

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ  
федеральное государственное автономное образовательное учреждение высшего  
образования

«САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
АЭРОКОСМИЧЕСКОГО ПРИБОРОСТРОЕНИЯ»

КАФЕДРА № 43

ОТЧЕТ  
ЗАЩИЩЕН С ОЦЕНКОЙ  
ПРЕПОДАВАТЕЛЬ

Старший преподаватель		Фоменкова А.А.
Ассистент		Величко М.В.
должность, уч. степень, звание	подпись, дата	инициалы, фамилия

ОТЧЕТ О ЛАБОРАТОРНОЙ РАБОТЕ № 7

«Оценка количества информации в сообщении и эффективное кодирование»

по курсу: ИНФОРМАТИКА

РАБОТУ ВЫПОЛНИЛ

СТУДЕНТ ГР. №	4134К		Иванов И.В.
		подпись, дата	инициалы, фамилия

Санкт-Петербург 2021

## Цель работы:

Получение практических навыков численного определения количества информации, содержащегося в сообщении. Освоение методов построения кодов дискретного источника информации используя конструктивный метод, предложенный К.Шенноном и Н.Фано, и метод Хаффмана. На примере показать однозначность раскодирования имеющегося сообщения.

### ЧАСТЬ 1.

#### «Определение количества информации, содержащейся в сообщении»

#### Порядок выполнения части 1 лабораторной работы

1. Создать таблицу (50 рабочих строк) в Excel аналогичную рис.1.

<i>№ n/n</i>	<i>Символ</i>	<i>Код символа</i>	<i>Число вхождений символа в текст</i>	<i>Вероятность вхождения символа (<math>p_i</math>)</i>	<i><math>I_i</math></i>
1	0				
2	1				
...	...				
50	я				
		<i>Всего символов в тексте (K)</i>			

<i>Полная вероятность(P)</i>	(должна получиться «1»)	
	<i>Энтропия источника (<math>I_{ср}</math>)</i>	

Рис.1.

2. Заполнить столбец Символ следующими значениями:

- 33 буквы русского алфавита;
- 10 цифр (0 — 9);
- Знаки препинания – «.», «,», «:», «;», «-», « », «(».

3. Заполнить столбец Код символа используя функцию «КОДСИМВ(...)», находящуюся в категории «Текстовые».

4. Открыв каскадом текст по варианту и таблицу и используя в Word «Правка ⇒ Заменить» заполнить столбец Число вхождений символа в текст. (Предполагается, что других символов в тексте НЕТ.) Сосчитать общее число символов.

5. По формулам заполнить столбцы « $p_i$ » и « $I_i$ ». Сосчитать полную вероятность и энтропию источника.

6. Создать таблицу, аналогичную рис.2 и заполнить ее по формулам.

	Неопределенность	Разрядность кода	Абсолютная избыточность	Относительная избыточность
Стандартная кодовая таблица ASCII				
Мера Хартли				

Рис.2.

7. Выписать применяемые формулы с расшифровкой используемых символов

## Ход выполнения:

### Таблица №1

	кодировка	число вхождений	$p = k/K$	$I = \log_2(1/p) = -\log_2 p$
1	0	48	0,017676768	5,822001698
2	1	49	0,018181818	5,781359714
3	2	50	0,01969697	5,665882496
4	3	51	0,024747475	5,336574871
5	4	52	0,01969697	5,665882496
6	5	53	0,018181818	5,781359714
7	6	54	0,018181818	5,781359714
8	7	55	0,018686869	5,741831349
9	8	56	0,017171717	5,863821874
10	9	57	0,019191919	5,703357202
11 .		46	0,017171717	5,863821874
12 ,		44	0,01969697	5,665882496
13 :		58	0,017171717	5,863821874
14 ;		59	0,024242424	5,366322214
15 -		45	0,01969697	5,665882496
16		32	0,018686869	5,741831349
17 (		40	0,022222222	5,491853096
18 а		224	0,015656566	5,997088405
19 б		225	0,023232323	5,427722759
20 в		226	0,018686869	5,741831349
21 г		227	0,021717172	5,52501996
22 д		228	0,022727273	5,459431619
23 е		229	0,01969697	5,665882496
24 ё		222	0,023232323	5,427722759
25 ж		230	0,019191919	5,703357202
26 з		231	0,019191919	5,703357202
27 и		232	0,01969697	5,665882496
28 й		233	0,022727273	5,459431619
29 к		234	0,020707071	5,59373271
30 л		235	0,017676768	5,822001698

