

Отчет по лабораторной работе №2

Уманский Александр и Межуев Владислав, ПИН-43

Вс сен 22 15:36:18 MSK 2019

В данной лабораторной работе мы будем исследовать следующие файлы:

- С английским текстом:
 1. Frankenstein, by Mary Wollstonecraft (Godwin) Shelley.txt
 2. Puck of Pook's Hill, by Rudyard Kipling.txt
- С русским текстом:
 1. Керниган, Ричи. Язык C.txt
 2. Рассел Сейдж. Приемы профессиональной работы в UNIX.txt
- Я пример

Урезанные таблицы частот

Для английских текстов

Название	Частота	Символ
Frankenstein, by Mary Wollstonecraft (Godwin) Shelley.txt, Всего символов 441034
	25225	o
	26743	a
	30364	t
	46043	e

Соотношения:

- Пробела и самой частой буквы:

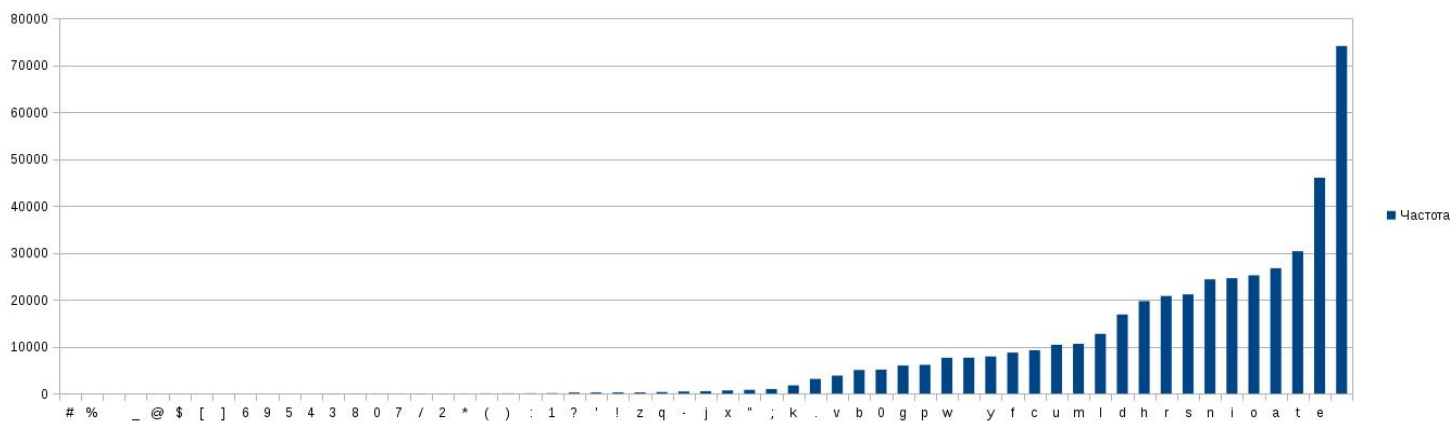
$$(7653 + 74143)/46043 = 1.7765$$

- Самой частой буквы и второй по популярности:

$$46043/30364 = 1.5163$$

- Второй и третьей по популярности буквы:

$$30364/26743 = 1.1353$$



Название	Частота	Символ
Puck of Pook's Hill, by Rudyard Kipling.txt, Всего символов 332910
	18822	o
	21016	a
	21684	t
	30952	e

Соотношения:

- Пробега и самой частой буквы:

