

CC8210 – NCA210

Programação Avançada I

Prof. Reinaldo A. C. Bianchi
Prof. Isaac Jesus da Silva
Prof. Danilo H. Perico

Manipulação de Strings

Caracteres

- A representação de um **caractere** é dada por um **número** inteiro, hexadecimal ou binário
- Esse número segue um **padrão conhecido** entre diversos sistemas computacionais:
 - ASCII - American Standard Code for Information Interchange (lê-se ASC2 OU asci)
 - ISO 8859 - é extensão ao código ASCII, inclui caracteres acentuados.
 - UTF - Unicode Transformation Format

Tabela ASCII

- **Tabela ASCII:**

- 7 bits (números de 0 a 127)

- **Tabela ASCII Estendida**

- 8 bits
 - Igual a tabela ASCII, porém contém mais caracteres:
 - Além do 0 ao 127,
 - Contém do 128 até o 255 (inclui os caracteres com acentos)

Tabela ASCII - 7 bits

Dec	Hex	Oct	Chr	Dec	Hex	Oct	HTML	Chr	Dec	Hex	Oct	HTML	Chr	Dec	Hex	Oct	HTML	Chr
0	0	000	NULL	32	20	040	 	Space	64	40	100	@	@	96	60	140	`	`
1	1	001	Start of Header	33	21	041	!	!	65	41	101	A	A	97	61	141	a	a
2	2	002	Start of Text	34	22	042	"	"	66	42	102	B	B	98	62	142	b	b
3	3	003	End of Text	35	23	043	#	#	67	43	103	C	C	99	63	143	c	c
4	4	004	End of Transmission	36	24	044	$	\$	68	44	104	D	D	100	64	144	d	d
5	5	005	Enquiry	37	25	045	%	%	69	45	105	E	E	101	65	145	e	e
6	6	006	Acknowledgment	38	26	046	&	&	70	46	106	F	F	102	66	146	f	f
7	7	007	Bell	39	27	047	'	'	71	47	107	G	G	103	67	147	g	g
8	8	010	Backspace	40	28	050	((72	48	110	H	H	104	68	150	h	h
9	9	011	Horizontal Tab	41	29	051))	73	49	111	I	I	105	69	151	i	i
10	A	012	Line feed	42	2A	052	*	*	74	4A	112	J	J	106	6A	152	j	j
11	B	013	Vertical Tab	43	2B	053	+	+	75	4B	113	K	K	107	6B	153	k	k
12	C	014	Form feed	44	2C	054	,	,	76	4C	114	L	L	108	6C	154	l	l
13	D	015	Carriage return	45	2D	055	-	-	77	4D	115	M	M	109	6D	155	m	m
14	E	016	Shift Out	46	2E	056	.	.	78	4E	116	N	N	110	6E	156	n	n
15	F	017	Shift In	47	2F	057	/	/	79	4F	117	O	O	111	6F	157	o	o
16	10	020	Data Link Escape	48	30	060	0	0	80	50	120	P	P	112	70	160	p	p
17	11	021	Device Control 1	49	31	061	1	1	81	51	121	Q	Q	113	71	161	q	q
18	12	022	Device Control 2	50	32	062	2	2	82	52	122	R	R	114	72	162	r	r
19	13	023	Device Control 3	51	33	063	3	3	83	53	123	S	S	115	73	163	s	s
20	14	024	Device Control 4	52	34	064	4	4	84	54	124	T	T	116	74	164	t	t
21	15	025	Negative Ack.	53	35	065	5	5	85	55	125	U	U	117	75	165	u	u
22	16	026	Synchronous idle	54	36	066	6	6	86	56	126	V	V	118	76	166	v	v
23	17	027	End of Trans. Block	55	37	067	7	7	87	57	127	W	W	119	77	167	w	w
24	18	030	Cancel	56	38	070	8	8	88	58	130	X	X	120	78	170	x	x
25	19	031	End of Medium	57	39	071	9	9	89	59	131	Y	Y	121	79	171	y	y
26	1A	032	Substitute	58	3A	072	:	:	90	5A	132	Z	Z	122	7A	172	z	z
27	1B	033	Escape	59	3B	073	;	;	91	5B	133	[[123	7B	173	{	{
28	1C	034	File Separator	60	3C	074	<	<	92	5C	134	\	\	124	7C	174	|	
29	1D	035	Group Separator	61	3D	075	=	=	93	5D	135]]	125	7D	175	}	}
30	1E	036	Record Separator	62	3E	076	>	>	94	5E	136	^	^	126	7E	176	~	~
31	1F	037	Unit Separator	63	3F	077	?	?	95	5F	137	_	_	127	7F	177		Del

