50	P	112	70	р
17	11	[DEVICE CONTROL 1]	49	31	1	81	51	Q	113	71	q
18	12	[DEVICE CONTROL 2]	50	32	2	82	52	R	114	72	r
19	13	[DEVICE CONTROL 3]	51	33	3	83	53	S	115	73	s
20	14	[DEVICE CONTROL 4]	52	34	4	84	54	T	116	74	t
21	15	[NEGATIVE ACKNOWLEDGE]	53	35	5	85	55	U	117	75	u
22	16	[SYNCHRONOUS IDLE]	54	36	6	86	56	V	118	76	v
23	17	[ENG OF TRANS. BLOCK]	55	37	7	87	57	W	119	77	w
24	18	[CANCEL]	56	38	8	88	58	X	120	78	x
25	19	[END OF MEDIUM]	57	39	9	89	59	Υ	121	79	У
26	1A	[SUBSTITUTE]	58	3A	:	90	5A	Z	122	7A	z
27	1B	[ESCAPE]	59	3B	;	91	5B	[123	7B	{
28	1C	[FILE SEPARATOR]	60	3C	<	92	5C	\	124	7C	Ĭ
29	1D	[GROUP SEPARATOR]	61	3D	=	93	5D	1	125	7D	}
30	1E	[RECORD SEPARATOR]	62	3E	>	94	5E	^	126	7E	~
31	1F	[UNIT SEPARATOR]	63	3F	?	95	5F	_	127	7F	[DEL]



ASCII: Overview

- Punctuation, spaces and other special control characters are also encoded
 - each encoded item is sometimes called a code point
- ASCII uses 7 bits to represent each character, which allows for 128 (2⁷) unique characters (from 0 to 127)
 - Why 7 bits?
 - An extra "check" bit to detect certain errors that might arise
- Extended ASCII uses 8 bits (or one byte), allowing for characters with accents (Á, ë and others)



ASCII: Overview

ASCII uses **7 bits** to represent each character:

- Control characters (0–31): For managing hardware (like line breaks or bell sounds).
- **Printable characters** (32–126): Letters, digits, punctuation, and symbols.
- Delete character (127).



ASCII: Control characters (0-31)

Decimal	Hex	Char	Decimal	Нех	Char	Decimal	Нех	Char	Decimal	Нех	Char
0	0	[NULL]	32	20		64	40		96		
1	1	[START OF HEADING]	33	21	1	65	41	Α	97	61	a
2	2	[START OF TEXT]	34	22		66	42	В	98	62	b
3	3	[END OF TEXT]	35	23		67	43	C	99	63	С
4	4	[END OF TRANSMISSION]		24		68	44		100	64	
5	5	[ENQUIRY]	37	25		69	45	E	101	65	е
6	6	[ACKNOWLEDGE]		26	& c	70	46	F	102	66	f
7	7	[BELL]	39	27		71	47	G	103	67	g
8	8	[BACKSPACE]	40	28		72	48	H	104		h
9	9	[HORIZONTAL TAB]	41	29		73			105		i i
10	Α	[LINE FEED]	42	2A	*	74	4A	J	106	6A	j
11	В	[VERTICAL TAB]	43	2B	+	75	4B	K	107	6B	k
12	С	[FORM FEED]	44	2C	,	76	4C	L	108	6C	
13	D	[CARRIAGE RETURN]	45	2D	-	77	4D	M	109		m
14	E	[SHIFT OUT]	46	2E		78		N	110		
15	F	[SHIFT IN]	47	2F		79	4F	0	111		0
16	10	[DATA LINK ESCAPE]	48			80		P	112	70	р
17	11	[DEVICE CONTROL 1]	49	31	1	81	51	Q	113	71	q
18	12	[DEVICE CONTROL 2]	50	32	2	82	52	R	114	72	r
19	13	[DEVICE CONTROL 3]	51			83			115		
20	14	[DEVICE CONTROL 4]	52	34	4	84	54	T	116	74	t
21	15	[NEGATIVE ACKNOWLEDGE]	53		5	85		U	117		u
22	16	[SYNCHRONOUS IDLE]	54		6	86		V	118	76	V
23	17	[ENG OF TRANS. BLOCK]	55	37	7	87	57	W	119	77	W
24	18	[CANCEL]				88		X	120		X
25	19	[END OF MEDIUM]	57	39	9	89	59	Υ	121	79	У
26	1A	[SUBSTITUTE]	58	3A	:	90	5A	Z	122	7A	Z
27	1B	[ESCAPE]	59		;	91		[123		{
28	1C	[FILE SEPARATOR]	60		<	92			124		
29	1D	[GROUP SEPARATOR]	61			93]	125		}
30	1E	[RECORD SEPARATOR]	62		>	94		^	126		
31	1F	[UNIT SEPARATOR]	63		?	95			127	7F	



