

Table 6. Frequencies of the Characters in the Italian 1.5M 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	248,071	16.54	78.46	6.05	1.27
i	105	69	L	132,077	8.81	11.62	11.36	8.61
a	97	61	L	126,461	8.43	11.13	11.86	8.99
e	101	65	L	126,425	8.43	11.12	11.86	8.99
o	111	6f	L	105,062	7.00	9.24	14.28	10.82
n	110	6e	L	83,389	5.56	7.34	17.99	13.63
t	116	74	L	78,006	5.20	6.86	19.23	14.57
r	114	72	L	76,712	5.11	6.75	19.55	14.82
l	108	6c	L	72,419	4.83	6.37	20.71	15.69
s	115	73	L	56,964	3.80	5.01	26.33	19.95

Table 6. (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	70	46	U	1,260	0.08	3.24	1,190.48	30.90
ù	249	f9	L	1,096	0.07	0.10	1,368.61	1,037.02
*	42	2a	S	1,075	0.07	0.34	1,395.35	294.11
2	50	32	N	937	0.06	11.28	1,600.85	8.86
U	85	55	U	931	0.06	2.39	1,611.17	41.82
3	51	33	N	767	0.05	9.24	1,955.67	10.83
8	56	38	N	760	0.05	9.15	1,973.68	10.93
V	86	56	U	756	0.05	1.94	1,984.13	51.51
k	107	6b	L	713	0.05	0.06	2,103.79	1,594.07
9	57	39	N	692	0.05	8.33	2,167.63	12.00
[91	5b	S	690	0.05	0.22	2,173.91	458.22
]	93	5d	S	685	0.05	0.22	2,189.78	461.56
5	53	35	N	644	0.04	7.75	2,329.19	12.90
ì	236	ec	L	641	0.04	0.06	2,340.09	1,773.12
ò	242	f2	L	623	0.04	0.05	2,407.70	1,824.35
4	52	34	N	572	0.04	6.89	2,622.38	14.52
6	54	36	N	517	0.03	6.23	2,901.35	16.06
?	63	3f	S	496	0.03	0.16	3,024.19	637.44
y	121	79	L	483	0.03	0.04	3,105.59	2,353.15
;	59	3b	S	476	0.03	0.15	3,151.26	664.23
7	55	37	N	460	0.03	5.54	3,260.87	18.05
H	72	48	U	377	0.03	0.97	3,978.78	103.29
Q	81	51	U	368	0.02	0.95	4,076.09	105.81
é	233	e9	L	350	0.02	0.03	4,285.71	3,247.35
J	74	4a	U	258	0.02	0.66	5,813.95	150.93
x	120	78	L	252	0.02	0.02	5,952.38	4,510.20
w	119	77	L	232	0.02	0.02	6,465.52	4,899.01
%	37	25	S	222	0.01	0.07	6,756.76	1,424.20
"	34	22	S	214	0.01	0.07	7,009.35	1,477.44
K	75	4b	U	207	0.01	0.53	7,246.38	188.11
W	87	57	U	194	0.01	0.50	7,731.96	200.72
Z	90	5a	U	163	0.01	0.42	9,202.45	238.89
j	106	6a	L	150	0.01	0.01	10,000.00	7,577.14
È	200	c8	U	108	0.01	0.28	13,888.89	360.55
Y	89	59	U	79	0.01	0.20	18,987.34	492.90
/	47	2f	S	60	0.00	0.02	25,000.00	5,269.53
!	33	21	S	51	0.00	0.02	29,411.76	6,199.45
°	176	b0	S	31	0.00	0.01	48,387.10	10,199.10
X	88	58	U	29	0.00	0.07	51,724.14	1,342.72
,	184	b8	S	8	0.00	0.00	187,500.00	39,521.50

