ДЕПАРТАМЕНТ ОБРАЗОВАНИЯ ИВАНОВСКОЙ ОБЛАСТИ
ОБЛАСТНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ПРОФЕССИОНАЛЬНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
«КИНЕШЕМСКИЙ ТЕХНОЛОГИЧЕСКИЙ КОЛЛЕДЖ»

УТВЕРЖДАЮ
Директор ОГБПОУ КТК
_________ /И.А. Смирнов/
«31» августа 2022г.

Фонд оценочных средств
по учебной дисциплине
ОП.12 Основы теории информации
по специальности среднего профессионального образования
программа подготовки специалистов среднего звена
технологического профиля
09.02.06 Сетевое и системное администрирование

Срок обучения 3 года 10 месяцев

Кинешма, 2022

Фонд оценочных средств по учебной дисциплине ОП.12 Основы теории информации разработан в соответствии с Федеральным государственным образовательным стандартом среднего профессионального образования по специальности 09.02.06 Сетевое и системное администрирование.

Разработчик: Шаронов Максим Олегович – преподаватель ОГБПОУ «Кинешемский технологический колледж»

Фонд оценочных средств по учебной дисциплине ОП.12 Основы теории информации рассмотрен и одобрен на заседании методической комиссии учебно-методического объединения по укрупненным группам специальностей 09.00.00 Информатика и вычислительная техника, 13.00.00 Электро - и теплоэнергетика, 15.00.00 Машиностроение, 18.00.00 Химические технологии

Протокол № 1 от «31» августа 2022г.

Председатель ___________ Киселева Е.В.

Паспорт

фонда оценочных средств по дисциплине *Основы теории информации*

Контролируемые разделы (темы) дисциплины*	Код формируемой компетенции	Результат освоения (умения и знания)		Оценочные средства
		уметь	знать	
Раздел 1. Базовые понятия теории информации.	ОК-01, ОК-02, ОК-04, ОК-05, ОК-09, ОК-10, ПК 1.3	✓ Использовать формулу Шеннона.	✓ Виды и формы представления информации. ✓ Методы и средства определения количества информации. ✓ Способы передачи цифровой информации.	Устный опрос, практическая работа № 1, практическая работа № 2, практическая работа № 3, самостоятельная работа студента, контрольная работа
Раздел 2. Информация и энтропия.	ОК-01, ОК-02, ОК-04, ОК-05, ОК-09, ОК-10, ПК 1.3	✓ Применять теорему Котельникова. ✓ Применять закон аддитивности информации.	✓ Виды и формы представления информации. ✓ Методы и средства определения количества информации.	Устный опрос, Практическая работа № 4, Практическая работа № 5, Практическая работа № 6, самостоятельная работа студента
Раздел 3. Защита и передача информации.	ОК-01, ОК-02, ОК-04, ОК-05, ОК-09, ОК-10, ПК 1.3	✓ Применять закон аддитивности информации. ✓ Применять теорему Котельникова. ✓ Использовать формулу Шеннона.	✓ Принципы кодирования и декодирования информации. ✓ Методы повышения помехозащищенности передачи и приема данных, основы теории сжатия данных.	Устный опрос, Практическая работа № 7 Практическая работа № 8, Практическая работа № 9, Практическая работа № 10, Практическая работа № 11, Практическая работа № 12, Самостоятельная работа студента, контрольная работа

Раздел 4. Основы теории защиты информации.	ОК-01, ОК-02, ОК-04, ОК-05, ОК-09, ОК-10, ПК 1.3	✓ Применять закон аддитивности информации. ✓ Применять теорему Котельникова. ✓ Использовать формулу Шеннона.	✓ Методы криптографической защиты информации. ✓ Способы генерации ключей.	Устный опрос, Практическая работа № 13, Практическая работа № 14, Практическая работа № 15, Практическая работа № 16, Самостоятельная работа студента, контрольная работа

Приложение 1

ДЕПАРТАМЕНТ ОБРАЗОВАНИЯ ИВАНОВСКОЙ ОБЛАСТИ

ОБЛАСТНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ПРОФЕССИОНАЛЬНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ

«КИНЕШЕМСКИЙ ТЕХНОЛОГИЧЕСКИЙ КОЛЛЕДЖ»

Вопросы к экзамену

Учебная дисциплина ОП. 12 Основы теории информации

специальность 09.02.06 Сетевое и системное администрирование

Раздел 1. Базовые понятия теории информации

Тема 1.1. Формальное представление знаний. Виды информации.

1 Теория информации – дочерняя наука кибернетики. Информация, канал связи, шум, кодирование. Принципы хранения, измерения, обработки и передачи информации.

2 Информация в материальном мире, информация в живой природе, информация в человеческом обществе, информация в науке, классификация информации.

Тема 1.2. Способы измерения информации.

3 Измерение количества информации, единицы измерения информации, носитель информации.

4 Передача информации, скорость передачи информации.

5 Способы хранения обработки и передачи информации.

Тема 1.3. Вероятностный подход к измерению информации.

6 Вероятностный подход к измерению дискретной и непрерывной информации Клода Шеннона.

7 Теория вероятности, функция распределения, дисперсия случайной величины

8 Измерение количества информации.

9 Определение пропускной способности канала.

Раздел 2. Информация и энтропия

Тема 2.1. Теорема отсчетов

10 Теорема отсчетов Котельникова и Найквиста — Шеннона, математическая модель системы передачи информации.

11 Интерполяционная формула Уиттекера-Шеннона, частота Найквиста.

Тема 2.2 Понятие энтропии. Виды энтропии

12 Понятие энтропии. Формула Хартли. Виды условной энтропии, энтропия объединения двух источников.

13 Ƅ-нарная энтропия, взаимная энтропия.

14 Поиск энтропии случайных величин.

15 Энтропийное кодирование.

Тема 2.3. Смысл энтропии Шеннона

16 Статистический подход к измерению информации.