ASCII: Printable characters (32-126)

Decimal	Нех	Char	Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char
			32	20	[SPACE]	64	40	@	96	60	`
1	1		33	21	1	65	41	A	97	61	a
2	2		34	22		66	42	В	98	62	b
3			35	23	#	67	43	С	99	63	c
4	4		36	24	\$	68	44	D	100	64	d
5			37	25	%	69	45	E	101	65	e
6	6		38	26	&	70	46	F	102	66	f
7	7		39	27	1	71	47	G	103	67	g
			40	28	(72	48	Н	104	68	h
9			41	29)	73	49	1	105	69	i i
10	Α		42	2A	*	74	4A	J	106	6A	j
11			43	2B	+	75	4B	K	107	6B	k
12			44	2C	,	76	4C	L	108	6C	1
13	D		45	2D	-	77	4D	M	109	6D	m
14			46	2E	4	78	4E	N	110	6E	n
15	F		47	2F	1	79	4F	0	111	6F	0
16	10		48	30	0	80	50	Р	112	70	р
17	11		49	31	1	81	51	Q	113	71	q
18	12		50	32	2	82	52	R	114	72	r
19	13		51	33	3	83	53	S	115	73	S
20	14		52	34	4	84	54	Т	116	74	t
21	15		53	35	5	85	55	U	117	75	u
22	16		54	36	6	86	56	V	118	76	V
23	17		55	37	7	87	57	W	119	77	w
24	18		56	38	8	88	58	X	120	78	X
25	19		57	39	9	89	59	Y	121	79	У
26	1A		58	3A	:	90	5A	Z	122	7A	Z
27	1B		59	3B	;	91	5B	[123	7B	{
28	1C		60	3C	<	92	5C	\	124	7C	
29	1D		61	3D	=	93	5D	1	125	7D	}
30	1E		62	3E	>	94	5E	^	126	7E	~
31	1F		63	3F	?	95	5F	_	127	7F	[DEL]



ASCII: Delete character (127)

Decimal	Нех	Char									
			32	20		64	40		96		
1	1		33	21		65	41	Α	97	61	a
2	2		34	22		66	42	В	98	62	b
3			35	23		67	43	C	99	63	С
4	4			24		68	44	D	100	64	
5			37	25		69	45	E	101	65	е
6	6			26	δι	70	46	F	102	66	f
7	7		39	27		71	47	G	103	67	g
			40	28		72	48	H	104	68	h
9			41	29		73		1	105		i
10	Α		42	2A	*	74	4A	J	106	6A	j
11			43	2B	+	75	4B	K	107	6B	k
12			44	2C	,	76	4C	L	108	6C	1
13	D		45	2D		77	4D	M	109		m
14			46	2E		78		N	110		
15	F		47	2F		79	4F	0	111		0
16	10		48			80		P	112	70	р
17	11		49	31	1	81	51	Q	113	71	q
18	12		50	32	2	82	52	R	114	72	r
19			51			83			115	73	
20	14		52	34	4	84	54	T	116	74	t
21	15		53		5	85		U	117	75	u
22	16		54		6	86		V	118	76	V
23	17		55	37	7	87	57	W	119	77	W
24						88		X	120		X
25	19		57	39	9	89	59	Υ	121	79	У
26	1A		58	3A	:	90	5A	Z	122	7A	Z
27	1B		59		;	91		[123	7B	{
28	1C		60		<	92			124	7C	
29			61			93]	125		
30	1E		62		>	94		^	126	7E	
31	1F		63		?	95			127	7F	[DEL]







