

Table 16. Frequencies of the Characters in the Latin 500K Sample.

Character	CODES		Character class code	Count	Percent in total sample	Percent in char. class	Occurrence in total sample	Occurrence in char. class
	ASCII	Unicode						
Blank	32	20	S	73,107	14.62	83.81	6.84	1.19
e	101	65	L	49,462	9.89	12.26	10.11	8.15
i	105	69	L	39,125	7.83	9.70	12.78	10.31
a	97	61	L	38,881	7.78	9.64	12.86	10.37
u	117	75	L	33,438	6.69	8.29	14.95	12.06
t	116	74	L	32,893	6.58	8.16	15.20	12.26
s	115	73	L	30,866	6.17	7.65	16.20	13.07
r	114	72	L	28,928	5.79	7.17	17.28	13.94
n	110	6e	L	24,524	4.90	6.08	20.39	16.45
m	109	6d	L	20,647	4.13	5.12	24.22	19.53
o	111	6f	L	20,341	4.07	5.04	24.58	19.83
c	99	63	L	15,849	3.17	3.93	31.55	25.45
l	108	6c	L	13,400	2.68	3.32	37.31	30.10
p	112	70	L	10,208	2.04	2.53	48.98	39.51
d	100	64	L	10,072	2.01	2.50	49.64	40.05
,	44	2c	S	7,310	1.46	8.38	68.40	11.93
q	113	71	L	7,185	1.44	1.78	69.59	56.14
v	118	76	L	6,191	1.24	1.53	80.76	65.15
b	98	62	L	5,850	1.17	1.45	85.47	68.95
g	103	67	L	5,379	1.08	1.33	92.95	74.98

Table 16. (continued)

Character	CODES		Character class code	Count	Percent in total sample	Percent in char. class	Occurrence in total sample	Occurrence in char. class
	ASCII	Unicode						
f	102	66	L	4,292	0.86	1.06	116.50	93.97
.	46	2e	S	3,082	0.62	3.53	162.23	28.30
h	104	68	L	3,011	0.60	0.75	166.06	133.96
x	120	78	L	1,952	0.39	0.48	256.15	206.63
A	65	41	U	1,280	0.26	13.59	390.63	7.36
T	84	54	U	1,095	0.22	11.63	456.62	8.60
;	59	3b	S	1,079	0.22	1.24	463.39	80.85
I	73	49	U	965	0.19	10.25	518.13	9.76
:	58	3a	S	909	0.18	1.04	550.06	95.97
y	121	79	L	752	0.15	0.19	664.89	536.36
P	80	50	U	637	0.13	6.76	784.93	14.79
S	83	53	U	600	0.12	6.37	833.33	15.70
H	72	48	U	555	0.11	5.89	900.90	16.97
N	78	4e	U	517	0.10	5.49	967.12	18.22
C	67	43	U	480	0.10	5.10	1,041.67	19.62
E	69	45	U	464	0.09	4.93	1,077.59	20.30
L	76	4c	U	435	0.09	4.62	1,149.43	21.65
D	68	44	U	423	0.08	4.49	1,182.03	22.27
`	96	60	S	408	0.08	0.47	1,225.49	213.81
'	39	27	S	394	0.08	0.45	1,269.04	221.40
M	77	4d	U	383	0.08	4.07	1,305.48	24.59
?	63	3f	S	370	0.07	0.42	1,351.35	235.76
-	45	2d	S	321	0.06	0.37	1,557.63	271.75
Q	81	51	U	286	0.06	3.04	1,748.25	32.93
O	79	4f	U	278	0.06	2.95	1,798.56	33.88
R	82	52	U	237	0.05	2.52	2,109.70	39.74
V	86	56	U	230	0.05	2.44	2,173.91	40.95
F	70	46	U	147	0.03	1.56	3,401.36	64.07
G	71	47	U	136	0.03	1.44	3,676.47	69.26
B	66	42	U	120	0.02	1.27	4,166.67	78.49
U	85	55	U	114	0.02	1.21	4,385.96	82.62
!	33	21	S	102	0.02	0.12	4,901.96	855.23
(	40	28	S	47	0.01	0.05	10,638.30	1,856.02
)	41	29	S	45	0.01	0.05	11,111.11	1,938.51
ë	235	eb	L	30	0.01	0.01	16,666.67	13,444.63
[	91	5b	S	28	0.01	0.03	17,857.14	3,115.46
]	93	5d	S	28	0.01	0.03	17,857.14	3,115.46
z	122	7a	L	26	0.01	0.01	19,230.77	15,513.04
i	239	ef	L	23	0.00	0.01	21,739.13	17,536.48
Z	90	5a	U	16	0.00	0.17	31,250.00	588.69

Table 16. (continued)

Character	CODES		Character class code	Count	Percent in total sample	Percent in char. class	Occurrence in total sample	Occurrence in char. class
	ASCII	Unicode						
X	88	58	U	12	0.00	0.13	41,666.67	784.92
K	75	4b	U	9	0.00	0.10	55,555.56	1,046.56
ü	252	fc	L	9	0.00	0.00	55,555.56	44,815.44
Ñ	209	d1	O	7	0.00	100.00	71,428.57	1.00
ö	246	f6	L	5	0.00	0.00	100,000.00	80,667.80
~	126	7e	S	2	0.00	0.00	250,000.00	43,616.50
l	49	31	N	2	0.00	100.00	250,000.00	1.00
¢	162	a2	S	1	0.00	0.00	500,000.00	87,233.00
Total				500,000				