17 Закон аддитивности информации. Формула Шеннона.

18 Дифференциальная энтропия.

Раздел 3. Защиты и передача информации

Тема 3.1. Сжатие информации

19 Простейшие алгоритмы сжатия информации, методы Лемпела-Зива. Особенности программ архиваторов.

20 Применение алгоритмов кодирования в архиваторах для обеспечения продуктивной работы в WINDOWS.

21 Практическое применение различных алгоритмов сжатия.

22 Сравнение и анализ архиваторов.

Тема 3.2. Кодирование

23 Помехоустойчивое кодирование. Адаптивное арифметическое кодирование.

24 Цифровое кодирование, аналоговое кодирование, таблично-символьное кодирование.

25 Числовое кодирование, дельта-кодирование.

26 Адаптивное арифметическое кодирование.

27 Дельта-кодирование.

28 Цифровое кодирование и аналоговое кодирование.

29 Таблично-символьное кодирование.

Раздел 4. Основы теории защиты информации

Тема 4.1. Стандарты шифрования данных. Криптография.

30 Понятие криптографии, использование ее на практике.

31 Различные методы криптографии, их свойства и методы шифрования.

32 Шифрование с использованием перестановок.

33 Шифрование с использованием замен.

34 Криптография с симметричным ключом, с открытым ключом.

Приложение 2

ДЕПАРТАМЕНТ ОБРАЗОВАНИЯ ИВАНОВСКОЙ ОБЛАСТИ

ОБЛАСТНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ПРОФЕССИОНАЛЬНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ

«КИНЕШЕМСКИЙ ТЕХНОЛОГИЧЕСКИЙ КОЛЛЕДЖ»

Билеты к экзамену

Экзаменационный билет № 1

1. Теория информации – дочерняя наука кибернетики. Информация, канал связи, шум, кодирование.
2. Принципы хранения, измерения, обработки и передачи информации.
3. Задача.

Преподаватель ___________М.О. Шаронов

Экзаменационный билет № 2

1. Информация в материальном мире, информация в живой природе, информация в человеческом обществе, информация в науке, классификация информации.
2. Измерение количества информации, единицы измерения информации, носитель информации.
3. Задача.

Преподаватель ___________М.О. Шаронов

Экзаменационный билет № 3

1 Передача информации, скорость передачи информации.
2 Способы хранения обработки и передачи информации.
3 Задача.

Преподаватель ___________М.О. Шаронов

Экзаменационный билет № 4

1 Вероятностный подход к измерению дискретной и непрерывной информации Клода Шеннона.
2 Способы хранения обработки и передачи информации.
3 Задача.

Преподаватель ___________М.О. Шаронов

Экзаменационный билет № 5

1 Теория вероятности, функция распределения, дисперсия случайной величины.
2 Измерение количества информации.
3 Задача.

Преподаватель ___________М.О. Шаронов

Экзаменационный билет № 6

1 Теорема отсчетов Котельникова и Найквиста — Шеннона, математическая модель системы передачи информации.
2 Интерполяционная формула Уиттекера-Шеннона, частота Найквиста.
3 Задача.

Преподаватель ___________М.О. Шаронов

Экзаменационный билет № 7

1 Понятие энтропии. Формула Хартли. Интерполяционная формула Уиттекера-Шеннона, частота Найквиста.
2 Виды условной энтропии, энтропия объединения двух источников.
3 Задача.

Преподаватель ___________М.О. Шаронов

Экзаменационный билет № 8

1 b-нарная энтропия, взаимная энтропия.
2 Поиск энтропии случайных величин.
3 Задача.

Преподаватель ___________М.О. Шаронов

Экзаменационный билет № 9

1 Энтропийное кодирование.
2 Статистический подход к измерению информации.
3 Задача.

Преподаватель ___________М.О. Шаронов

Экзаменационный билет № 10

1 Закон аддитивности информации. Формула Шеннона.
2 Статистический подход к измерению информации.
3 Задача.

Преподаватель ___________М.О. Шаронов

Экзаменационный билет № 11

1 Простейшие алгоритмы сжатия информации, методы Лемпела-Зива.
2 Особенности программ архиваторов.
3 Задача.

Преподаватель ___________М.О. Шаронов

Экзаменационный билет № 12

1 Применение алгоритмов кодирования в архиваторах для обеспечения продуктивной работы в WINDOWS.
2 Практическое применение различных алгоритмов сжатия.
3 Задача.

Преподаватель ___________М.О. Шаронов

Экзаменационный билет № 13

1 Сравнение и анализ архиваторов.
2 Помехоустойчивое кодирование.
3 Задача.

Преподаватель ___________М.О. Шаронов

Экзаменационный билет № 14

1 Адаптивное арифметическое кодирование.
2 Цифровое кодирование, аналоговое кодирование, таблично-символьное кодирование.
3 Задача.

Преподаватель ___________М.О. Шаронов

Экзаменационный билет № 15

1 Числовое кодирование, дельта-кодирование.
2 Адаптивное арифметическое кодирование.
3 Задача.

Преподаватель ___________М.О. Шаронов

Экзаменационный билет № 16

1 Дельта-кодирование.
2 Цифровое кодирование и аналоговое кодирование.
3 Задача.

Преподаватель ___________М.О. Шаронов

Экзаменационный билет № 17

1 Таблично-символьное кодирование.
2 Понятие криптографии, использование ее на практике.
3 Задача.

Преподаватель ___________М.О. Шаронов

Экзаменационный билет № 18

1 Различные методы криптографии, их свойства и методы шифрования.
2 Шифрование с использованием перестановок.
3 Задача.

Преподаватель ___________М.О. Шаронов

Экзаменационный билет № 19

1 Шифрование с использованием замен.
2 Криптография с симметричным ключом, с открытым ключом.
3 Задача.