Q: Convert Hi! from ASCII to Decimal



- A. 48, 69, 21
- B. 47, 68, 20
- C. 71, 104, 34
- D. 72, 105, 33
- E. None of the above



Q: Convert the following Hex to ASCII:



- A. BADGE
- B. *),/-
- C. A@CFD
- D.)(+.,
- E. None of the above

Binary	01000010	01000001	01000100	01000111	01000101
Hex	42	41	44	47	45
ASCII					

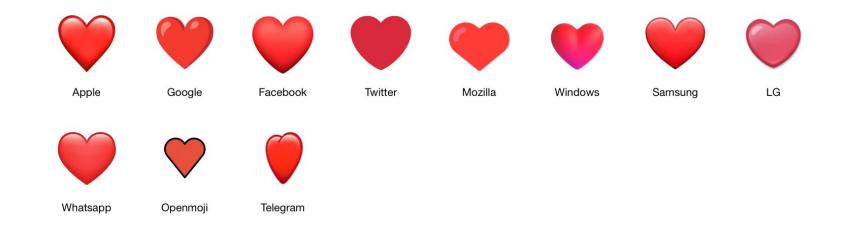


Can we do this faster?



Convert Text to ASCII

Emoji ♥ Red Heart on platforms



Technical Information			roperties	Encoding	Copy & Paste Codes	
Encoding	hex	dec (byte	s) dec	2	binary	
UTF-8	E2 9D A4	226 157 16	14851492	111	100010 10011101 10100100	

Source



Demo



Application of ASCII The Martian



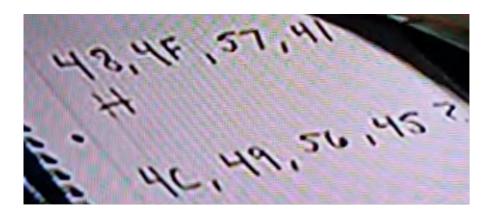
The Martian

- "Ridley Scott claimed that one of the most difficult scenes to direct was how to explain to the audience the hexadecimal system Watney uses as a code to communicate with Earth, which Scott admitted was hard for himself to understand." IMDb
- Youtube Video:
 - o The Martian, Hexadecimal Scene





The Martian



Decimal	Hex	Char
64	40	@
65	41	A
66	42	В
67	43	C
68	44	D
69	45	E
70	46	F
71	47	G
72	48	H
73	49	1
74	4A	J
75	4B	K
76	4C	L
77	4D	M
78	4E	N
79	4F	0
80	50	P
81	51	Q
82	52	R
83	53	S
84	54	т
85	55	U
86	56	V
87	57	W
88	58	X
89	59	Y Z
90	5A	Z
0.1	ED.	F .



What about other languages?

Arabic (火字) Chinese (汉字) Emojis (空, ※)



Unicode



Unicode



Unicode: Overview

Universal character encoding standard

 Designed to represent every character from every language in the world, as well as symbols, emojis, and special scripts, using a unified system.

Before unicode...

- Different encoding systems (like ASCII, ISO-8859-1, Shift-JIS)
 - Difficult to mix languages in one document
- Unicode solved this by creating one global codebook



How does Unicode Work?

- Unicode assigns a unique number called a "code point" to every character, regardless of platform, program, or language.
 - Code points are written like: U+0041 (which is 'A').