Extended ASCII - 8 bits

128	Ç	144	É	160	á	176	░	192	Ł	208	⌌	224	α	240	≡
129	ü	145	æ	161	í	177	▒	193	⊥	209	⌍	225	β	241	±
130	é	146	Æ	162	ó	178	▓	194	⌋	210	⌎	226	Γ	242	≥
131	â	147	ô	163	û	179		195	⌋	211	⌌	227	π	243	≤
132	ä	148	ö	164	ñ	180	⌋	196	—	212	⌍	228	Σ	244	∫
133	à	149	ò	165	Ñ	181	⌋	197	+	213	⌎	229	σ	245	∫
134	å	150	û	166	²	182	⌌	198	⌋	214	⌎	230	μ	246	÷
135	ç	151	ù	167	°	183	⌎	199	⌌	215	⌌	231	τ	247	≈
136	ê	152	ÿ	168	¿	184	⌋	200	⌌	216	⌋	232	Φ	248	°
137	ë	153	Ö	169	┐	185	⌌	201	⌎	217	┐	233	⊗	249	·
138	è	154	Ü	170	┐	186	⌌	202	⌌	218	┐	234	Ω	250	·
139	ï	155	◊	171	½	187	┐	203	⌍	219	■	235	δ	251	√
140	î	156	£	172	¾	188	┐	204	⌌	220	■	236	∞	252	∞
141	ì	157	⌘	173	¡	189	┐	205	=	221	■	237	φ	253	²
142	Ä	158	Ⓔ	174	«	190	┐	206	⌌	222	■	238	ε	254	■
143	Å	159	ƒ	175	»	191	┐	207	⌌	223	■	239	◊	255	

Source : www.LookupTables.com

ISO 8859

- ISO 8859 is a full series of 10 (and soon MORE) standardized multilingual single-byte coded (8bit) graphic character sets for writing in alphabetic languages:
 - ISO 8859-1: West European
 - ISO 8859-2: East European
 - ISO 8859-3: South European
 - ISO 8859-4: North European
 - ISO 8859-5: Cyrillic
 - ...

ISO 8859-1

- Latin 1

Codepage 819 - Latin 1 - ISO 8859-1

	-0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-A	-B	-C	-D	-E	-F
0-		0001	0002	0003	0004	0005	0006	0007	0008	0009	000A	000B	000C	000D	000E	000F
1-	0010	0011	0012	0013	0014	0015	0016	0017	0018	0019	001A	001B	001C	001D	001E	001F
2-		!	"	#	\$	%	&	'	()	*	+	,	-	.	/
3-	0020	0021	0022	0023	0024	0025	0026	0027	0028	0029	002A	002B	002C	002D	002E	002F
4-	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
5-	0030	0031	0032	0033	0034	0035	0036	0037	0038	0039	003A	003B	003C	003D	003E	003F
6-	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
7-	0040	0041	0042	0043	0044	0045	0046	0047	0048	0049	004A	004B	004C	004D	004E	004F
8-	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
9-	0050	0051	0052	0053	0054	0055	0056	0057	0058	0059	005A	005B	005C	005D	005E	005F
A-	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
B-	0060	0061	0062	0063	0064	0065	0066	0067	0068	0069	006A	006B	006C	006D	006E	006F
C-	p	q	r	s	t	u	v	w	x	y	z	{		}	~	
D-	0070	0071	0072	0073	0074	0075	0076	0077	0078	0079	007A	007B	007C	007D	007E	007F
E-																
F-	0080	0081	0082	0083	0084	0085	0086	0087	0088	0089	008A	008B	008C	008D	008E	008F
G-																
H-	0090	0091	0092	0093	0094	0095	0096	0097	0098	0099	009A	009B	009C	009D	009E	009F
I-		¡	¢	£	¤	¥	¦	§	¨	©	ª	«	¬	®	¯	
J-	00A0	00A1	00A2	00A3	00A4	00A5	00A6	00A7	00A8	00A9	00AA	00AB	00AC	00AD	00AE	00AF
K-	°	±	²	³	´	µ	¶	·	¸	¹	º	»	¼	½	¾	¿
L-	00B0	00B1	00B2	00B3	00B4	00B5	00B6	00B7	00B8	00B9	00BA	00BB	00BC	00BD	00BE	00BF
M-	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
N-	00C0	00C1	00C2	00C3	00C4	00C5	00C6	00C7	00C8	00C9	00CA	00CB	00CC	00CD	00CE	00CF
O-	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
P-	00D0	00D1	00D2	00D3	00D4	00D5	00D6	00D7	00D8	00D9	00DA	00DB	00DC	00DD	00DE	00DF
Q-	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
R-	00E0	00E1	00E2	00E3	00E4	00E5	00E6	00E7	00E8	00E9	00EA	00EB	00EC	00ED	00EE	00EF
S-	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ
T-	00F0	00F1	00F2	00F3	00F4	00F5	00F6	00F7	00F8	00F9	00FA	00FB	00FC	00FD	00FE	00FF

