

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное автономное бюджетное образовательное учреждение высшего
образования

Севастопольский государственный университет

кафедра Информационных систем

Институт информационных технологий и управления в технических системах

курс 1 группа ИС/м-11-о

Райков Сергей Сергеевич

09.04.02 Информационные системы и технологии

ОТЧЁТ

о лабораторном практикуме №1
Исчисление количества информации
по дисциплине «Дополнительные разделы теории информации»

Отметка о зачёте _____
(дата)

Руководитель практикума

доц.
(должность)

(подпись)

Заморенов М.В.
(инициалы, фамилия)

Севастополь
2017

1. ЦЕЛЬ РАБОТЫ

Изучить основные методы и средства исчисления количества информации. Научиться применять пакет Matlab для расчета количественных характеристик.

2. ВАРИАНТ ЗАДАНИЯ

Вариантом задания является ФИО студента и его дата рождения, а именно: «Райков Сергей Сергеевич 1994 01 25»

3. РАСЧЕТНАЯ ЧАСТЬ

При выполнении работы необходимо выделить две группы символов: отдельно буквы и отдельно цифры. Это связано с тем, что появление цифр в тексте носит равновероятный характер.

3.1 Вычисление количества информации и энтропии сообщения для случая равновероятных символов алфавита

$$I = \log_2 N, \quad (3.1)$$

В рассматриваемом случае весь алфавит для передачи буквенных сообщений на русском языке представлен 32 символами, а алфавит для передачи численных сообщений – 10 символами.

$$I_{16} = \log_2 32 = 5 \text{ (бит)}.$$

$$I_{10} = \log_2 10 = 3,322 \text{ (бит)}.$$

Для вычисления количества информации, которое приходится на все сообщение, необходимо воспользоваться выражением

$$\begin{aligned} I_6 &= N_6 I_{16}, \\ I_{10} &= N_{10} I_{10}, \end{aligned} \quad (3.2)$$

где N_6 – число букв в сообщении, N_{10} – число цифр в сообщении.

Наше сообщение:

Райков Сергей Сергеевич 1994 01 25

Количество букв в сообщении с учетом пробелов равно 26, количество цифр – 8, следовательно, количество информации буквенной части сообщения равно

$$\begin{aligned} I_6 &= 26 \times 5 = 130 \text{ (бит)} \\ I_{10} &= 8 \times 3,322 = 26,575 \text{ (бит)} \end{aligned}$$

Общее количество информации в сообщении, состоящем из равновероятных букв и цифр, определяется по формуле

$$I = I_{\text{б}} + I_{\text{ц}} = 130 + 26,575 = 156,575 \text{ (бит)} .$$

Для расчета энтропии сообщения следует использовать соотношение

$$H = \frac{I_{\text{б}} + I_{\text{ц}}}{N_{\text{б}} + N_{\text{ц}}} = \frac{156,575}{8 + 26} = 4,6 \left(\frac{\text{бит}}{\text{символ}} \right) .$$

3.2 Вычисление количества информации и энтропии сообщения для случая неравновероятных символов алфавита

Для случая неравновероятных символов алфавита расчеты необходимо начать с составления таблицы вероятностей появления символов алфавита в сообщении.

Символ сообщения	p_i	$p_i \times \log_2 p_i$
Р	0,02	0.11122402631316657
а	0,02	0.22244805262633313
й	0,04	0.40568041897815643
к	0,02	0.516904445291323
о	0,02	0.6281284716044895
в	0,04	0.8113608379563129
	0,1	1.1398409602194632
С	0,04	1.3230733265712866
е	0,1	1.651553448834437
р	0,04	1.8347858151862604
г	0,04	2.0180181815380838
е	0,1	2.346498303801234
й	0,04	2.5297306701530573
	0,1	2.8582107924162075
С	0,04	3.041443158768031
е	0,1	3.369923281031181
р	0,04	3.5531556473830044
г	0,04	3.7363880137348278
е	0,1	4.064868135997978
е	0,1	4.3933482582611285
в	0,04	4.5765806246129515

и	0,02	4.687804650926118
ч	0,02	4.799028677239285
	0,1	5.1275087995024355
	0,1	5.455988921765586
	0,1	5.784469044028737

После составления таблицы необходимо вычислить энтропию сообщения, которая приходится на его буквенную часть. Для этого необходимо воспользоваться формулой

$$H_{\text{б}} = - \sum_{i=1}^{N_{\text{б}}} p_i \times \log_2(p_i) . \quad (3.3)$$

Выполнив расчеты получим $H_{\text{б}} = 5,78 \left(\frac{\text{бит}}{\text{символ}} \right)$.

Далее, можно вычислить количество информации, содержащейся в буквенной части сообщения по формуле

$$I_{\text{б}} = N_{\text{б}} \times H_{\text{б}} . \quad (3.4)$$

Подставляя в (3.4) численные значения параметров, получим

$$I_{\text{б}} = 26 \times 5,78 = 150,39 \text{ (бит)} .$$

Общее количество информации для буквенно-цифрового сообщения рассчитывается из соотношения $I = I_{\text{б}} + I_{\text{ц}} = 150,39 + 20,73 = 171,13 \text{ (бит)}$. Отметим, что количество информации для буквенной части сообщения определяется точно также, как и в подразделе 3.1.