- Unicode itself is just a standard. To store the characters in files and transmit them over networks, you need encoding formats
 - There are different implementations, including UTF-8 and UTF-16 (UTF stands for Unicode Transformation Format)



Encoding Formats

- UTF-8 and UTF-16 are variable length encodings
- They use 1 byte (8 bits) for ASCII, but more for other characters

character	encoding				bits
A	UTF-8				01000001
A	UTF-16			00000000	01000001
A	UTF-32	00000000	00000000	00000000	01000001
あ	UTF-8		11100011	10000001	10000010
あ	UTF-16			00110000	01000010
あ	UTF-32	00000000	00000000	00110000	01000010



Emoji

 Notice how Emoji have the same "CLDR" names and Unicode values but different aesthetics

Smi	Smileys & Emotion											
face	-smiling											
Nº	Code	Browser	Sample	GMail	SB	DCM	KDDI	CLDR Short Name				
1	U+1F600			•	_	_	_	grinning face				
2	U+1F603			0	a	ä	(3)	grinning face with big eyes				
3	U+1F604		\(\text{\cond}\)		€	_	_	grinning face with smiling eyes				
4	U+1F601			Ö	뜐	200	≅	beaming face with smiling eyes				
5	U+1F606		25	V	_	∜	_	grinning squinting face				





Emoji

This is how skin-tones are varied!

hand	-fingers-o	pen		
Nº	Code	Browser	Sample	CLDR Short Name
1	U+1F44B U+1F3FB			waving hand: light skin tone
2	U+1F44B U+1F3FC			waving hand: medium-light skin tone
3	U+1F44B U+1F3FD			waving hand: medium skin tone
4	U+1F44B U+1F3FE			waving hand: medium-dark skin tone
5	U+1F44B U+1F3FF			waving hand: dark skin tone









Demo



How does Word store its data?

 Uploading a Word document into the online Hex editor suggests that the document is not in ASCII representation

- Most of the files that comprise a Word document are in XML
 (Extensible Markup Language) format; they describe metadata such as the font style and size, document creator, etc.
- The files may also contain information about tracked changes to the document, collaborators, privacy and security settings, and more



Activity

How does Microsoft Word store its data?

- Open Microsoft Word
- Write your name in the file and save it
- Visit https://hexed.it/
- Open the file using Open file feature
- What do you notice?



CPSC 100 Course Project



Class Activity

Today we will make groups for the project!

 First, a bit more about the CPSC 100 course project and Contract Grading

CPSC 100 Term Project

Contract Grading [DRAFT]

Introduction to Contract Grading

Contract Grading

- 1. Contract Grading involves a shared process of assessment between students and instructors:
 - Based on self-assessment of strengths and limitations, and will determine what students need to do to achieve a certain grade.
- 2. Contract grading can be used for a specific assignment or project, or for an entire course.

- 3. Course grades are determined by students completing a pre-determined set of tasks or demonstrate proficiency in skills that students pick from:
 - Tasks and Skills that require more work or are more challenging (or both)
 allow students to earn higher grades.

Contract Grading - Example

Number of texts read / blogposts written	Percentage Grade	Final Letter Grade
4 (Proust, Bombal, Perec, and Bolaño) / 6	60-63	C
5 (the basic four plus one) / 7	64-67	C+
6 (the basic four plus two) / 8	68-71	B-
7 (the basic four plus three) / 9	72-75	В
8 (the basic four plus four) / 10	76-79	B+
9 (the basic four plus five) / 11	80-84	A-
10 (the basic four plus six) / 12	85-89	A
11 (the basic four plus seven) / 13	90-100	A+

Contract Grading in CPSC 100

Contract Grading in Data 301

1. In CPSC 100, we will use Contract Grading for the course project.

- 2. At the beginning of the project (in Milestone 1), you will choose to contract (with me) a grade that you would like to aim for.
 - All members of the group MUST agree on the contracted grade!

- 3. You will have two more opportunities mid-way through the project to upgrade (or downgrade) your contracted grade.
 - All members of the group MUST agree on the change!