31	м	236	38	0,019191919	5,703357202	0,109458371
32	н	237	44	0,022222222	5,491853096	0,12204118
33	о	238	46	0,023232323	5,427722759	0,12609861
34	п	239	37	0,018686869	5,741831349	0,107296848
35	р	240	38	0,019191919	5,703357202	0,109458371
36	с	241	43	0,021717172	5,52501996	0,119987807
37	т	242	40	0,02020202	5,62935662	0,113724376
38	у	243	47	0,023737374	5,396695863	0,128103387
39	ф	244	36	0,018181818	5,781359714	0,105115631
40	х	245	27	0,013636364	6,196397213	0,084496326
41	ц	246	49	0,024747475	5,336574871	0,132066752
42	ч	247	46	0,023232323	5,427722759	0,12609861
43	ш	248	38	0,019191919	5,703357202	0,109458371
44	щ	249	34	0,017171717	5,863821874	0,100691891
45	ъ	250	32	0,016161616	5,951284715	0,096182379
46	ы	251	40	0,02020202	5,62935662	0,113724376
47	ь	252	48	0,024242424	5,366322214	0,13009266
48	э	253	44	0,022222222	5,491853096	0,12204118
49	ю	254	33	0,016666667	5,906890596	0,098448177
50	я	223	43	0,021717172	5,52501996	0,119987807
всего символов				1 энтропия:		5,631959196
1980						

**Таблица №2**

	Неопределенность	Разрядность кода	Абсолютная избыточность	Относительная избыточность
Стандартная кодовая таблица ASCII	8	8	2,368040804	3,378320164
Мера Хартли	5,64385619	6	0,011896994	504,3290657

## Формулы, примененные в работе:

$$p(i) = k/n$$

Где: 1) k – количество вхождения символа в текст

2) n – Количество символов в тексте

$$I(i) = p(i) \cdot \log_2(1/p(i)) - \text{Значение энтропии отдельного элемента}$$

Где: 1) p(i) – вероятность вхождение символа

2) i – символ

Если сообщение — число, понятие глубины числа будет трансформировано в понятие основания системы счисления. При заданных глубине и длине числа количество чисел, которое можно представить,  $N = A^n$ . Очевидно, что  $N$  однозначно характеризует степень исходной неопределенности. Исходная неопределенность по Хартли определяется

$$H_1 = \log_a N. \quad (4)$$

Неопределенность после получения сообщения, остаточная неопределенность,

$$H_2 = \log_a N^*, \quad (5)$$

где  $N^*$  — число возможных значений принятого слова после получения сообщения.

**Количество информации по Хартли:**

$$I = H_1 - H_2 = \log_a N - \log_a N^* n = \log_a N / N^* .$$

средняя информация, доставляемая одним опытом,

$$I_{cp} = (k_1 I_1 + k_2 I_2 + \dots + k_A I_A) / K.$$

K - кол-во символов алфавита.

A - количество исходов опыта.

k<sub>i</sub> - количество появления символа в тексте.

I<sub>i</sub> - количество вносимой этим символом информации.

## **Вывод:**

В ходе выполнения первой части лабораторной работы были изучены методы определения информации по Шеннону и Хартли, метод Хаффмана. Мера Хартли и неопределенность по Шеннону совпадают, так как все символы в таблице ASCII имеют одинаковую вероятность нахождения. В таком случае результаты будут аналогичными. Абсолютная избыточность равна нулю, так как предполагается передача без потерь.