Преподаватель ___________М.О. Шаронов

Экзаменационный билет № 20

1 Различные методы криптографии, их свойства и методы шифрования.
2 Теория информации – дочерняя наука кибернетики. Информация, канал связи, шум, кодирование. Принципы хранения, измерения, обработки и передачи информации.
3 Задача.

Преподаватель ___________М.О. Шаронов

Экзаменационный билет № 21

1 Информация в материальном мире, информация в живой природе, информация в человеческом обществе, информация в науке, классификация информации.
2 Шифрование с использованием перестановок.
3 Задача.

Преподаватель ___________М.О. Шаронов

Экзаменационный билет № 22

1 Шифрование с использованием замен.
2 Числовое кодирование, дельта-кодирование.
3 Задача.

Преподаватель ___________М.О. Шаронов

Экзаменационный билет № 23

1 Помехоустойчивое кодирование. Адаптивное арифметическое кодирование.
2 Адаптивное арифметическое кодирование.
3 Задача.

Преподаватель ___________М.О. Шаронов

Экзаменационный билет № 24

1 Статистический подход к измерению информации.
2 Дифференциальная энтропия.
3 Задача.

Преподаватель ___________М.О. Шаронов

Приложение 3

ДЕПАРТАМЕНТ ОБРАЗОВАНИЯ ИВАНОВСКОЙ ОБЛАСТИ

ОБЛАСТНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ПРОФЕССИОНАЛЬНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ

«КИНЕШЕМСКИЙ ТЕХНОЛОГИЧЕСКИЙ КОЛЛЕДЖ»

Задачик экзамену

Задача 1. Перевести десятичное число $A = 113$ в двоичную систему счисления, используя таблицу эквивалентов цифр и степеней основания

($q_2 = 2$).

Таблица 1 – Таблица эквивалентов

Десятичное число	Двоичное число
10^0	0001
10^1	1010
10^2	110 0100

Решение. Подставив значения двоичных эквивалентов десятичных цифр и степеней основания в (3), получим

$A = 113 = 1 \cdot 10^2 + 1 \cdot 10^1 + 3 \cdot 10^0 = 001 \cdot 1100100 + 0001 \cdot 1010 + 0011 \cdot 0001 = 1110001_2$.

Ответ: 1110001_2.

Задача 2

Представить в форматах H и F числа -127_{10} и 127_{10}

$127_{10}=1*2^6+1*2^5+1*2^4+1*2^3+1*2^2+1*2^1+1*2^0=0111\ 1111_2$.

$A_2^H=0000\ 0000\ 0111\ 1111_2$, $A_2^F=0000\ 00\ 7\ F_{16}$.

$-127_{10}=-(1*2^6+1*2^5+1*2^4+1*2^3+1*2^2+1*2^1+1*2^0)= -0111\ 1111_2$.

$A_2^H=1000\ 0000\ 0111\ 1111_2$, $A_2^F=8000\ 00\ 7\ F_{16}$.

Задача 3

Определить, какие из следующих шестнадцатиричных чисел положительные, а какие отрицательные: 9754, 157, ADF, 7654AD и DFEA.

Знак числа определяется по первой цифре: если она меньше 8 (1000), то число положительное, если значение от 8 до F, то отрицательное. Таким образом, получаем 9754<0, 157>0, ADF<0, 7654AD>0 и DFEA<0.

Задача 5

Представить в нормальной сетке Е числа $32001{,}5_{10}$ *и* $-32001{,}5_{10}$

Представим числа в шестнадцатиричном коде $32001{,}5_{10}=7D01{,}8_{16}$ и

$-32001{,}5_{10}.=-7D01{,}8_{16}$

Затем найдем нормализованные мантиссы и характеристики.

$m=7D01{,}8_{16} \rightarrow m=0{,}7D018,$

при этом характеристика становится равной Px=40+4=44

Знак m	Px	m	
0	100 0100	0111 1101 0000 0001 1000 0000	447*D0180>0*

$m=-7D01{,}816 \rightarrow m=-0{,}7D018,$

при этом характеристика становится равной Px=40+4=44

Знак m	Px	m	
1	100 0100	0111 1101 0000 0001 1000 0000	C47*D0180<0*

Задача 6

Определить количество информации, получаемое при реализации одного из событий, если бросают

а) несимметричную четырехгранную пирамидку;

б) симметричную и однородную четырехгранную пирамидку.

Решение.

а) Будем бросать несимметричную четырехгранную пирамидку.

Вероятность отдельных событий будет такова:

p1 = 1 / 2,

p2 = 1 / 4,

p3 = 1 / 8,

p4 = 1 / 8,

тогда количество информации, получаемой после реализации одного из этих событий, рассчитывается по формуле:

I = -(1 / 2 log2 1/2 + 1 / 4 log2 1/4 + 1 / 8 log2 1/8 + 1 / 8 log2 1/8) = 1 / 2 + 2 / 4 + + 3 / 8 + 3 / 8 = 14/8 = 1,75 (бит).

б) Теперь рассчитаем количество информации, которое получится при бросании симметричной и однородной четырехгранной пирамидки:

I = log2 4 = 2 (бит).

Задача 7

Сколько различных чисел можно закодировать с помощью 8 бит?

Решение: I=8 бит, K=2I=28=256 различных чисел.

Задача 8

Получить внутреннее представление целого числа 1607 в 2-х байтовой ячейке.

Решение:

$N=1607=11001000111_2$.

Внутреннее представление этого числа будет: 0000 0110 0100 0111. Шестнадцатеричная форма внутреннего представления числа: 0647.

Задача 9

Записать внутреннее представление числа 250,1875 в форме с плавающей точкой.

Решение:

1) Приведем его в двоичную систему счисления с 24 значащими цифрами: $250.1875_{10}=11111010,\ 0011000000000000000_2$.