ISO 8859-1

- Latin 1

	-0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-A	-B	-C	-D	-E	-F
0-		001	002	003	004	005	006	007	008	009	00A	00B	00C	00D	00E	00F
1-	0010	0011	0012	0013	0014	0015	0016	0017	0018	0019	001A	001B	001C	001D	001E	001F
2-		!	"	#	\$	%	&	'	()	*	+	,	-	.	/
3-	0020	0021	0022	0023	0024	0025	0026	0027	0028	0029	002A	002B	002C	002D	002E	002F
4-	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
5-	0030	0031	0032	0033	0034	0035	0036	0037	0038	0039	003A	003B	003C	003D	003E	003F
6-	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
7-	0040	0041	0042	0043	0044	0045	0046	0047	0048	0049	004A	004B	004C	004D	004E	004F
8-	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
9-	0050	0051	0052	0053	0054	0055	0056	0057	0058	0059	005A	005B	005C	005D	005E	005F
A-	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
B-	0060	0061	0062	0063	0064	0065	0066	0067	0068	0069	006A	006B	006C	006D	006E	006F
C-	p	q	r	s	t	u	v	w	x	y	z	{		}	~	
D-	0070	0071	0072	0073	0074	0075	0076	0077	0078	0079	007A	007B	007C	007D	007E	007F
E-																
F-	0080	0081	0082	0083	0084	0085	0086	0087	0088	0089	008A	008B	008C	008D	008E	008F
G-																
H-	0090	0091	0092	0093	0094	0095	0096	0097	0098	0099	009A	009B	009C	009D	009E	009F
I-		;	€	£	¤	¥	¦	§	¨	©	ª	«	¬	®	¯	
J-	00A0	00A1	00A2	00A3	00A4	00A5	00A6	00A7	00A8	00A9	00AA	00AB	00AC	00AD	00AE	00AF
K-	°	±	²	³	´	µ	¶	·	¸	¹	º	»	¼	½	¾	¿
L-	00B0	00B1	00B2	00B3	00B4	00B5	00B6	00B7	00B8	00B9	00BA	00BB	00BC	00BD	00BE	00BF
M-	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
N-	00C0	00C1	00C2	00C3	00C4	00C5	00C6	00C7	00C8	00C9	00CA	00CB	00CC	00CD	00CE	00CF
O-	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
P-	00D0	00D1	00D2	00D3	00D4	00D5	00D6	00D7	00D8	00D9	00DA	00DB	00DC	00DD	00DE	00DF
Q-	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
R-	00E0	00E1	00E2	00E3	00E4	00E5	00E6	00E7	00E8	00E9	00EA	00EB	00EC	00ED	00EE	00EF
S-	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ
T-	00F0	00F1	00F2	00F3	00F4	00F5	00F6	00F7	00F8	00F9	00FA	00FB	00FC	00FD	00FE	00FF