## **Часть 2.**

### ***«Кодирование дискретных источников информации методом Шеннона-Фано»***

#### ***Порядок выполнения части 2 лабораторной работы***

Исходными данными для данной лабораторной работы являются результаты статистической обработки текста, выполненной в предыдущей лабораторной работе. Из лабораторной работы «Определение количества информации, содержащегося в сообщении» для данной работы необходимо взять: 1) список символов данного текста; 2) оценку вероятностей появления символов в тексте; 3) значение энтропии источника. Расчеты рекомендуется выполнять в табличной форме, используя MS Excel. 1. Отсортировать символы в порядке убывания их вероятности появления в тексте. 2. Построить один из возможных вариантов по правилу Шеннона-Фано для посимвольного кодирования заданного текста. 3. Определить энтропию и среднее количество двоичных разрядов, необходимых для передачи текста при использовании эффективных кодов. 4. Проверить возможность однозначного декодирования полученных кодов, рассмотрев пример передачи слова, состоящего из не менее 10 символов.

## ***Ход выполнения:***

**Таблица для 50-ти символов, содержащая список символов, значения вероятностей, кодовые комбинации, ступени.**

[illegible]

**Значение средней информации в битах = 5,86**

**Применялись формулы подсчета вероятности и энтропии каждого элемента.**

## Закодированное сообщение:

с	д	з	ю	л	а	о	у	
10101	1100000	0000001	000010	0100001	00011	0000001	11010	11011
Номер с	Кол-во разрд.	Символ	КОД					
1	3	-	101					
2	4	-	1010					
3	5	С	10101					
1	3	-	110					
2	4	-	1100					
3	5	-	11000					
4	6	-	110000					
5	7	Д	1100000					
1	3	-	000					
2	4	-	0000					
3	5	-	00000					
4	6	-	000000					
5	7	а	0000001					
1	3	-	000					
2	4	-	0000					
3	5	-	00001					
5	6	ю	000010					
1	3	-	010					
2	4	-	0100					
3	5	-	01000					
4	6	-	010000					
5	7		0100001					
1	3	-	000					
2	4	-	0001					
3	5	л	00011					
1	3	-	1110110					
1	3	-	000					
2	4	-	0000					
3	5	-	00000					
4	6	-	000000					
5	7	а	0000001					
1	3	-	110					
2	4	-	1101					
3	5	б	11010					
1	3	-	110					
2	4	-	1101					
3	5	в	11011					

## Вывод:

За вторую часть работы я воспользовался методом Шеннона-Фано, собрав все двоичные коды для каждого символа. Проверил на корректность и составил таблицу ступеней.

## Часть 3.

### *«Кодирование дискретных источников информации по методике Хаффмана»*

#### *Порядок выполнения части 3 лабораторной работы*

Из лабораторной работы «Кодирование дискретных источников информации методом Шеннона-Фано» необходимо взять вычисленное значение средней информации. Расчеты рекомендуется выполнять в табличной форме, используя MSExcel. 1. Отсортировать символы в порядке убывания их вероятности появления в тексте. 2. Построить таблицу по правилу Д. Хаффмана для посимвольного кодирования заданного текста (См. Табл.3.1). 3. Определить энтропию и среднее количество двоичных разрядов, необходимых для передачи текста при использовании эффективных кодов. 4. Построить кодовое дерево (См.рис.3.1). 5. Создать таблицу кодов. 6. Проверить возможность однозначного декодирования полученных кодов, рассмотрев пример передачи слова, состоящего из не менее 10 символов.

символ	вхождения	вероятность	
3	49	0,024747475	01100
ц	49	0,024747475	01101
;	48	0,024242424	01010
ь	48	0,024242424	01011
у	47	0,023737374	01001
б	46	0,023232323	00101
ё	46	0,023232323	00100
о	46	0,023232323	00101
ч	46	0,023232323	00111
д	45	0,022727273	00010
й	45	0,022727273	00011
(	44	0,022222222	00001
н	44	0,022222222	111111
э	44	0,022222222	00000
г	43	0,021717172	111110
с	43	0,021717172	111100
я	43	0,021717172	111101
к	41	0,020707071	111011
т	40	0,02020202	111001
ы	40	0,02020202	111010
2	39	0,01969697	110011
4	39	0,01969697	110110
,	39	0,01969697	111000
-	39	0,01969697	110101
е	39	0,01969697	110100
и	39	0,01969697	110111
9	38	0,019191919	00010
ж	38	0,019191919	110000
з	38	0,019191919	110001
м	38	0,019191919	110010
р	38	0,019191919	101110
ш	38	0,019191919	101111
7	37	0,018686869	101001
	37	0,018686869	101011
в	37	0,018686869	101010
п	37	0,018686869	101100
1	36	0,018181818	100101
5	36	0,018181818	100111
6	36	0,018181818	100111
ф	36	0,018181818	101000
0	35	0,017676768	100011
л	35	0,017676768	100100
8	34	0,017171717	011111
.	34	0,017171717	100010
:	34	0,017171717	100000
щ	34	0,017171717	100001
ю	33	0,016666667	011110
ъ	32	0,016161616	011101
а	31	0,015656566	011100
х	27	0,013636364	010001