2) Запишем в форме нормализованного двоичного числа с плавающей точкой: $0,111110100011000000000000*10_2^{1000}$. Здесь мантисса, основание системы счисления ($2_{10}=10_2$) и порядок ($8_{10}=1000_2$) записаны в двоичной системе.

3) Вычислим машинный порядок в двоичной системе счисления: Mp_2= 1000 + 100 0000 =100 1000.

4) Запишем представление числа в 4-х байтовой ячейке памяти с учетом знака числа:

0	1001000	11111010	00110000	00000000

31	24	23		0		

Шестнадцатеричная форма: 48FA3000.

Задача 10

По шестнадцатеричной форме внутреннего представления числа в форме с плавающей точкой C9811000 восстановить само число.

Решение: 1) Перейдем к двоичному представлению числа в 4-х байтовой ячейке, заменив каждую шестнадцатеричную цифру 4-мя двоичными цифрами:

1100 1001 1000 0001 0001 0000 0000 0000

1	1001001	10000001		00010000	00000000	
31		23		0		

2) Заметим, что получен код отрицательного числа, поскольку в старшем разряде с номером 31 записана 1. Получим порядок числа: $p=1001001_2 - 1000000_2 = 1001_2 = 9_{10}$.

3) Запишем в форме нормализованного двоичного числа с плавающей точкой с учетом знака числа:

$-0{,}100000010001000000000000 * 2^{1001}$

4) Число в двоичной системе счисления имеет вид: -100000010.001_2.

5) Переведем число в десятичную систему счисления:

$-100000010.001_2 = -(1*2^8+1*2^1+1*2^{-3}) = -258.125_{10}$

Задача 11

В последовательности из 6 двоичных символов имеется 3 единицы. При передаче данной последовательности сохраняется 3 символа, остальные теряются. Какова вероятность того, что среди сохранившихся будет не более 2 –х единиц?

Решение.

Пусть А – событие, состоящее в том, что среди двоичных символов будет не более 2-х единиц, т.е. 2 или 1, или ни одной. Тогда вероятность события А определяется как сумма:

$$P(A)=P(X\leq 2)=P(X=2)+P(X=1)+P(X=0)$$

Вероятность каждого слагаемого можно рассчитать, используя гипергеометрическое распределение дискретной случайной величины

$$P(X=m)=\frac{C_M^m C_{N-M}^{n-m}}{C_N^n}$$

Общее число возможных комбинаций выбора символов равно числу сочетаний 3 из 6, т.е. C_6^3.

Число благоприятных исходов для X=2 определяется как произведение $C_3^2 C_3^1$, где первый сомножитель это число комбинаций выбора 2-х «единиц» из общего числа «единиц» в последовательности. Но с каждой такой комбинацией могут встретиться символы, не являющиеся «единицами». Число таких комбинаций будет C_3^1. Поэтому искомая вероятность запишется в виде

$$P(X=2)=\frac{C_3^2 C_3^1}{C_6^3}=0{,}45$$

Аналогично для $P(X=1)=\frac{C_3^1 C_3^2}{C_6^3}=0{,}45$ и $P(X=0)=\frac{C_3^0 C_3^3}{C_6^3}=0{,}05$

Таким образом, $P(A)=0{,}95$.

Задача 12

По каналу связи с помехами передается одна из двух команд управления в виде 11111 и 00000, вероятности передачи этих команд соответственно равны 0,7 и 0,3. Вероятность правильного приема каждого из символов 0 и 1 равна 0,6. Символы искажаются помехами независимо друг от друга. На выходе канала имеем кодовую комбинацию 10110. Определить какая комбинация была передана.

Решение.

Пусть событие А состоит в приеме комбинации 10110. Это событие может произойти в совокупности с событиями B_1 (передавалась комбинация 11111) и B_2 (передавалась комбинация 00000). При этом $P(B_1)=0{,}7$, $P(B_2)=0{,}3$.

Условная вероятность приема комбинации 10110 при условии, что передавалась команда 11111 равна

$$P(A/B_1)=P(1/1)\cdot P(0/1)\cdot P(1/1)\cdot P(1/1)\cdot P(0/1),$$

где P(1/1)=0,6, P(0/1)=1- P(1/1)=0,4

$$P(A/B_1)=0{,}6\cdot 0{,}4\cdot 0{,}6\cdot 0{,}6\cdot 0{,}4=0{,}035.$$

Условная вероятность приема комбинации 10110 при условии, что передавалась команда 00000 равна

$$P(A/B_2)=P(1/0)\cdot P(0/0)\cdot P(1/0)\cdot P(1/0)\cdot P(0/0),$$

где P(0/0)=0,6, P(1/0)=1- P(0/0)=0,4

$$P(A/B_2)=0{,}4\cdot 0{,}6\cdot 0{,}4\cdot 0{,}4\cdot 0{,}6=0{,}023.$$

По формуле полной вероятности $P(A)=P(B_1)P(A/B_1)+P(B_2)P(A/B_2)=$

$=0{,}7\cdot 0{,}035+0{,}3\cdot 0{,}023=0{,}0314$

По формуле Байеса $P(B_1/A)=\dfrac{P(B_1)P(A/B_1)}{P(A)}=\dfrac{0{,}7\cdot 0{,}035}{0{,}0314}=0{,}78,$

$$P(B_2/A)=\frac{P(B_2)P(A/B_2)}{P(A)}=\frac{0{,}3\cdot 0{,}023}{0{,}0314}=0{,}22.$$

Сравнивая найденные результаты, заключаем, что более вероятна передача команды 11111.

Приложение 4

ДЕПАРТАМЕНТ ОБРАЗОВАНИЯ ИВАНОВСКОЙ ОБЛАСТИ

ОБЛАСТНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ПРОФЕССИОНАЛЬНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ

«КИНЕШЕМСКИЙ ТЕХНОЛОГИЧЕСКИЙ КОЛЛЕДЖ»

Практические занятия

Практическое занятие «Способы измерения информации»

Цель работы: познакомиться с основными подходами к измерению информации и использовать их при решении задач.