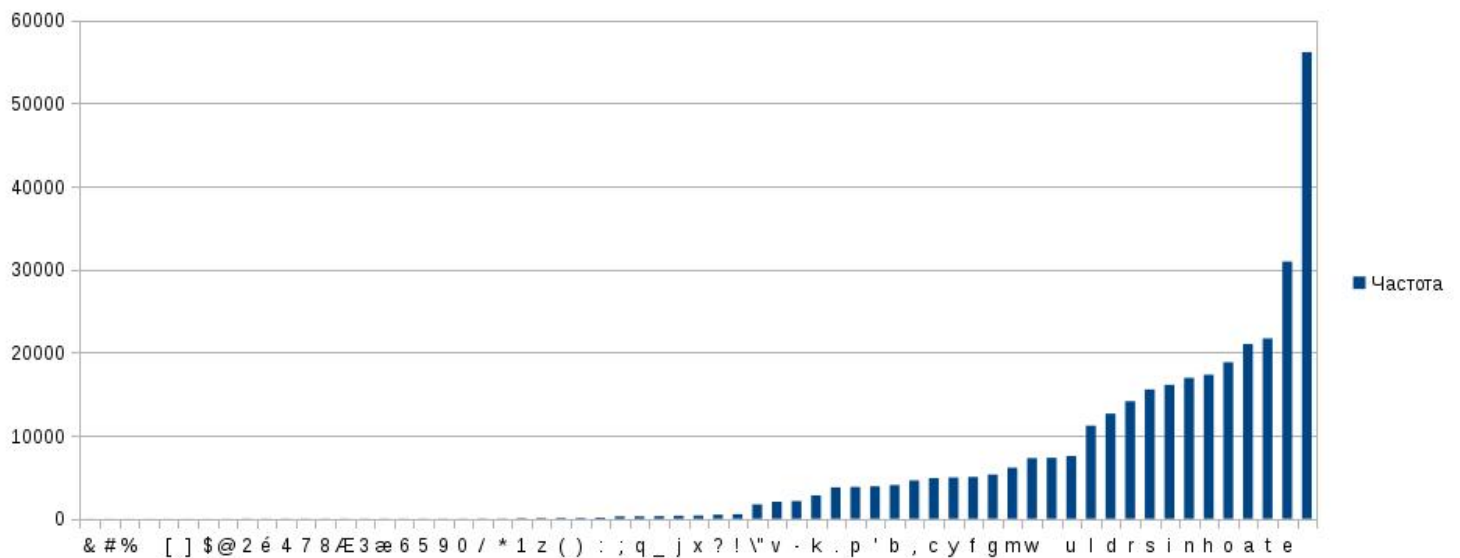
$$(7325 + 56154)/30952 = 2.0508$$

- Самой частой буквы и второй по популярности:

$$30952/21684 = 1.4274$$

- Второй и третьей по популярности буквы:

$$21684/21016 = 1.0317$$



Для русских текстов

Название	Частота	Символ
Керниган, Ричи. Язык C.txt, Всего символов 461275
	24053	а
	24307	и
	27488	е
	29507	о

Соотношения:

- Пробега и самой частой буквы:

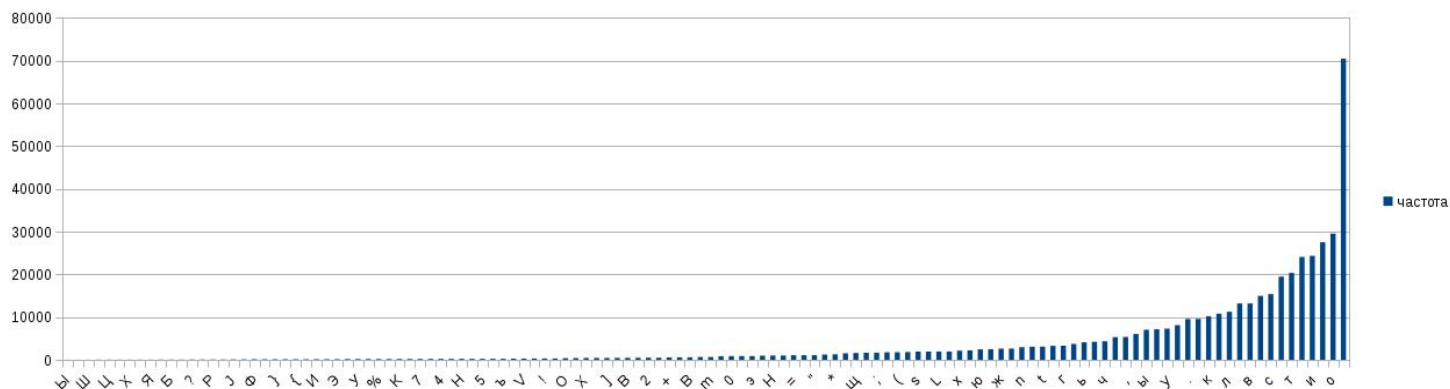
$$(13168 + 70419)/29507 = 2.8327$$

- Самой частой буквы и второй по популярности:

$$29507/27488 = 1.0734$$

- Второй и третьей по популярности буквы:

$$27488/24307 = 1.1308$$



Название	Частота	Символ
Рассел Сейдж. Приемы профессиональной работы в UNIX.txt, Всего символов 1029582
	45742	т
	52556	а
	54101	е
	68833	о

Соотношения:

- Пробега и самой частой буквы:

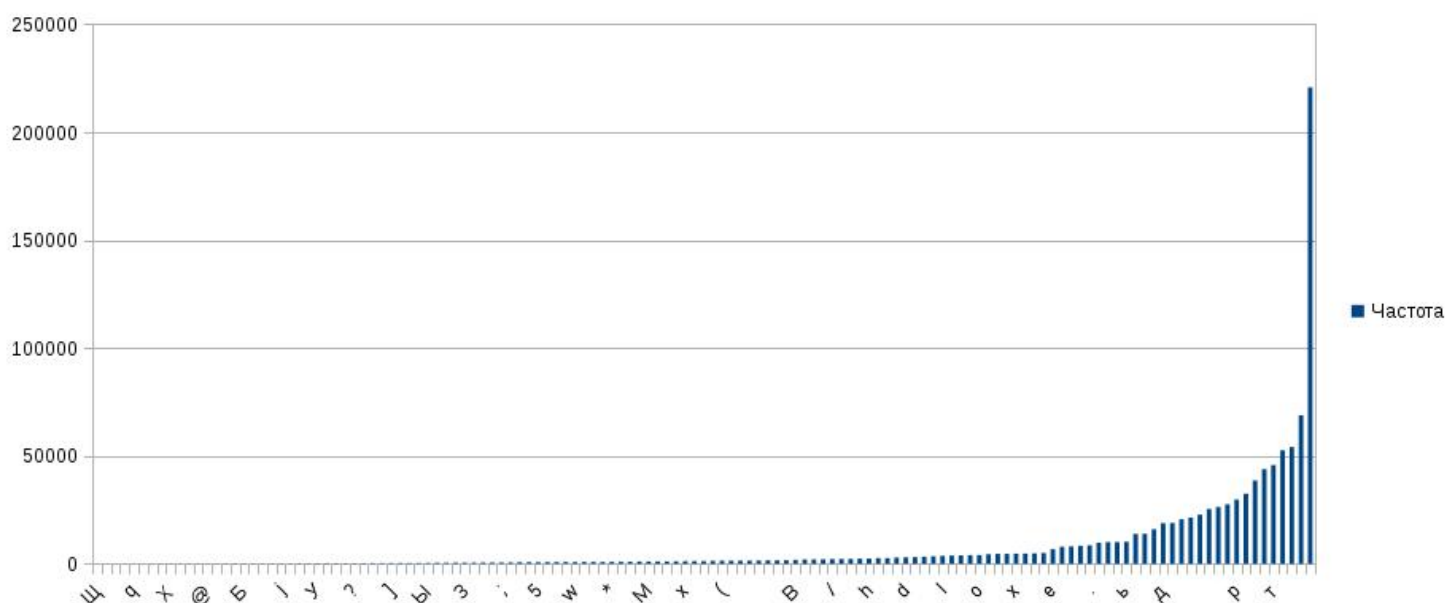
$$(22768 + 220704)/68833 = 3.5371$$

- Самой частой буквы и второй по популярности:

$$68833/54101 = 1.2723$$

- Второй и третьей по популярности буквы:

$$54101/52556 = 1.0293$$



Вывод

Соотношения частот букв не сильно отличаются в зависимости от языка. Соотношения пробельных символов и самой частой буквы разнятся сильно даже в пределах одного языка.