**Зчение средней информации в битах = 5,72**

**Были применены формулы:**

$$p(i) = k/n$$

Где: 1) k – количество вхождения символа в текст

2) n – Количество символов в тексте

$$I(i) = p(i) \cdot \log_2(1/p(i))$$
 – Значение энтропии отдельного элемента

Где: 1) p(i) – вероятность вхождение символа

2) i – символ



Если сообщение — число, понятие глубины числа будет трансформировано в понятие основания системы счисления. При заданных глубине и длине числа количество чисел, которое можно представить,  $N = A^n$ . Очевидно, что  $N$  однозначно характеризует степень исходной неопределенности. Исходная неопределенность по Хартли определяется

$$H_1 = \log_a N. \quad (4)$$

Неопределенность после получения сообщения, остаточная неопределенность,

$$H_2 = \log_a N^*, \quad (5)$$

где  $N^*$  — число возможных значений принятого слова после получения сообщения.

**Количество информации по Хартли:**

$$I = H_1 - H_2 = \log_a N - \log_a N^* = \log_a (N/N^*).$$

средняя информация, доставляемая одним опытом,

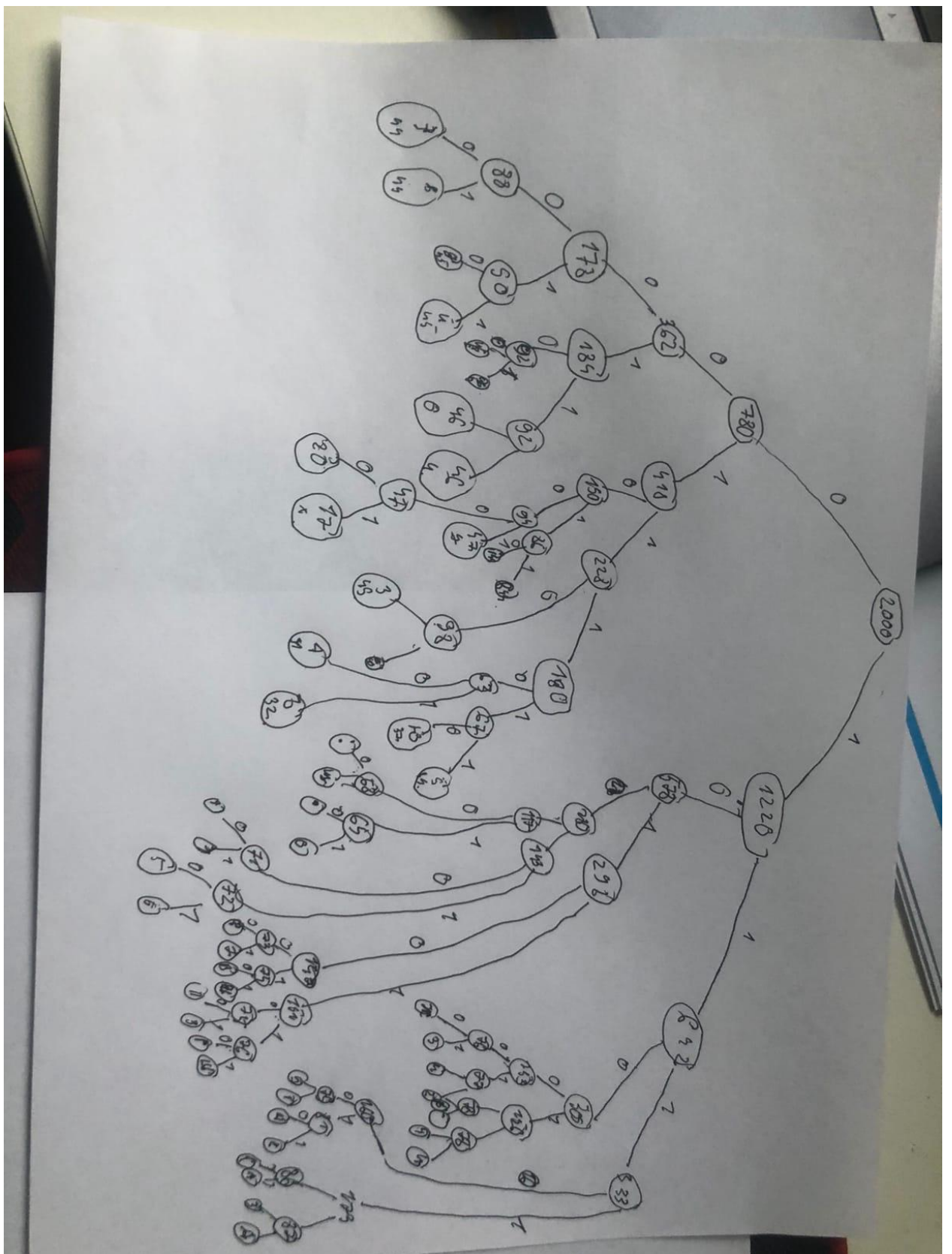
$$I_{cp} = (k_1 I_1 + k_2 I_2 + \dots + k_A I_A) / K.$$