Теоретические сведения к практической работе

Единицы измерения количества информации

Минимальную порцию информации о каком-либо свойстве объекта принято называть битом (binary digit – двоичная цифра). **Бит** – единица измерения информации, представляющая собой выбор из двух равновозможных вариантов. Бит представляет собой обозначение одного двоичного разряда, способного, в зависимости от сделанного выбора, принимать значение 1 или 0.

Таблица степеней двойки показывает, сколько комбинаций можно закодировать с помощью некоторого количества бит:

Количество бит	1	2	3	4	5	6	7	8	9	10
Количество комбинаций	2	4	8	16	32	64	128	256	512	1024

Байт – единица измерения информации, представляющая собой последовательность, состоящую из 8 бит: **1 байт = 23 бит = 8 бит.**

Каждый бит имеет определенное место внутри байта, которое называется **разрядом**. Разряды принято нумеровать справа налево. Например, третий бит в байте на самом деле находится в пятом разряде байта.

Для измерения больших объемов информации принято использовать производные единицы измерения, представленные в таблице:

Название	**Степень**	**Условное обозначение**
Килобайт	210 (1024 байт)	Кбайт, KB
Мегабайт	220 (1024 Кбайт)	Мбайт, MB
Гигабайт	230(1024 Мбайт)	Гбайт, GB
Терабайт	240 (1024 Гбайт)	Тбайт, TB
Петабайт	250 (1024 Тбайт)	Пбайт, PB
Эксабайт	260 (1024 Пбайт)	Эбайт, EB
Зеттабайт	270 (1024 Эбайт)	Збайт, ZB
Йоттабайт	280 (1024 Збайт)	Йбайт, YB

Содержательный подход к измерению количества информации

Новые сведения о свойствах объектов окружающего нас мира содержат информацию для человека и, следовательно, пополняют его знания. При содержательном подходе возможна качественная оценка полученной информации, например, насколько она для нас полезна, важна или наоборот – вредна.

Неопределенность знания о некотором событии – это количество возможных результатов события (бросания монеты, кубика; вытаскивания жребия и пр.). Уменьшение неопределенности знания человека в 2 раза, несет для него **1 бит** информации.

Количество информации *(I)* для событий с различными вероятностями определяется по формуле *К.Шеннона*:

$$\mathbf{I=\sum_{i=1}^{N} p_i \log 2 p_i}$$

где *N* – количество возможных событий; – вероятности отдельных событий. p_i

Заметим, что сумма вероятностей равна 1.

Если события равновероятны, то количество информации *(I)* определяется по формуле *Р.Хартли*:

$\mathbf{I = \log 2\ N}$ или $\mathbf{2^I = N}$

где *N* – количество равновероятных событий.

Алфавитный подход к измерению количества информации

Алфавит – множество символов, используемых при записи текста. Полное количество символов в алфавите называется **размером** (*мощностью*) алфавита.

Алфавитный подход позволяет определить количество информации в тексте. Данный подход является **объективным**, т.е. он не зависит от человека, воспринимающего текст.

Если допустить, что все символы алфавита встречаются в тексте с одинаковой частотой (равновероятно), то **мощность *(N)*** алфавита вычисляется по формуле:

$N=2^i$

где *i* – информационный вес одного символа в используемом алфавите.

Если весь текст состоит из *К* символов, то при алфавитном подходе размер содержащейся в нем информации равен:

I=K* i

Задания для самостоятельного решения

1. Расположите величины в порядке убывания: 1024 бита, 1000 байтов, 1 бит, 1 байт, 1 Кбайт.

2. Переведите:
2,5 байта = _____ бит
20 Кб = ________ байт
2048 байт = ______ Кб
2560 Кбайт = _____ Мб

3. Сравните (поставьте знак отношения):
3 байта ____ 24 бита;
1536 битов _____ 1,5Килобайта;
8192 байта _____ 1 Кбайт.

4. Заполните пропуски (степени двойки).

1 байт	2^3 битов					
1 Кбайт	2— битов	2^{10} байтов				
1 Мбайт	2— битов	2— байтов	2^{10} Кбайт			
1 Гбайт	2— битов	2— байтов	2— Кбайт	2^{10} Мбайт		
1 Тбайт	2— битов	2— байтов	2— Кбайт	2—Мбайт	2^{10} Гбайт	
1 Пбайт	2— битов	2— байтов	2— Кбайт	2—Мбайт	2— Гбайт	2^{10} Тбайт

5. *Найдите x.
1) 8^x битов = 32 Кбайт   2) 16^x битов = 128 Кбайт

6. *Найдите *x и y*
512 Кбайт = 2^x байт = 2^y бит.

7. *Сколько Кбайт информации содержат сообщения следующего объёма:
1) 216 битов ____
2) 216 байтов ____
3) 1/4 Мбайт ____

8. Сделайте вывод по работе.

Практическое занятие
«Вероятностный подход к измерению дискретной и непрерывной информации»

Цель: научиться вычислять вероятности событий (появление символов в сообщении) и рассчитывать энтропию.
Время выполнения: 2 часа
Оборудование: ПК.
Программное обеспечение: операционная система, калькулятор, текстовый редактор.