Sempre igual

Muda com a língua

ISO 8859-1: West European

A0	A1	A2	A3	A4	A5	A6	A7	A8	A9	AA	AB	AC	AD	AE	AF
	í	ç	£	¥	¥	!	§		©	≡	«	¬	-	®	_
B0	°	±	²	³	µ	¶	·	¸	¹	º	»	¼	½	¾	¿
C0	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î
D0	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ
E0	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î
F0	ä	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	ÿ

French (fr), Spanish (es), Catalan (ca), Basque (eu), Portuguese (pt), Italian (it), Albanian (sq), Rhaeto-Romanic (rm), Dutch (nl), German (de), Danish (da), Swedish (sv), Norwegian (no), Finnish (fi), Faroese (fo), Icelandic (is), Irish (ga), Scottish (gd), and English (en), incidentally also Afrikaans (af) and Swahili (sw), thus in effect also the entire American continent, Australia and much of Africa.

ISO 8859-2: East European

A0	A1	A2	A3	A4	A5	A6	A7	A8	A9	AA	AB	AC	AD	AE	AF
	À	Á	Â	Ã	Ä	Å	Š	Ŝ	Š	Ť	Ž	-	Ž	Ž	
B0	B1	B2	B3	B4	B5	B6	B7	B8	B9	BA	BB	BC	BD	BE	BF
	à	á	â	ã	ä	å	š	ŝ	š	ť	ž		ž	ž	
C0	C1	C2	C3	C4	C5	C6	C7	C8	C9	CA	CB	CC	CD	CE	CF
Ř	Ā	Ā	Ā	Ā	Ĺ	Č	Ç	Č	Ě	Ě	Ě	Ě	Ī	Ī	Ď
D0	D1	D2	D3	D4	D5	D6	D7	D8	D9	DA	DB	DC	DD	DE	DF
Đ	Ñ	Ñ	Ō	Ô	Ō	Ö	×	Ř	Ů	Ů	Ů	Ů	Ý	Ť	ß
E0	E1	E2	E3	E4	E5	E6	E7	E8	E9	EA	EB	EC	ED	EE	EF
ř	ā	ā	ā	ā	ĺ	č	ç	č	ě	ě	ě	ě	ī	ī	ď
F0	F1	F2	F3	F4	F5	F6	F7	F8	F9	FA	FB	FC	FD	FE	FF
đ	ñ	ñ	ō	ô	ō	ö	÷	ř	ů	ů	ů	ů	ý	ť	.

Czech (cs), Hungarian (hu), Polish (pl), Romanian (ro), Croatian (hr), Slovak (sk), Slovenian (sl), Sorbian.

ISO 8859-3: South European

A0	A1	A2	A3	A4		A6	A7	A8	A9	AA	AB	AC	AD		AF
	Å	Ä	£	¤		Â	Š	¨	Í	Š	Ğ	Ĵ	-		Ž
B0	°	ħ	²	³	µ	ĥ	·	,	ı	Ş	ğ	ĵ	¾		Ž
C0	À	Á	Â		Ä	Å	Ç	È	É	Ê	Ë	Ĭ	Í	Î	Ï
	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ğ	Ù	Ú	Û	Ü	Ý	Ş	ß
E0	à	á	â		ä	å	ç	è	é	ê	ë	î	í	î	ï
	ñ	ò	ó	ô	õ	ö	÷	ğ	ù	ú	û	ü	ý	ş	.

ISO 8859-5: Cyrillic

A0	A1	A2	A3	A4	A5	A6	A7	A8	A9	AA	AB	AC	AD	AE	AF
	Ё	Ђ	Ѓ	Є	Ѕ	І	Ї	Ј	Љ	Њ	Ћ	Ќ	–	Ў	Ч
B0	А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О
C0	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю
D0	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о
E0	р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю
F0	ё	ђ	ѓ	є	ѕ	і	ї	ј	љ	њ	ћ	ќ	ѕ	ў	ч

Bulgarian (bg), Byelorussian (be), Macedonian (mk), Russian (ru), Serbian

Unicode

- Unicode é um padrão que permite aos computadores representar e manipular, de forma consistente qualquer sistema de escrita existente.
- O padrão consiste de quase 138 mil caracteres:
 - Define uma metodologia para codificação e um conjunto de codificações padrões de caracteres, uma enumeração de propriedades de caracteres como caixa alta e caixa baixa, além de regras para normalização, decomposição, ordenação alfabética e renderização.
 - Os 256 primeiros códigos Unicode são idênticos aos do padrão [ISO 8859-1](#)


































































































UTF-8

- Um dos padrões mais utilizados da atualidade é o UTF-8 (8-bit Unicode Transformation Format)
- É um tipo de codificação binária (Unicode) de comprimento variável.
- UTF-8 pode representar qualquer caractere universal padrão do Unicode, sendo também compatível com o ASCII
 - Por esta razão, está lentamente a ser adaptado como tipo de codificação padrão para e-mail, páginas web, e outros locais onde os caracteres são armazenados.