Первый

00000000: 6f20 6164 6d69 7265 2070 6561 6365 6162 6c65 206c o admire peaceable l
00000014: 6177 6769 7665 7273 2c20 4e75 6d61 2c0a 536f 6c6f awgivers, Numa,.Solo
00000028: 6e2c 2061 6e64 204c 7963 7572 6775 732c 2069 6e20 n, and Lycurgus, in
0000003c: 7072 6566 6572 656e 6365 2074 6f20 526f 6d75 6c75 preference to Romulu
00000050: 7320 616e 6420 5468 6573 6575 732e 2020 5468 650a s and Theseus. The.
00000064: 7061 7472 6961 7263 6861 6c20 6c69 7665 7320 6f66 patriarchal lives of
00000078: 206d 7920 7072 6f74 6563 746f 7273 2063 6175 7365 my protectors cause
0000008c: 6420 7468 6573 6520 696d 7072 6573 7369 6f6e 7320 d these impressions
000000a0: 746f 2074 616b 6520 610a 6669 726d 2068 6f6c 6420 to take a.firm hold
000000b4: 6f6e 206d 7920 6d69 6e64 3b20 7065 7268 6170 732c on my mind; perhaps,
000000c8: 2069 6620 6d79 2066 6972 7374 2069 6e74 726f 6475 if my first introdu
000000dc: 6374 696f 6e20 746f 2068 756d 616e 6974 7920 6861 ction to humanity ha
000000f0: 640a 6265 656e 206d 6164 6520 6279 2061 2079 6f75 d.been made by a you
00000104: 6e67 2073 6f6c 6469 6572 2c20 6275 726e 696e 6720 ng soldier, burning
00000118: 666f 7220 676c 6f72 7920 616e 6420 736c 6175 6768 for glory and slaugh
0000012c: 7465 722c 2049 2073 686f 756c 640a 6861 7665 2062 ter, I should.have b
00000140: 6565 6e20 696d 6275 6564 2077 6974 6820 6469 6666 een imbued with diff
00000154: 6572 656e 7420 7365 6e73 6174 696f 6e73 2e0a 0a22 erent sensations..."
00000168: 4275 7420 5061 7261 6469 7365 204c 6f73 7420 6578 But Paradise Lost ex
0000017c: 6369 7465 6420 6469 6666 6572 656e 7420 616e 6420 cited different and
00000190: 6661 7220 6465 6570 6572 2065 6d6f 7469 6f6e 732e far deeper emotions.
000001a4: 2020 4920 7265 6164 0a69 742c 2061 7320 4920 6861 I read.it, as I ha
000001b8: 6420 7265 6164 2074 6865 206f 7468 6572 2076 6f6c d read the other vol
000001cc: 756d 6573 2077 6869 6368 2068 6164 2066 616c 6c65 umes which had falle
000001e0: 6e20 696e 746f 206d 7920 6861 6e64 732c 2061 730a n into my hands, as.

Частота	Символ
0.002	;
0.002	"
0.002	k
0.002	x
0.006	.
0.006	w
0.008	v
0.012	g
0.014	b
0.018	
0.018	,
0.018	c
0.018	p
0.018	y
0.024	f
0.03	m
0.032	l
0.032	u
0.038	h
0.048	d
0.054	n
0.054	o
0.054	t
0.056	r
0.056	s
0.06	i
0.068	a
0.094	e
0.156	