Теоретические основы

Количество информации по Хартли и Шеннону
Понятие количество информации отождествляется с понятием информация. Эти два понятия являются синонимами. Мера информации должна монотонно возрастать с увеличением длительности сообщения (сигнала), которую естественно измерять числом символов в дискретном сообщении и временем передачи в непрерывном случае. Кроме того, на содержание количества информации должны влиять и статистические характеристики, так как сигнал должен рассматриваться как случайный процесс.
При этом наложено ряд ограничений:
1. Рассматриваются только дискретные сообщения.
2. Множество различных сообщений конечно.
3. Символы, составляющие сообщения равновероятны и независимы.
Хартли впервые предложил в качестве меры количества информации принять логарифм числа возможных последовательностей символов.
$I=\log m^k=\log N$ (1)
К.Шеннон попытался снять те ограничения, которые наложил Хартли. На самом деле в рассмотренном выше случае равной вероятности и независимости символов при любом k все возможные сообщения оказываются также равновероятными, вероятность каждого из таких сообщений равна P=1/N. Тогда количество информации можно выразить через вероятности появления сообщений $I=-\log P$.
В силу статистической независимости символов, вероятность сообщения длиной в k символов равна

$$P=\prod_{i=1}^{k} p_i$$

Если i–й символ повторяется в данном сообщении k_i раз, то $P=\prod_{i=1}^{m} p_i^{k_i}$

так как при повторении i символа k_i раз k уменьшается до m. Из теории вероятностей известно, что, при достаточно длинных сообщениях (большое число символов k) $k_i \approx k \cdot p_i$ и тогда вероятность сообщений будет равняться $P = \prod_{i=1}^{m} p_i^{k_i}$

Тогда окончательно получим

$$I = -\log P = -k\sum_{i=1}^{m} p_i \log p_i$$

Данное выражение называется формулой Шеннона для определения количества информации. Формула Шеннона для количества информации на отдельный символ сообщения совпадает с энтропией. Тогда количество информации сообщения, состоящего из k символов будет равняться $I=k \cdot H$

Количество информации, как мера снятой неопределенности

При передаче сообщений, о какой-либо системе происходит уменьшение неопределенности. Если о системе все известно, то нет смысла посылать сообщение. Количество информации измеряют уменьшением энтропии.

Количество информации, приобретаемое при полном выяснении состояния некоторой физической системы, равно энтропии этой системы:

$$I = -\sum_{i=1}^{n} p_i \log p_i$$

Количество информации I – есть осредненное значение логарифма вероятности состояния. Тогда каждое отдельное слагаемое $-\log p_i$ необходимо рассматривать как частную информацию, получаемую от отдельного сообщения, то есть

$$I_i = -\log p_i$$

Избыточность информации

Если бы сообщения передавались с помощью равновероятных букв алфавита и между собой статистически независимых, то энтропия таких сообщений была бы максимальной. На самом деле реальные сообщения строятся из не равновероятных букв алфавита с наличием статистических связей между буквами. Поэтому энтропия реальных сообщений $-H_p$, оказывается много меньше оптимальных сообщений $- H_o$. Допустим, нужно передать сообщение, содержащее количество информации, равное I. Источнику, обладающему энтропией на букву, равной H_p, придется затратить некоторое число n_p, то есть

$$I = n_p H_p$$

Если энтропия источника была бы H_0, то пришлось бы затратить меньше букв на передачу этого же количества информации

$I= n_0 H_0$ $\quad n_0 = \frac{I}{H_0} < n_p$

Таким образом, часть букв $n_p - n_o$ являются как бы лишними, избыточными. Мера удлинения реальных сообщений по сравнению с оптимально закодированными и представляет собой избыточность D.

$$D = 1 - \frac{H_p}{H_0} = 1 - \frac{n_0}{n_p} = \frac{n_p - n_0}{n_p} \qquad (3)$$

Но наличие избыточности нельзя рассматривать как признак несовершенства источника сообщений. Наличие избыточности способствует повышению помехоустойчивости сообщений. Высокая избыточность естественных языков обеспечивает надежное общение между людьми.

Частотные характеристики текстовых сообщений

Важными характеристиками текста являются повторяемость букв, пар букв (биграмм) и вообще m-ок (m-грамм), сочетаемость букв друг с другом, чередование гласных и согласных и некоторые другие. Замечательно, что эти характеристики являются достаточно устойчивыми. Идея состоит в подсчете чисел вхождений каждой n^m возможных m-грамм в достаточно длинных открытых текстах $T=t_1t_2\ldots t_l$, составленных из букв алфавита $\{a_1, a_2, ..., a_n\}$. При этом просматриваются подряд идущие m-граммы текста

$t_1t_2...t_m$, $t_2t_3... t_{m+1}$, ..., $t_{i-m+1}t_{l-m+2}...t_l$.

Если $\Phi(a_{i_1}a_{i_2}...a_{i_m})$ – число появлений m-граммы $a_{i1}a_{i2}...a_{im}$ в тексте T, а L общее число подсчитанных m-грамм, то опыт показывает, что при достаточно больших L частоты

$\frac{\Phi(a_{i_1}a_{i_2}...a_{i_m})}{L}$ для данной m-граммы мало отличаются друг от друга.

Таблица 1. Частоты появления русских букв

Буква	Относительная частота	Буква	Относительная частота
а	0,062	р	0,040
б	0,014	с	0,045
в	0,038	т	0,053
г	0,013	у	0,021
д	0.025	ф	0,002
е, ё	0,072	х	0,009
ж	0,007	ц	0,004
з	0,016	ч	0,012
и	0,062	ш	0,006
й	0,010	щ	0,003
к	0,028	ы	0,016
л	0,035	ь, ъ	0,014
м	0,026	э	0,003
н	0,053	ю	0,006
о	0,090	я	0,018
п	0,023		

В силу этого, относительную частоту считают приближением вероятности P ($a_{i1}a_{i2}...a_{im}$) появления данной m-граммы в случайно выбранном месте текста (такой подход принят при статистическом определении вероятности).

Для русского языка частоты (в порядке убывания) знаков алфавита, в котором отождествлены Е с Ё, Ь с Ъ, а также имеется знак пробела (-) между словами, приведены в таблице 1.

Некоторая разница значений частот в приводимых в различных источниках таблицах объясняется тем, что частоты существенно зависят не только от длины текста, но и от его характера.

Устойчивыми являются также частотные характеристики биграмм, триграмм и четырехграмм осмысленных текстов.