4. All members of the group will earn the same group grade in the project

	Requirements (Draft)
C	 A bunch of requirements Demonstrate an EXCELLENT level of understanding of 3 course topics
	Everything in the buckets above PLUS
В	 A few more requirements Demonstrate an EXCELLENT level of understanding of +1 course topic
	Everything in the buckets above PLUS
A	 A few more requirements Demonstrate an EXCELLENT level of understanding of +1 course topic
	Everything in the buckets above PLUS
A +	- A few more requirements - Demonstrate an EXCELLENT level of understanding of +3 course topics



Q: Which topic are you most interested in (not finalized)?

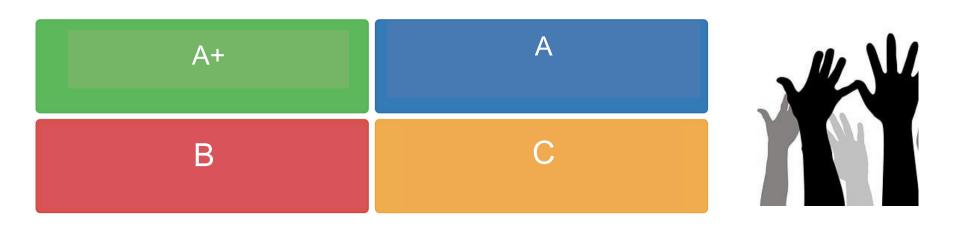


- A. Generative Al and student learning
- B. Technological inequality
- C. Algorithmic Bias
- D. Financial and Environmental impacts of technology
- E. Technology and Crime



Participation Question

What grade do you intend to contract for?





Take Home Practice



Q: Convert the following hexadecimal sequence to ASCII: 53 54 41 52 53



- A. STARS
- B. 56)45
- C. !@#\$\$
- D. HELLO
- E. WORLD

Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char
0	0	[NULL]	32	20	[SPACE]	64	40	@	96	60	*
1	1	[START OF HEADING]	33	21	1	65	41	A	97	61	a
2	2	ISTART OF TEXT!	34	22		66	42	В	98	62	b
3	3	IEND OF TEXTI	35	23	#	67	43	C	99	63	c
4	4	[END OF TRANSMISSION]	36	24	\$	68	44	D	100	64	d
5	5	[ENOUIRY]	37	25	%	69	45	E	101	65	e
6	6	[ACKNOWLEDGE]	38	26	84	70	46	F	102	66	f
7	7	[BELL]	39	27	.1	71	47	G	103	67	a
8	8	[BACKSPACE]	40	28	(72	48	н	104	68	h
9	9	[HORIZONTAL TAB]	41	29)	73	49	1	105	69	1
10	A	[LINE FEED]	42	2A	*	74	4A	J	106	6A	i
11	В	[VERTICAL TAB]	43	2B	+	75	4B	K	107	6B	k
12	C	[FORM FEED]	44	2C	,	76	4C	L	108	6C	1
13	D	[CARRIAGE RETURN]	45	2D	-	77	4D	M	109	6D	m
14	E	[SHIFT OUT]	46	2E		78	4E	N	110	6E	n
15	F	(SHIFT IN)	47	2F	1	79	4F	0	111	6F	0
16	10	[DATA LINK ESCAPE]	48	30	0	80	50	P	112	70	p
17	11	[DEVICE CONTROL 1]	49	31	1	81	51	Q	113	71	q
18	12	[DEVICE CONTROL 2]	50	32	2	82	52	R	114	72	r
19	13	[DEVICE CONTROL 3]	51	33	3	83	53	S	115	73	S
20	14	[DEVICE CONTROL 4]	52	34	4	84	54	Т	116	74	t
21	15	[NEGATIVE ACKNOWLEDGE]	53	35	5	85	55	U	117	75	u
22	16	[SYNCHRONOUS IDLE]	54	36	6	86	56	V	118	76	V
23	17	[END OF TRANS. BLOCK]	55	37	7	87	57	W	119	77	w
24	18	[CANCEL]	56	38	8	88	58	X	120	78	x
25	19	[END OF MEDIUM]	57	39	9	89	59	Y	121	79	V
26	1A	[SUBSTITUTE]	58	3A	:	90	5A	Z	122	7A	z
27	1B	[ESCAPE]	59	3B	;	91	5B	1	123	7B	-
28	1C	[FILE SEPARATOR]	60	3C	<	92	5C	1	124	7C	T.
29	1D	[GROUP SEPARATOR]	61	3D	-	93	5D	1	125	7D	}
30	1E	[RECORD SEPARATOR]	62	3E	>	94	5E	^	126	7E	~
31	1F	[UNIT SEPARATOR]	63	3F	?	95	5F		127	7F	[DEL]