Второй

00000000: 6174 6f72 3b20 6865 2077 6173 2061 6c6c 6f77 6564 ator; he was allowed
 00000014: 2074 6f20 636f 6e76 6572 7365 2077 6974 6820 616e to converse with an
 00000028: 6420 6163 7175 6972 6520 6b6e 6f77 6c65 6467 6520 d acquire knowledge
 0000003c: 6672 6f6d 0a62 6569 6e67 7320 6f66 2061 2073 7570 from.beings of a sup
 00000050: 6572 696f 7220 6e61 7475 7265 2c20 6275 7420 4920 erior nature, but I
 00000064: 7761 7320 7772 6574 6368 6564 2c20 6865 6c70 6c65 was wretched, helple
 00000078: 7373 2c20 616e 6420 616c 6f6e 652e 0a4d 616e 7920 ss, and alone..Many
 0000008c: 7469 6d65 7320 4920 636f 6e73 6964 6572 6564 2053 times I considered S
 000000a0: 6174 616e 2061 7320 7468 6520 6669 7474 6572 2065 atan as the fitter e
 000000b4: 6d62 6c65 6d20 6f66 206d 7920 636f 6e64 6974 696f mblem of my conditio
 000000c8: 6e2c 2066 6f72 0a6f 6674 656e 2c20 6c69 6b65 2068 n, for.often, like h
 000000dc: 696d 2c20 7768 656e 2049 2076 6965 7765 6420 7468 im, when I viewed th
 000000f0: 6520 626c 6973 7320 6f66 206d 7920 7072 6f74 6563 e bliss of my protec
 00000104: 746f 7273 2c20 7468 6520 6269 7474 6572 0a67 616c tors, the bitter.gal
 00000118: 6c20 6f66 2065 6e76 7920 726f 7365 2077 6974 6869 l of envy rose withi
 0000012c: 6e20 6d65 2e0a 0a22 416e 6f74 6865 7220 6369 7263 n me..."Another circ
 00000140: 756d 7374 616e 6365 2073 7472 656e 6774 6865 6e65 umstance strengthene
 00000154: 6420 616e 6420 636f 6e66 6972 6d65 6420 7468 6573 d and confirmed thes
 00000168: 6520 6665 656c 696e 6773 2e20 2053 6f6f 6e0a 6166 e feelings. Soon.af
 0000017c: 7465 7220 6d79 2061 7272 6976 616c 2069 6e20 7468 ter my arrival in th
 00000190: 6520 686f 7665 6c20 4920 6469 7363 6f76 6572 6564 e hovel I discovered
 000001a4: 2073 6f6d 6520 7061 7065 7273 2069 6e20 7468 6520 some papers in the
 000001b8: 706f 636b 6574 206f 660a 7468 6520 6472 6573 7320 pocket of.the dress
 000001cc: 7768 6963 6820 4920 6861 6420 7461 6b65 6e20 6672 which I had taken fr
 000001e0: 6f6d 2079 6f75 7220 6c61 626f 7261 746f 7279 2e20 om your laboratory.

Частота	Символ
0.002	;
0.002	"
0.002	q
0.008	.
0.008	k
0.01	g
0.012	b
0.012	p
0.012	u
0.012	v
0.014	,
0.014	y
0.016	
0.02	w
0.026	c
0.026	f
0.028	m
0.03	l
0.032	d
0.04	h
0.05	s
0.052	a
0.054	n
0.058	i
0.058	r
0.064	t
0.066	o
0.112	e
0.16	

Вывод

Частоты в фрагментах и самом файле примерно одинаковы. Порядок символов различается из-за того, что различается точность относительных частот

Полные таблицы частот

Относительная таблица частот

Название	Частота	Символ
Frankenstein, by Mary Wollstonecraft (Godwin) Shelley.txt, всего символов 441034		
	2.2674e-06	#
	2.2674e-06	%
	2.2674e-06	
	4.5348e-06	_
	4.5348e-06	@
	4.5348e-06	\$
	9.0696e-06	[
	9.0696e-06]
	2.2674e-05	6
	2.94762e-05	9
	3.62784e-05	5
	3.85458e-05	4
	4.08132e-05	3
	4.5348e-05	8
	4.76154e-05	0
	5.21502e-05	7
	5.44176e-05	/
	5.44176e-05	2
	6.34872e-05	*
	8.84286e-05	(
	8.84286e-05)
	0.000154183	:
	0.000208601	1
	0.000498828	?
	0.000519234	'
	0.000541908	!
	0.000550978	z
	0.000734637	q
	0.00100899	-
	0.00114277	j
	0.00153503	x
	0.00180485	"
	0.00219938	;
	0.00397929	k
	0.00708335	.
	0.00869094	v
	0.0113959	b
	0.0115569	,
	0.0135454	g
	0.0138787	p
	0.0173184	w
	0.0173524	
	0.0179442	y
	0.0197967	f
	0.0209576	c
	0.0235968	u
	0.0240435	m
	0.0288844	l
	0.0382351	d
	0.0447244	h
	0.0472027	r
	0.0479668	s
	0.0552497	n