$K$  - кол-во символов алфавита.

$A$  - количество исходов опыта.

$k_i$  - количество появления символа в тексте.

$I_i$  - количество вносимой этим символом информации.



**Закодированное сообщение:**

00111	010100	011001	010101	011001	10001	01101	0100001	00111	010100	110001	01001	11011
п	р	и	м	и	т	е		п	р	о	ш	у
				Номер с	Кол-во разрд.	Символ	КОД					
				1	3	-	001					
				2	4	-	0011					
				3	5	п	00111					
				1	3	-	010					
				2	4	-	0101					
				3	5	-	01010					
				4	6	р	010100					
				1	3	-	011					
				2	4	-	0110					
				3	5	-	01100					
				4	6	и	011001					
				1	3	-	010					
				2	4	-	0101					
				3	5	-	01010					
				4	6	м	010101					
				1	3	-	011					
				2	4	-	0110					
				3	5	-	01100					
				4	6	и	011001					
				1	3	-	100					
				2	4	-	1000					
				3	5	т	10001					
				1	3	-	011					
				2	4	-	0110					
				3	5	е	01101					
				1	3	-	010					
				2	4	-	0100					
				3	5	-	01000					
				4	6	-	010000					
				5	7		0100001					
				1	3	-	001					
				2	4	-	0011					
				3	5	п	00111					
				1	3	-	010					
				2	4	-	0101					
				3	5	-	01010					
				4	6	р	010100					
				1	3	-	110					
				2	4	-	1100					
				3	5	-	11000					
				4	6	о	110001					
	1	3	-	010								
	2	4	-	0100								
	3	5	ш	01001								
	1	3	-	110								
	2	4	-	1101								
	3	5	у	11011								

## Вывод:

Кодировка методом Хаффмана оказалась труднее, но более эффективной. Средняя длина бит стремиться к энтропии с большим успехом. В ходе третьей части я научился строить дерево Хаффмана, проверил наличие условия единственной раскодировки сообщения.

## Текст (lorem)

н918оч3бицнвиэ3ю7иу-дфджбцйэ:(ч9(:щиксит4з8( 3д;е-9дж;зёюл(бтбц7жняу42бпэ-  
зл:ёятфр80ё54ь8д;нпк53л

,ёп8у...п8 рзэцд,8оэ;ыч19пгяцэ5жпыншы7ё9;(а1бул;в7::рагэабс901йбтн9я,щ  
сбщ1ом9ея7,юс77;,(щд(ш4кьнык