A:A B:B C:C D:D E:E

Unicode

U+1F600 : 😄	U+263A : 😊	U+1F61F : 😞	U+1F61E : 😞
U+1F603 : 😊	U+1F61A : 😞	U+1F641 : 😞	U+1F648 : 🙈
U+1F604 : 😊	U+1F619 : 😞	U+2639 : 😞	U+1F649 : 🙈
U+1F601 : 😊	U+1F60B : 😊	U+1F62E : 😞	U+1F64A : 🙈
U+1F606 : 😊	U+1F61B : 😊	U+1F62F : 😞	U+1F44B : 🙋
U+1F605 : 😊	U+1F61C : 😊	U+1F632 : 😞	U+1F91A : 🙋
U+1F923 : 🙋	U+1F92A : 😊	U+1F633 : 😞	U+1F590 : 🙋
U+1F602 : 😊	U+1F61D : 😊	U+1F626 : 😞	U+270B : 🙋
U+1F642 : 😊	U+1F911 : 🙋	U+1F627 : 😞	U+1F596 : 🙋
U+1F643 : 😊	U+1F917 : 🙋	U+1F628 : 😞	U+1F44C : 🙋
U+1F609 : 😊	U+1F92D : 🙋	U+1F630 : 🙋	U+270C : 🙋
U+1F60A : 😊	U+1F92B : 🙋	U+1F625 : 🙋	U+1F91E : 🙋
U+1F607 : 😊	U+1F914 : 🙋	U+1F622 : 🙋	U+1F91F : 🙋
U+1F60D : 😊	U+1F60E : 🙋	U+1F62D : 🙋	U+1F918 : 🙋
U+1F929 : 🙋	U+1F913 : 🙋	U+1F631 : 🙋	U+1F919 : 🙋
U+1F618 : 🙋	U+1F9D0 : 🙋	U+1F616 : 🙋	U+1F44D : 🙋
U+1F617 : 🙋	U+1F615 : 🙋	U+1F623 : 🙋	U+1F44E : 🙋

№	Code	Brow.	Chart	Apple	Goog.	One	Twtr.	Wind.	GMail	DCM	KDDI	SB	Name
1	U+1F600									missing	missing	missing	GRINNING FACE
2	U+1F601												GRINNING FACE WITH SMILING EYES
3	U+1F602									missing			FACE WITH TEARS OF JOY
4	U+1F603												SMILING FACE WITH OPEN MOUTH
5	U+1F604									missing	missing		SMILING FACE WITH OPEN MOUTH AND SMILING EYES
6	U+1F605										missing	missing	SMILING FACE WITH OPEN MOUTH AND COLD SWEAT
7	U+1F606										missing	missing	SMILING FACE WITH OPEN MOUTH AND TIGHTLY-CLOSED EYES
8	U+1F609												WINKING FACE
9	U+1F60A									missing			SMILING FACE WITH SMILING EYES
10	U+1F60B										missing	missing	FACE SAVOURING DELICIOUS FOOD

Unicode em Python

- *print()* com unicode:

```
print(u'\u00ae')
```

®

```
print(u'\u0061')
```

a

```
print(u'\u00c7')
```

ç

- No Python, aspas simples ou duplas são intercambiáveis para representar strings

Unicode em Python

```
# grinning face  
print("\U0001f600")
```



```
# grinning squinting face  
print("\U0001F606")
```



```
# rolling on the floor laughing  
print("\U0001F923")
```



ord()

- Imprimir o valor ASCII/Unicode de um caractere - função ord():

```
print( ord("a") )
```