Порядок выполнения работы

Определить количество информации (по Хартли), содержащееся в заданном сообщении, при условии, что значениями являются буквы кириллицы.

«Фамилия Имя Отчество» завершил ежегодный съезд эрудированных школьников, мечтающих глубоко проникнуть в тайны физических явлений и химических реакций

Построить таблицу распределения частот символов, характерные для заданного сообщения. Производится так называемая частотная селекция, текст сообщения анализируется как поток символов и высчитывается частота встречаемости каждого символа. Сравнить с имеющимися данными в табл 1.

На основании полученных данных определить среднее и полное количество информации, содержащееся в заданном сообщении. Оценить избыточность сообщения.

1. Построить таблицу распределения частот символов, характерных для заданного сообщения путём деления количества определённого символа в данном сообщении на общее число символов

По формуле

$$\sum_{i=1}^{m} p_i \log p_i$$

H= вычислил энтропию сообщения

2. Далее по формуле Шеннона для определения кол-ва информации

$$I = -\log P = -k\sum_{i=1}^{m} p_i \log p_i$$ вычислил кол-во информации в передаваемом сообщении

3. Вычислил избыточность D по формуле

$$D = 1 - \frac{H_T}{H_r} = 1 - \frac{x_v}{x_r} = \frac{x_r - x_v}{x_r}$$

Отчет

Отчет должен быть оформлен в текстовом редакторе и содержать:

- наименование работы;
- цель работы;
- задание;
- последовательность выполнения работы;
- ответы на контрольные вопросы;
- вывод о проделанной работе.

Контрольные вопросы

1. Дать определение понятие энтропия?
2. Что означает вероятностный способ измерения информации?
3. Что означает статическое определение вероятности?
4. Запишите уравнение Хартли?
5. Какие основные разработки внес в основу теории информации Шеннон?

Практическое занятие

«Поиск энтропии случайных величин. Энтропийное кодирование»

Цель: научиться вычислять энтропию случайной величины.
Оборудование: ПК.
Программное обеспечение: операционная система, калькулятор, текстовый редактор.

Теоретические основы

Энтропия в теории информации — мера хаотичности информации, неопределённость появления какого-либо символа первичного алфавита. При отсутствии информационных потерь численно равна количеству информации на символ передаваемого сообщения.

Так, возьмём, например, последовательность символов, составляющих какое-либо предложение на русском языке. Каждый символ появляется с разной частотой, следовательно, неопределённость появления для некоторых символов больше, чем для других. Если же учесть, что некоторые сочетания символов встречаются очень редко, то неопределённость ещё более уменьшается (в этом случае говорят об энтропии n-ого порядка. Концепции информации и энтропии имеют глубокие связи друг с другом, но, несмотря на это, разработка теорий в статистической механике и теории информации заняла много лет, чтобы сделать их соответствующими друг другу.

Энтропия независимых случайных событий *x* с *n* возможными состояниями (от 1 до *n*) рассчитывается по формуле:

$$H(x) = -\sum_{i=1}^{n} p(i)\log_2 p(i)$$

Эта величина также называется *средней энтропией сообщения*.

Величина $\log_2 \frac{1}{p(i)}$ называется *частной энтропией*, характеризующей только i-е состояние.
Таким образом, энтропия события *x* является суммой с противоположным знаком всех произведений относительных частот появления события *i*, умноженных на их же двоичные логарифмы (основание 2 выбрано только для удобства работы с информацией, представленной в двоичной форме). Это определение для дискретных случайных событий можно расширить для функции распределения вероятностей.
Шеннон вывел это определение энтропии из следующих предположений:

- мера должна быть непрерывной; т. е. изменение значения величины вероятности на малую величину должно вызывать малое результирующее изменение энтропии;
- в случае, когда все варианты (буквы в приведенном примере) равновероятны, увеличение количества вариантов (букв) должно всегда увеличивать полную энтропию;
- должна быть возможность сделать выбор (в нашем примере букв) в два шага, в которых энтропия конечного результата должна будет является суммой энтропий промежуточных результатов.

Шеннон показал, что любое определение энтропии, удовлетворяющее этим предположениям, должно быть в форме:

$$-K\sum_{i=1}^{n} p(i)\log_2 p(i)$$

где *K* — константа (и в действительности нужна только для выбора единиц измерения).
Шеннон определил, что измерение энтропии ($H = -\,p_1 \log_2 p_1 - \ldots - p_n \log_2 p_n$), применяемое к источнику информации, может определить требования к минимальной пропускной способности канала, требуемой для надежной передачи информации в виде закодированных двоичных чисел.
Для вывода формулы Шеннона необходимо вычислить математическое ожидания «количества информации», содержащегося в цифре из источника информации. Мера энтропии Шеннона выражает неуверенность реализации случайной переменной. Таким образом, энтропия является разницей между информацией, содержащейся в сообщении, и той частью информации, которая точно известна (или хорошо предсказуема) в сообщении. Примером этого является избыточность языка — имеются явные статистические закономерности в появлении букв, пар последовательных букв, троек и т.д.
В общем случае **b-арная энтропия** (где b равно 2,3,...) источника $\mathcal{S}$ = (S,P) с исходным алфавитом $S = \{a_1, \ldots, a_n\}$ и дискретным распределением вероятности $P = \{p_1, \ldots, p_n\}$ где p_i является вероятностью a_i ($p_i = p(a_i)$) определяется формулой:

$$H_b(\mathcal{S}) = -\sum_{i=1}^{n} p_i \log_b p_i$$

Определение энтропии Шеннона очень связано с понятием термодинамической энтропии. Больцман и Гиббс проделали большую работу по статистической термодинамике, которая способствовала принятию слова «энтропия» в информационную теорию. Существует связь между понятиями энтропии в термодинамике и теории информации. Например, демон Максвелла также противопоставляет термодинамическую энтропию информации, и получение какого-либо количества информации равно потерянной энтропии.