55 э 45(к,64ц;шёэ8чирэуз7йи5йцкх3;07дю:йъ;лкочзлзельчнбьбь5йгкггёш.-

цяё,32рщкв3м79япашь3а8ммг3эи:зъч

тс-учэп5чкщ(брчбн0ю5лт3еечъг1ёф;з1ряжткчнсь.ив(ж3мь,:иязэб-

жмоу,йкз4рлц3явгие0еаордюб64тёньфё4рыюю7

ж-6фекьбцуньш9от,йфоть4ы яыягц8ш.жб2вичтшчж4

ёужикй;0ж1йдс3ьжъя30хжэлрпсш1пнийщшгемщъшущсж9пи(п9ир

ф;9уэммбьяф2лнщп-;2с8,ё;аа в (ф332лгагурф62р1.зрбм-

щм71:лэёйрч;хишщпуд1м24тдаё4ю8г0л9 фкъ-лк;дйл:(7ц

гкфх9ньё1в:дллыяу8апщюп0б9ыбщб::т00ёщм8чёбым(

унш4928(у4ха.2ь8д762ё(5звкъёж1(уд7ымйяпк2хех7ж-ьсх67;

лбмхобибузп2ч,чгймг3йя35й-9ч-жкь(йэ,вэяюувр(;;ц црне6006нд7жм36ьо.2ёпвфа386о,.  
в4тсщвш.к1ы79в-.ъэ;;  
оо.:бдъ-вюцю1фрж89л1юу.щгс8,лоркзкд:тёпж(йщу9щ;щмцйн-юярцт-ыд-  
м662р0н;359цж10яиэт2жтрд.;-эмхшя45 ле  
ы;0у;ххюм9м044сэ9и жёмх ьучь4е эщ30вдч(дъхн5ед8:ё46н.м3цв(сявсд:аль-5чз  
ыщнгткноыхяшжэ(хужбт33(фшй.  
феубёж3цбм3елр1нлбтп.обыхх;шщяияшт3мбрбк4б;3а5о-.вмс(9мчэс91:0б.ьн7аеы  
пиощ20н(быщпнвяльзодем1(я(-  
2яу,к;е3с-  
дёт.ж,иёйт:уьисйелфвеб34эищщ;ф2а,60ш;7дтр8юкбияъуь4чкь7з;цс3зк:юцьы,ъц;х5ьй-  
5(о3п3иёа9э,  
сь5;и3ё,й;бцзв4,ч.е;9къче2букёгф(в5 улмешг-ф3чм-ьж1аш44ьцр1ъ.гч286;фс-.пчяфьаш2-  
ству1йчюбафть42ботж  
о:(2а-7м .г-37-7шзбхпбач7эё7д0.у,бъъжвьееёцм вфци,кеач  
:уо41ээйтгчш9ьн8.;гьыгк(,яй:йтпу1ьсё.ядзчбшч  
убтэп9еэк1й8цыб::9щю,;зю:5ол(зо-ррюк0оорю йуо-,еаг2я734фс,.ч0щ2о5очк2хык;5д  
ёл0ьныщ2с;67чгь2лыд3.жг  
фян(яюуфьййцлпслц5цёй9д9:шшйдозисьшьубяфн  
4оыбг,ъ2хй8кс(ёечгё2ов7ш;ш,г;ф8э31ьгб5лгйю:э:йьошщ4,а75 ь  
хдниофц5з7яэ-кыю  
3пггжнн(2уыву03ыяьйзшрмнний0т4фы:0кйлб1цньё.ё(ьшв89опя3б1йвдьб5  
з3ззёп1яецою0ыф,ер2  
(б8ва82иш4сз;теё(ынсгю7с8эцп720н1ё-9(ссобхтщхшб78,йжф5048чнифы9у;вшт :дсщпзог-  
цуёгтщс.(5сжг4ююжщнж5  
,чэрюч5-юуьв в2.асди-  
бэп12ьа7йг.гю.;чрорсшеьй7ьызтл4и5сд1и3э2трёгч,чф:44ягц5дцаыцисцхзёизкйъз бмр.  
пвд 21ёбьюнс3дршэ8оы50ц(8цу0эо;(-э;ецб,дзьо:уи  
зт(ф9т,ов39ьыиехмдёпздмгбцбоызястбч1е0ббьну0жбэзцу