97

```
print( ord("A") )
```

65

```
print( ord("á") )
```

225

Unicode – ord()

```
print(ord('a'))
```

97

```
print(ord('â'))
```

225

```
print(ord('ç'))
```

231

```
print(ord('😊'))
```

128512

```
print(hex(ord('😊')))
```

0x1f600

chr()

- Imprimir o **caractere** a partir de um valor **ASCII/Unicode** - função **chr()**:

```
print( chr(97) )
```

a

```
print( chr(65) )
```

A

```
print( chr(225) )
```

á

Exemplo: entrada em hexadecimal

```
print( chr(0xe1) )
```

á

Strings

- Strings são cadeias de caracteres:
 - Sequência de caracteres
- String em Python são cercados por aspas simples ou aspas duplas:
 - 'hello' é igual a "hello".
- Exemplo:
 - `print("Hello")`
 - `print('Hello')`

Strings – aspas triplas

- O Python também têm a opção de aspas triplas: `"""texto"""`
- Este comando pode ser chamado de **bloco de string**
- Aspas triplas são bastante úteis para textos com múltiplas linhas
- Exemplo:

```
print("""Este  
texto  
tem  
muitas  
linhas""")
```

```
Este  
texto  
tem  
muitas  
linhas
```


Atribuir string a uma variável

- Para atribuir uma string a uma variável se faz como se fosse qualquer outro literal:
 - Com o nome variável seguido por um sinal igual e a sequência:
- Exemplo:
 - `a = "Hello"`
 - `print(a)`

Caracteres de escape

- Para inserir caracteres que são ilegais em uma sequência, use um caractere de escape – *Escape Character*
- Um caractere de escape é uma barra invertida seguido pelo caractere que você deseja inserir.
- Um exemplo de um caractere ilegal é uma citação dupla dentro de uma String que é cercada por citações duplas:
 - txt = "We are the so-called "Vikings" from the north."
 - Gera erro! - Deve ser usado \"

Caracteres de escape

- `\'` Single Quote
- `\\` Backslash
- `\n` New Line
- `\r` Carriage Return
- `\t` Tab
- `\b` Backspace
- `\f` Form Feed
- `\xhh` Hex value

```
txt = "We are the so-called \"Vikings\" from the north."  
print(txt)
```

```
We are the so-called "Vikings" from the north.
```

Strings são Arrays !!!

- Strings are Arrays
- Como muitas outras linguagens de programação populares, strings em Python são arrays, **tuplas** de bytes representando caracteres Unicode.
- No entanto, Python não tem um tipo de dados de caractere:
 - um único caractere é simplesmente uma sequência com um comprimento de 1.
- Colchetes podem ser usados para acessar elementos da String.

Strings – Índices

- De uma maneira geral, *strings* são **tuplas** de caracteres
- Podemos acessar cada caractere utilizando o índice de sua posição dentro da *String*
- Exemplos:

```
s = "spam"
print(s[0])
```

s

```
s = "spam"
print(s[2])
```

a

```
s = "spam"
print(s[-2])
```

a

Strings – Fatiamento (*slicing*)

- Podemos também utilizar o fatiamento (*slicing*) nas strings
- O fatiamento serve para extrairmos uma seção específica da *string*
- Exemplos:

```
s = "spam"  
print(s[1:3])
```

pa

```
s = "spam"  
print(s[:3])
```

spa

```
s = "spam"  
print(s[1:])
```

pam

Strings – Imutabilidade

- As strings seguem o conceito de imutabilidade
- Isso quer dizer que não é possível alterar um único caractere de uma string
- Por exemplo: Alterar o “s” por “z” na palavra “spam”

```
s = "spam"
```

```
s[0] = "z"
```

```
-----  
TypeError                                Traceback (most recent call last)
```

```
<ipython-input-13-6ae494b2e9c0> in <module>
```

```
1 s = "spam"
```

```
2
```

```
----> 3 s[0] = "z"
```

```
TypeError: 'str' object does not support item assignment
```

Strings – Imutabilidade

- Para alterarmos um ou mais caracteres de uma string, normalmente criamos outra variável
- Exemplo: alterar 's' por 'z' na palavra "spam"

```
s = "spam"

s_novo = "z" + s[1:]

print(s_novo)
```

zpam

Strings – Métodos

- Existem muitos métodos próprios para serem utilizados com strings
- Alguns exemplos:
 - `capitalize()`
 - `count()`
 - `find()`
 - `lower()`
 - `islower()`
 - `isdigit()`
 - `isalpha()`
 - `isupper()`
 - `split()`
 - `strip()`
 - `replace()`