Название	Частота	Символ
	0.0558075	i
	0.0571951	o
	0.060637	a
	0.0688473	t
	0.104398	e
	0.168112	

Для английских текстов

Название	Частота	Символ
Frankenstein, by Mary Wollstonecraft (Godwin) Shelley.txt, Всего символов 441034		
	1	%
	1	
	2	-
	2	@
	2	\$
	4	[
	4]
	10	6
	13	9
	16	5
	17	4
	18	3
	20	8
	21	0
	23	7
	24	/
	24	2
	28	*
	39	(
	39)
	68	:
	92	1
	220	?
	229	'
	239	!
	243	z
	324	q
	445	-
	504	j
	677	x
	796	"
	970	;
	1755	k
	3124	.
	3833	v
	5026	b
	5097	,
	5974	g
	6121	p
	7638	w
	7653	
	7914	y
	8731	f
	9243	c
	10407	u

Название	Частота	Символ
	10604	m
	12739	l
	16863	d
	19725	h
	20818	r
	21155	s
	24367	n
	24613	i
	25225	o
	26743	a
	30364	t
	46043	e
	74143	

Название	Частота	Символ
Puck of Pook's Hill, by Rudyard Kipling.txt, Всего символов 332910	1	&
	1	#
	1	%
	1	
	2	[
	2]
	2	\$
	3	@
	9	2
	9	é
	10	4
	11	7
	11	8
	12	Æ
	13	3
	13	æ
	14	6
	18	5
	18	9
	21	0
	27	/
	28	*
	66	1
	73	z
	93	(
	93)
	122	:
	274	;
	275	q
	290	_
	349	j
	370	x
	474	?
	534	!
	1716	"
	2027	v
	2117	-
	2801	k
	3765	.
	3814	p
	3885	'
	4041	b
	4599	,

Название	Частота	Символ
	4878	c
	4946	y
	5015	f
	5314	g
	6123	m
	7283	w
	7325	
	7528	u
	11191	l
	12639	d
	14130	r
	15550	s
	16105	i
	16935	n
	17315	h
	18822	o
	21016	a
	21684	t
	30952	e
	56154	

Для русских текстов

Название	Частота	Символ
Керниган, Ричи. Язык C.txt, Всего символов 461275	1	Ы
	2	@
	2	Ш
	17	Г
	20	Ц
	26	Q
	27	X
	30	^
	31	Я
	33	Л
	41	Б
	41	Ч
	55	?
	62	А
	67	Р
	68	З
	81	Ј
	96	9
	112	Ф
	114	Д
	117	}
	119	#
	120	{
	120	8
	130	И
	130	М
	142	Э
	157	Z
	159	У
	162	<
	162	%
	167	Е
	191	К

Название	Частота	Символ
	192	T
	193	7
	205	6
	211	4
	218	–
	222	H
	227	K
	234	5
	238	&
	238	ъ
	252	C
	275	V
	294	3
	306	!
	311	>
	375	O
	395	П
	420	X
	425	[
	425]
	443	:
	457	B
	464	W
	480	2
	481	Y
	557	+
	568	,
	578	B
	636	\
	842	g
	854	0
	862	/
	871	э
	960	ш
	991	H
	1004	1
	1049	=
	1071	u
	1079	"
	1219	d
	1261	*
	1508	P
	1589	щ
	1640	f
	1647	;
	1749)
	1761	(
	1790	c
	1904	s
	1907	ф
	1931	L
	1934	o
	2140	x
	2214	a
	2438	ю
	2476	R
	2607	ж
	2631	ц
	2951	n