Общая энтропия сообщения рассчитывается аналогично

$$H = \frac{I_{\text{б}} + I_{\text{ц}}}{N_{\text{б}} + N_{\text{ц}}} = \frac{171,13}{26 + 8} = 5,03 \left(\frac{\text{бит}}{\text{символ}} \right) .$$

3.3 Вычисление количества информации и энтропии сообщения для случая дву- и трехбуквенных сочетаний

Для русского алфавита, с учетом пробела и двубуквенных сочетаний

$$H_2 = 3,52 \left(\frac{\text{бит}}{\text{символ}} \right) ; \text{ с учетом трехбуквенных сочетаний } H_3 = 3,01 \left(\frac{\text{бит}}{\text{символ}} \right) .$$

Исходя из этого можно записать выражение для количества информации

$$I_{2x} = H_2 N_6 + I_y = 112.25 \text{ (бит)} ,$$

$$I_{3x} = H_3 N_6 + I_y = 98.99 \text{ (бит)} .$$

Энтропию сообщений каждого вида вычислим по уже знакомой формуле

$$H_{2x} = \frac{I_{2x}}{N_6 + N_y} = 3,3 \left(\frac{\text{бит}}{\text{символ}} \right) ,$$

$$H_{3x} = \frac{I_{3x}}{N_6 + N_y} = 2,9 \left(\frac{\text{бит}}{\text{символ}} \right) .$$

ВЫВОДЫ

В ходе выполнения данной лабораторной работы для заданного исходного буквенно-цифрового сообщения были вычислены такие характеристики, как количество информации отдельно для букв и цифр и общее, энтропия.

Данные количественные характеристики были вычислены как для равновероятных, так и для неравновероятных символов алфавита, а также для двухбуквенных и трехбуквенных сочетаний.

Таким образом, были изучены основные методы и средства исчисления количества информации, для чего был применен пакет Matlab.

ПРИЛОЖЕНИЕ А

Текст программы

```
$ = console.log

var fname = "Райков Сергей Сергеевич 1994 01 25";

function getLetters(str) {

    return fname.split("").filter(e => isNaN(parseFloat(e)));

}

function getDigits(str) {

    return str.split("").filter(e => !isNaN(parseFloat(e)));

}

function hartley(N) {
    return Math.log2(N);
}

var Ione_ch = hartley(32);
var Ione_dig = hartley(10);

var digits = getDigits(fname);
var chars = getLetters(fname);

var digits_count = digits.length;
var letters_count = chars.length;

$("Количество цифр - " + digits_count, "Кол-во букв " +
letters_count);

var Ich = letters_count * Ione_ch;
var Idig = digits_count * Ione_dig;

$("Количество информации (буквы) " + Ich + " бит", "Количество
информации (цифры) " + Idig + " бит");

$("Общее кол-во информации ", Ich + Idig + " бит");

$("Энтропия сообщения ", (Ich + Idig) / (digits_count +
letters_count));

$("Неравновероятный алфавит");

function frequencies(chars) {
```

```

var str = chars.join();

var result = [];

chars.forEach((ch) => {

    var matches = str.match(new RegExp(ch, "g")).length;

    result.push([ch, matches/str.length]);

});

return result;

}

// ДЛЯ БУКВ

var char_freq = frequencies(getLetters(fname));

$(char_freq);

var Hch = 0;

for(var i = 0; i < char_freq.length; i++) {
    Hch += char_freq[i][1]*Math.log2(char_freq[i][1]);

    $(-Hch);
}

Hch = -Hch;

$("Энтропия буквенной части сообщения = ", Hch, " бит/символ");

Ich_p = letters_count * Hch;

$("Количество информации буквенной части сообщения ", Ich_p, "
бит");

// ДЛЯ ЦИФР

var digits_freq = frequencies(getDigits(fname));

$(digits_freq);

var Hdig = 0;

for(var i = 0; i < digits_freq.length; i++) {
    Hdig += digits_freq[i][1]*Math.log2(digits_freq[i][1]);
}

```

```

Hdig = -Hdig;

$("Энтропия цифровой части сообщения = ", Hdig, " бит/символ");

Idig_p = digits_count * Hdig;

$("Количество информации цифровой части сообщения ", Idig_p, "
бит");

$("Общее количество информации для буквенно-цифрового сообщения
", Ich_p+Idig_p, " бит");

$("Общая энтропия для буквенно-цифрового сообщения сообщения ",
(Ich_p+Idig_p) / (digits_count + letters_count), " бит/символ");

// Двухбуквенные и трехбуквенные сочетания
var H2 = 3.52;
var H3 = 3.01;

var I2x = H2*letters_count + Idig_p;
var I3x = H3*letters_count + Idig_p;

$("Количество информации для двухбуквенных сочетаний ", I2x, "
бит");
$("Количество информации для трехбуквенных сочетаний ", I3x, "
бит");

$("Энтропия для двухбуквенных сочетаний ", I2x / (letters_count
+ digits_count) , "бит/символ");
$("Энтропия для трехбуквенных сочетаний ", I3x / (letters_count
+ digits_count) , "бит/символ");

```