Strings – Métodos

```
s = "spammy"  
# conta as ocorrências de "m"  
s.count("m")
```

2

```
# coloca a primeira letra em maiúscula  
s.capitalize()
```

'Spammy'

```
# transforma tudo em maiúsculo  
s.upper()
```

'SPAMMY'

```
# retorna o índice de "a"  
s.find("a")
```

2

```
# verifica se os caracteres são minúsculos  
s.islower()
```

True

```
# substitui "p" por "d"  
s.replace("p", "d")
```

'sdammy'

```
# divide uma frase no argumento utilizado  
s = "teste*de*spam"  
s.split("*")
```

['teste', 'de', 'spam']

```
# remove o espaço vazio no começo e no fim da string  
s = "teste de spam\n"  
s.strip()
```

'teste de spam'

Strings – Métodos

```
s = "22"  
s.isdigit()
```

True

```
s = "22"  
s.isalpha()
```

False

```
s = "Aula"  
s.isdigit()
```

False

```
s = "Aula"  
s.isalpha()
```

True

```
s = "Aula de Python, Python 3"  
# conta as ocorrências da palavra "Python"  
s.count("Python")
```

2

```
s = "  Aula de Python, Python 3  "  
# remove o espaço vazio no começo e no fim da String  
s.strip()
```

'Aula de Python, Python 3'

```
s = "Aula de Python, Python 3"  
# encontra a primeira ocorrência da palavra "Python"  
s.find("Python")
```

8

```
s = "Aula de Python, Python 3"  
# substitui "Python" por "C++"  
s.replace("Python", "C++")
```

'Aula de C++, C++ 3'

Strings – Método len()

- Assim como nas listas ou tuplas, podemos utilizar também o método len() para obter o tamanho da string.
- Tamanho significa o número de caracteres
- Exemplo:

```
s = "olá mundo!"  
print(len(s))  
  
10
```

String – Método format()

- Python 3 introduziu uma nova maneira de fazer a formatação de strings que também foi posteriormente portado para Python 2.7.
- Esta formatação de string "novo estilo" se livra da sintaxe usando o “%” e torna a sintaxe para formatação de strings mais regular.
- A formatação agora é tratada chamando .format() em um objeto de sequência.

String – Método format()

```
name = "John"  
'Hello, {}'.format(name)  
  
'Hello, John'
```

```
age = 36  
txt = "My name is John, and I am {}"  
print(txt.format(age))  
  
My name is John, and I am 36
```

String – Método format()

```
name = "Bob"  
error = 0xbadc0ffee  
print('Hey {name}, there is a 0x{errno:x} error!'.format(name=name, errno=error))
```

```
Hey Bob, there is a 0xbadc0ffee error!
```

String – Método format()

```
quantity = 3
itemno = 567
price = 49.95
myorder = "I want {} pieces of item {} for {} dollars."
print(myorder.format(quantity, itemno, price))

I want 3 pieces of item 567 for 49.95 dollars.
```


Concatenação de Strings

- Para concatenar, ou combinar, duas strings você pode usar o operador +.
- Exemplo:

```
a = "Hello"  
b = "World"  
c = a + b  
print(c)
```

```
a = "Hello"  
b = "World"  
c = a + b  
print(c)
```

```
HelloWorld
```

Conclusão

- Vimos na aula de hoje alguns comandos básicos para trabalhar com strings em Python...
- Já podemos começar a fazer exercícios!

Exercício em aula de teoria

- Peça ao usuário uma string e imprima se essa string é um palíndromo ou não.
- Palíndromo é uma palavra ou frase (normalmente, ignorando-se os espaços em branco) que se pode ler, indiferentemente, da esquerda para a direita ou vice-versa.
- Exemplos: "ovo"; "a grama é amarga"; "A sacada da casa", "Luz azul".

Exercício em aula de teoria

- Faça um algoritmo que conte a quantidade de incidências de todas as palavras em uma String, assim listando todas as palavras e suas quantidades, considere como palavras as que tenha uma quantidade igual ou maior que duas letras.

Fim



light roast comics