Table 6. (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						
c	99	63	L	47,968	3.20	4.22	31.27	23.69
d	100	64	L	43,191	2.88	3.80	34.73	26.31
u	117	75	L	32,708	2.18	2.88	45.86	34.75
p	112	70	L	31,418	2.09	2.76	47.74	36.18
m	109	6d	L	28,517	1.90	2.51	52.60	39.86
g	103	67	L	18,815	1.25	1.66	79.72	60.41
v	118	76	L	17,098	1.14	1.50	87.73	66.47
+	43	2b	S	16,852	1.12	5.33	89.01	18.76
,	44	2c	S	14,405	0.96	4.56	104.13	21.95
h	104	68	L	11,612	0.77	1.02	129.18	97.88
f	102	66	L	11,597	0.77	1.02	129.34	98.01
z	122	7a	L	11,096	0.74	0.98	135.18	102.43
.	46	2e	S	10,514	0.70	3.33	142.67	30.07
b	98	62	L	9,945	0.66	0.88	150.83	114.29
'	39	27	S	6,575	0.44	2.08	228.14	48.09
q	113	71	L	4,230	0.28	0.37	354.61	268.69
I	73	49	U	3,291	0.22	8.45	455.79	11.83
A	65	41	U	3,258	0.22	8.37	460.41	11.95
è	232	e8	L	3,242	0.22	0.29	462.68	350.58
)	41	29	S	3,166	0.21	1.00	473.78	99.86
C	67	43	U	3,163	0.21	8.12	474.23	12.31
(40	28	S	3,129	0.21	0.99	479.39	101.05
à	224	e0	L	3,076	0.21	0.27	487.65	369.50
S	83	53	U	3,050	0.20	7.83	491.80	12.77
«	171	ab	S	2,908	0.19	0.92	515.82	108.72
»	187	bb	S	2,882	0.19	0.91	520.47	109.71
L	76	4c	U	2,788	0.19	7.16	538.02	13.97
M	77	4d	U	2,740	0.18	7.04	547.45	14.21
P	80	50	U	2,569	0.17	6.60	583.88	15.16
N	78	4e	U	2,103	0.14	5.40	713.27	18.52
-	45	2d	S	2,019	0.13	0.64	742.94	156.60
E	69	45	U	2,009	0.13	5.16	746.64	19.38
R	82	52	U	1,783	0.12	4.58	841.28	21.84
:	58	3a	S	1,626	0.11	0.51	922.51	194.45
G	71	47	U	1,572	0.10	4.04	954.20	24.77
O	48	30	N	1,528	0.10	18.40	981.68	5.44
B	66	42	U	1,512	0.10	3.88	992.06	25.75
D	68	44	U	1,506	0.10	3.87	996.02	25.86
T	84	54	U	1,434	0.10	3.68	1,046.03	27.15
O	79	4f	U	1,429	0.10	3.67	1,049.69	27.25
I	49	31	N	1,428	0.10	17.19	1,050.42	5.82

Table 6. (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						
a	170	aa	S	7	0.00	0.00	214,285.71	45,167.43
â	226	e2	O	4	0.00	30.77	375,000.00	3.25
>	62	3e	S	4	0.00	0.00	375,000.00	79,043.00
ˆ	168	a8	S	3	0.00	0.00	500,000.00	105,390.67
í	237	ed	L	3	0.00	0.00	500,000.00	378,857.00
ñ	241	f1	O	3	0.00	23.08	500,000.00	4.33
ó	243	f3	O	2	0.00	15.38	750,000.00	6.50
ê	234	ea	O	2	0.00	15.38	750,000.00	6.50
ü	252	fc	O	2	0.00	15.38	750,000.00	6.50
	124	7c	S	1	0.00	0.00	1,500,000.00	316,172.00
İ	204	cc	U	1	0.00	0.00	1,500,000.00	38,939.00
É	201	c9	U	1	0.00	0.00	1,500,000.00	38,939.00
◦	186	ba	S	1	0.00	0.00	1,500,000.00	316,172.00
`	96	60	S	1	0.00	0.00	1,500,000.00	316,172.00
Total				1,500,000				