Q: Convert the following hexadecimal sequence to ASCII: 53 54 41 52 53



A. STARS

B. 56)45

C. !@#\$\$

D. HELLO

E. WORLD

Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char
0	0	(NULL)	32	20	[SPACE]	64	40	@	96	60	*
1	1	[START OF HEADING]	33	21	1	65	41	A	97	61	a
2	2	[START OF TEXT]	34	22	н	66	42	В	98	62	b
3	3	[END OF TEXT]	35	23	#	67	43	C	99	63	C
4	4	[END OF TRANSMISSION]	36	24	\$	68	44	D	100	64	d
5	5	[ENQUIRY]	37	25	%	69	45	E	101	65	e
6	6	[ACKNOWLEDGE]	38	26	Se.	70	46	F	102	66	f
7	7	[BELL]	39	27		71	47	G	103	67	q
В	8	[BACKSPACE]	40	28	(72	48	н	104	68	h
9	9	[HORIZONTAL TAB]	41	29)	73	49	1	105	69	1
10	A	[LINE FEED]	42	2A	*	74	4A	J	106	6A	i
11	В	[VERTICAL TAB]	43	2B	+	75	4B	K	107	6B	k
12	C	[FORM FEED]	44	2C		76	4C	L	108	6C	1
13	D	[CARRIAGE RETURN]	45	2D	-	77	4D	M	109	6D	m
14	E	[SHIFT OUT]	46	2E		78	4E	N	110	6E	n
15	F	[SHIFT IN]	47	2F	1	79	4F	0	111	6F	0
16	10	[DATA LINK ESCAPE]	48	30	0	80	50	P	112	70	P
17	11	[DEVICE CONTROL 1]	49	31	1	81	51	Q	113	71	q
18	12	IDEVICE CONTROL 21	50	32	2	82	52	R	114	72	r
19	13	[DEVICE CONTROL 3]	51	33	3	83	53	S	115	73	S
20	14	IDEVICE CONTROL 41	52	34	4	84	54	Т	116	74	t
21	15	[NEGATIVE ACKNOWLEDGE]	53	35	5	85	55	U	117	75	u
22	16	[SYNCHRONOUS IDLE]	54	36	6	86	56	V	118	76	V
23	17	[END OF TRANS. BLOCK]	55	37	7	87	57	W	119	77	w
24	18	[CANCEL]	56	38	8	88	58	X	120	78	х
25	19	[END OF MEDIUM]	57	39	9	89	59	Y	121	79	У
26	1A	[SUBSTITUTE]	58	3A	:	90	5A	Z	122	7A	z
27	18	[ESCAPE]	59	3B	;	91	5B	1	123	7B	-
28	1C	[FILE SEPARATOR]	60	3C	<	92	5C	1	124	7C	T
29	1D	[GROUP SEPARATOR]	61	3D	=	93	5D	1	125	7D	}
30	1E	[RECORD SEPARATOR]	62	3E	>	94	5E	^	126	7E	~
31	1F	IUNIT SEPARATORI	63	3F	?	95	5F		127	7F	[DEL]



Convert your name into its ASCII values



Convert your name into its Hex format



Convert your name into its Binary format



Use https://symbl.cc/ to find the unicode for the following characters:









- ê