

Отчет по лабораторной работе №2

Уманский Александр и Межуев Владислав, ПИН-43

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

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:

- Пробега и самой частой буквы:

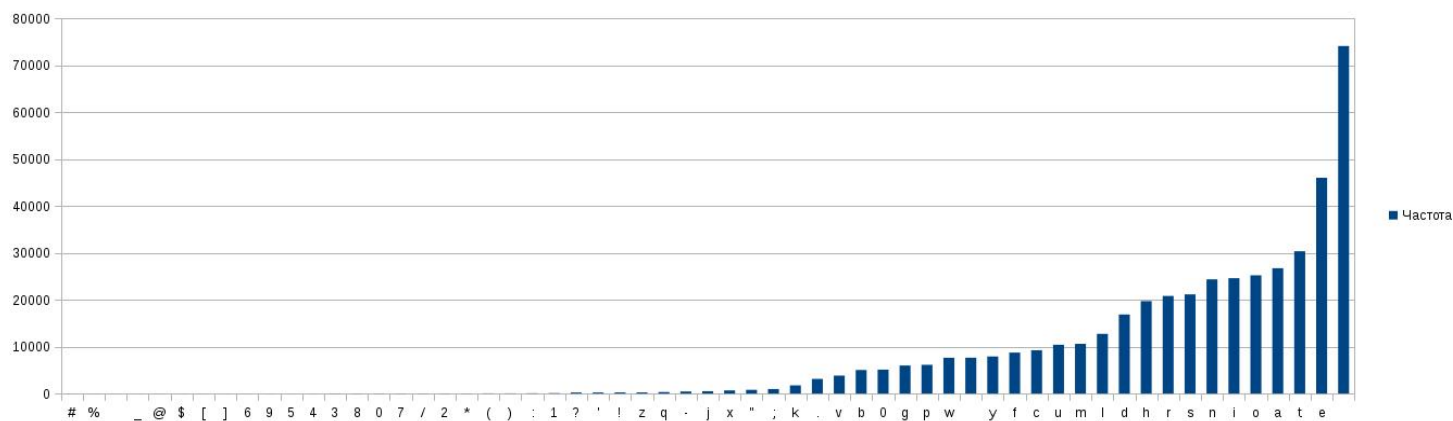
$$(7653 + 74143)/46043 = 1.7765$$

- 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:

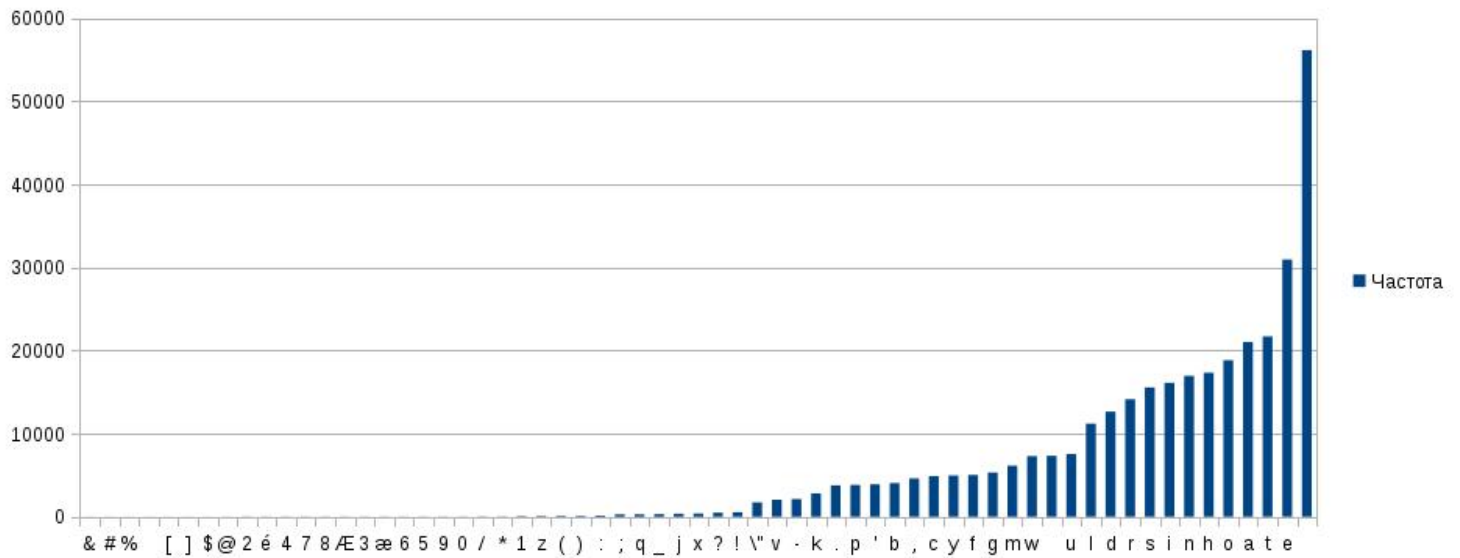
$$(7325 + 56154)/30952 = 2.0508$$

- 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:

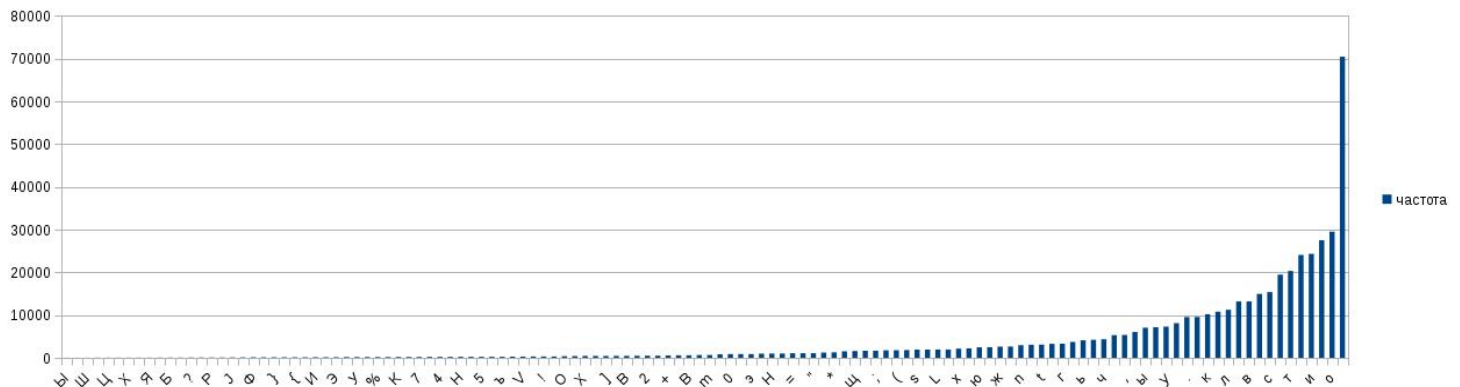
$$(13168 + 70419)/29507 = 2.8327$$

- 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:

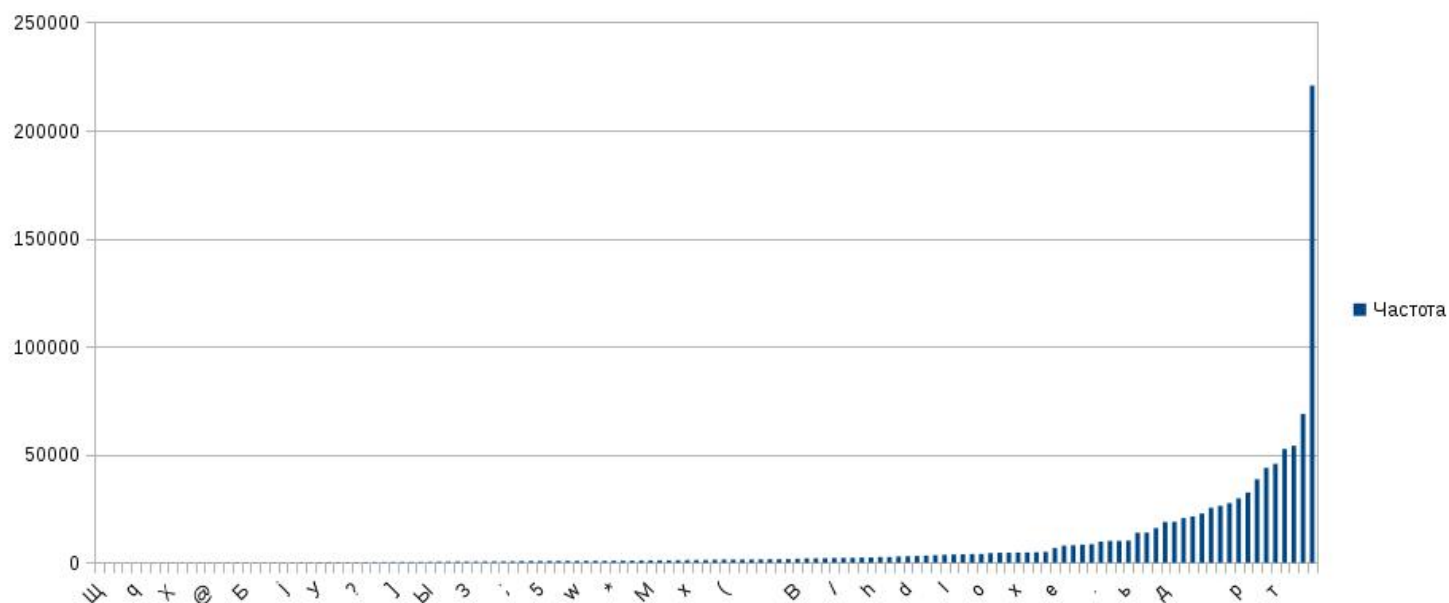
$$(22768 + 220704)/68833 = 3.5371$$

- 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.

Полные таблицы частот

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
	10604	m
	12739	l
	16863	d
	19725	h
	20818	r
	21155	s
	24367	n
	24613	i

Название	Частота	Символ
	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	,
	4878	c
	4946	y
	5015	f
	5314	g
	6123	m
	7283	w
	7325	
	7528	u

Название	Частота	Символ
	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
	192	Т
	193	7
	205	6
	211	4
	218	_
	222	Н
	227	К
	234	5

Название	Частота	Символ
	238	&
	238	ъ
	252	С
	275	V
	294	3
	306	!
	311	>
	375	О
	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
	3048	i
	3060	t
	3260	й
	3280	г
	3705	e
	4096	ь
	4208	б
	4339	ч

Название	Частота	Символ
	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	ь
	105	@
	129	<
	141	Г
	145	!
	151	Б
	152	}
	153	{
	167	&
	177	j
	184	Й
	205	^
	236	Ф
	251	У
	252	'
	260	Ч
	269	Я
	273	?
	295	>
	308	#
	351	,

Название	Частота	Символ
	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
	2398	h
	2624	ш
	2635	щ
	2914	"
	2950	d
	3104	u
	3261	ц
	3452	a

Название	Частота	Символ
	3655	l
	3818	с
	3912	г
	3992	п
	4069	о
	4428	ф
	4578	ю
	4596	i
	4690	х
	4745	s
	4815	ж
	5036	t
	6752	е
	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	