Название	Частота	Символ
	3048	і
	3060	т
	3260	й
	3280	г
	3705	е
	4096	ь
	4208	б
	4339	ч
	5292	з
	5338	,
	6025	-
	7021	ы
	7138	д
	7296	у
	8092	я
	9533	.
	9549	п
	10188	к
	10774	м
	11229	л
	13168	
	13174	в
	14930	р
	15360	с
	19451	н
	20309	т
	24053	а
	24307	и
	27488	е
	29507	о
	70419	

Название	Частота	Символ
Рассел Сейдж. Приемы профессиональной работы в UNIX.txt, Всего символов 1029582	17	Щ
	21	Ю
	23	~
	30	Ш
	40	q
	57	%
	82	Ь
	86	z
	88	X
	88	Ц
	92	Ж
	93	ъ
	105	@
	129	<
	141	Г
	145	!
	151	Б
	152	}
	153	{
	167	&
	177	j
	184	Й
	205	^
	236	Ф

Название	Частота	Символ
	251	У
	252	‘
	260	Ч
	269	Я
	273	?
	295	>
	308	#
	351	,
	353]
	380	[
	382	Л
	385	+
	431	Ы
	493	7
	511	8
	524	9
	542	3
	544	=
	597	6
	608	k
	625	;
	638	\
	735	Д
	744	5
	812	4
	812	Р
	814	Т
	821	w
	833	у
	834	–
	849	v
	903	*
	905	К
	915	3
	1031	b
	1061	М
	1075	А
	1095	И
	1115	\$
	1222	x
	1226	g
	1263	С
	1355	:
	1407	(
	1407	0
	1446	Н
	1480	П
	1523	
	1606	2
	1621)
	1670	О
	1793	В
	1999	э
	2001	Е
	2043	f
	2172	/
	2216	р
	2266	m
	2314	1

Название	Частота	Символ
	2398	h
	2624	ш
	2635	щ
	2914	"
	2950	d
	3104	u
	3261	ц
	3452	a
	3655	l
	3818	c
	3912	г
	3992	n
	4069	o
	4428	ф
	4578	ю
	4596	i
	4690	x
	4745	s
	4815	ж
	5036	t
	6752	e
	7860	г
	8055	ч
	8321	б
	8576	.
	9701	,
	10066	й
	10072	з
	10246	ь
	13840	я
	13899	у
	16019	ы
	18830	д
	18946	п
	20630	-
	21422	к
	22768	
	25370	л
	26288	м
	27587	в
	29761	р
	32428	с
	38614	н
	43875	и
	45742	т
	52556	а
	54101	е
	68833	о
	220704	

Криптограммы

Будем исследовать криптограмму *gold-bug-en.txt* ~~~ 53†††305))6;4826)4#)4#;806;48†8¶60))85;1-(:;8-83(88)5†;46(;8896?;8) #(;484)8¶8;4069285);)6†8)4;1(†9;48081;8:8†1;48†85;4)485†528806*81(†9;48; (88;4(†?34;48)4†;161;:188;†?; ~~~

Частота	Символ	Возможные символы
0.01		

Частота	Символ	Возможные символы
0.01	¶	
0.015	-	{v,b}
0.015	?	{v,b}
0.02	:	
0.02	3	
0.025	†	
0.025	2	
0.025	9	
0.03	0	
0.04	1	
0.05	(
0.055	6	h
0.06	5	r
0.065	*	s
0.07	‡	i
0.075)	o
0.095	4	a
0.13	;	t
0.165	8	e

53‡‡‡305))6*;4826)4‡)4‡;806*;48‡8¶60))85;1-(;:*8-83(88)5*‡
;46(;88*96*?;8)*‡(;485);5*‡2:*‡(;4956*2(5*- 4)8¶8*;4069285);
)6‡8)4;1(‡9;48081;8:8‡1;48‡85;4)485‡528806*81(‡9;48;
(88;4(‡?34;48)4‡;161;:188;‡?;