Frequency of Character Pairs in English Language Text

Row x, column y of the table below gives an estimate of the relative frequency of the two-character sequence xy in English language text. Specifically, it estimates the number of occurrences of xy per 10,000 characters of text. The estimate applies to English language text with characters other than letters deleted, and upper and lower case letters treated as identical It is based on seven English language novels -- approximately 3.2 million characters.

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
a	1	20	33	52	0	12	18	5	39	1	12	57	26	181	1	20	1	75	95	104	9	20	13	1	26	1
b	11	1	0	0	47	0	0	0	6	1	0	17	0	0	19	0	0	11	2	1	21	0	0	0	11	0
c	31	0	4	0	38	0	0	38	10	0	18	9	0	0	45	0	1	11	1	15	7	0	0	0	1	0
d	48	20	9	13	57	11	7	25	50	3	1	11	14	16	41	6	0	14	35	56	10	2	19	0	10	0
e	110	23	45	126	48	30	15	33	41	3	5	55	47	111	33	28	2	169	115	83	6	24	50	9	26	0
f	25	2	3	2	20	11	1	8	23	1	0	8	5	1	40	2	0	16	5	37	8	0	3	0	2	0
g	24	3	2	2	28	3	4	35	18	1	0	7	3	4	23	1	0	12	9	16	7	0	5	0	1	0
h	114	2	2	1	302	2	1	6	97	0	0	2	3	1	49	1	0	8	5	32	8	0	4	0	4	0
i	10	5	32	33	23	17	25	6	1	1	8	37	37	179	24	6	0	27	86	93	1	14	7	2	0	2
j	2	0	0	0	2	0	0	0	3	0	0	0	0	0	3	0	0	0	0	0	8	0	0	0	0	0
k	6	1	1	1	29	1	0	2	14	0	0	2	1	9	4	0	0	0	5	4	1	0	2	0	2	0
l	40	3	2	36	64	10	1	4	47	0	3	56	4	2	41	3	0	2	11	15	8	3	5	0	31	0
m	44	7	1	1	68	2	1	3	25	0	0	1	5	2	29	11	0	3	10	9	8	0	4	0	18	0
n	40	7	25	146	66	8	92	16	33	2	8	9	7	8	60	4	1	3	33	106	6	2	12	0	11	0
o	16	12	13	18	5	80	7	11	12	1	13	26	48	106	36	15	0	84	28	57	115	12	46	0	5	1
p	23	1	0	0	30	1	0	3	12	0	0	15	1	0	21	10	0	18	5	11	6	0	1	0	1	0
q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0
r	50	7	10	20	133	8	10	12	50	1	8	10	14	16	55	6	0	14	37	42	12	4	11	0	21	0
s	67	11	17	7	74	11	4	50	49	2	6	13	12	10	57	20	2	4	43	109	20	2	24	0	4	0
t	59	10	11	7	75	9	3	330	76	1	2	17	11	7	115	4	0	28	34	56	17	1	31	0	16	0
u	7	5	12	7	7	2	14	2	8	0	1	34	8	36	1	16	0	44	35	48	0	0	2	0	1	0
v	5	0	0	0	65	0	0	0	11	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	1	0
w	66	1	1	2	39	1	0	44	39	0	0	2	1	12	29	0	0	3	4	4	1	0	2	0	1	0
x	1	0	2	0	1	0	0	0	2	0	0	0	0	0	0	3	0	0	0	3	0	0	0	0	0	0
y	18	7	6	6	14	7	3	10	11	1	1	4	6	3	36	4	0	3	19	20	1	1	12	0	2	0
z	1	0	0	0	3	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

The most common two-character sequences are:

Sequence	Frequency (per 10,000 chars)
th	330
he	302
an	181
in	179
er	169
nd	146
re	133
ed	126
es	115
ou	115
to	115
ha	114
en	111
ea	110
st	109
nt	106
on	106
at	104
hi	97
as	95
it	93
ng	92
is	86
or	84
et	83
of	80
ti	76

Sequence	Frequency (per 10,000 chars)
ar	75
te	75
se	74
me	68
sa	67
ne	66
wa	66
ve	65
le	64
no	60
ta	59
al	57
de	57
ot	57
so	57
dt	56
ll	56
tt	56
el	55
ro	55
ad	52
di	50
ew	50
ra	50
ri	50
sh	50