СВОЙСТВА ЭНТРОПИИ

*1.*Энтропия является вещественной и неотрицательной величиной.
*2.*Энтропия – величина ограниченная.
*3.*Энтропия обращается в нуль лишь в том случае, если вероятность одного из состояний равна единице; тогда вероятности всех остальных состояний, естественно, равны нулю. Это положение соответствует случаю, когда состояние источника полностью определено.

*4.*Энтропия максимальна, когда все состояния источника равновероятны.
5. Энтропия источника и с двумя состояниями u1 и u2 изменяется от нуля до единицы, достигая максимума при равенстве их вероятностей:
p(u1) = p= p(u2) = 1 — p = 0,5.
*6.*Энтропия объединения нескольких статистически независимых источников информации равна сумме энтропии исходных источников.
7. Энтропия характеризует среднюю неопределенность выбора одного состояния из ансамбля. При ее определении используют только вероятности состояний, полностью игнорируя их содержательную сторону. Поэтому энтропия не может служить средством решения любых задач, связанных с неопределенностью.
8. Энтропия как мера неопределенности согласуется с экспериментальными данными, полученными при изучении психологических реакций человека, в частности реакции выбора. Установлено, что время безошибочной реакции на последовательность беспорядочно чередующихся равновероятных раздражителей (например, загорающихся лампочек) растет с увеличением их числа так же, как энтропия. Это время характеризует неопределенность выбора одного раздражителя. Замена равновероятных раздражителей неравновероятными приводит к снижению среднего времени реакции ровно настолько, насколько уменьшается энтропия.
Дифференциальной энтропией случайной величины X называется величина:

$$H_Д(x)=H(x)-H(y)= -\int_{-\infty}^{+\infty} p_x(x)*\log_2 d*p_x(x)dx$$

Если произвести квантование случайных величин X_1, $X_2 \ldots X_n$ по уровню с числом уровней квантования равным m , то возможное число реализаций длительностью T_n станет конечным и равным $M = m^n$.
Каждая из реализаций $C_1, C_2, \ldots C_i, \ldots C_m$ будет иметь определенную вероятность появления в эксперименте по наблюдению реализаций. Тогда неопределенность (энтропия) и количество информации в реализации (в среднем по всем реализациям) определяются равенством

$$H_n = -\sum_{i=1}^{M} P(C_i)\log(P(C_i)) \qquad H = \frac{H_n}{n} = -\frac{1}{n}\sum_{i=1}^{M} P(C_i)\log(P(C_i))$$

Энтропия и количество информации на одну степень свободы (на одну выборку) равны

$$H = \frac{H_n}{n} = -\frac{1}{n}\sum_{i=1}^{M} P(C_i)\log(P(C_i))$$

Избыточность показывает, какая доля максимально возможной при заданном объеме алфавита неопределенности не используется источником.
$\mu=(H_{max}-H_u)/H_{max}$,
Где H_u – энтропия рассматриваемого источника, H_{max} – максимально возможное значение его энтропии, которое может быть достигнуто подбором распределения и ликвидацией взаимозависимости элементов алфавита. Так, для дискретного источника с M элементами $H_{max}=\log M$

Выполнение расчетных задач

Задача №1
Доказать, что H(x,y)≤H(x)+H(y).
Решение:
По определению,

$$H(x,y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j)\log p(x_i, y_j).$$

Для статистически зависимых событий $p(x_i, y_j) = p(x_i)\, p(y_j / x_i)$.

$$H(x,y)=-\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_i)\,p(y_j/x_i)\operatorname{Log}\left[p(x_i)\,p(y_j/x_i)\right]=$$

$$=-\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_i)\,p(y_j/x_i)\left[\log p(x_i)+\log p(y_j/x_i)\right]=$$

$$=-\sum_{i=1}^{n}p(x_i)\log p(x_i)\underbrace{\sum_{j=1}^{m}p(y_j/x_i)}_{1}+\sum_{i=1}^{n}p(x_i)\left[\underbrace{-\sum_{j=1}^{m}p(y_j/x_i)\log p(y_j/x_i)}_{H(y/x_i)}\right]=$$

$$=\underbrace{-\sum_{i=1}^{n}p(x_i)\log p(x_i)}_{H(x)}+\underbrace{\sum_{i=1}^{n}p(x_i)\,H(x,y/x_i)}_{H(y/x)}=H(x)+H(y/x).$$

$H(y/x_i)$ - это частная энтропия Y при условии, что известно состояние X=x_i.
Наличие информации о состоянии X не может увеличить неопределенность состояния Y, но может уменьшить его в случае зависимости Y от X. Значит, условная энтропия $H(y/x_i)$ не больше безусловной энтропии $H(y)$, то есть $H(y/x_i)\le H(y)$. Тогда средняя условная энтропия

$$H(y/x)=\sum_{i=1}^{n}p(x_i)H(y/x_i)\le\sum_{i=1}^{n}p(x_i)H(y)=H(y)\underbrace{\sum_{i=1}^{n}p(x_i)}_{1}=H(y),$$

то есть $H(y/x)\le H(y)$.

Значит, $H(x,y)=H(x)+H(y/x)\le H(x)+H(y).$

Задача №2

Показать, что для регулярной марковской цепи энтропия $H(x)^{(r)}$ за r шагов равняется энтропии за один шаг, умноженной на число шагов r.

Решение:

Регулярная цепь Маркова полностью характеризуется матрицей переходных вероятностей

$$P=\begin{bmatrix} p_{11} & \dots & p_{1m} \\ . & . & . \\ p_{m1} & & p_{mm}\end{bmatrix}$$

и предельным стационарным распределением вероятностей состояний $(p_1, p_2, \dots, p_m)$.

В стационарном режиме энтропия за один шаг не зависит от номера шага и равна

$$H(X)^{(1)}=\sum_{k=1}^{m}p_k H_k(x),$$

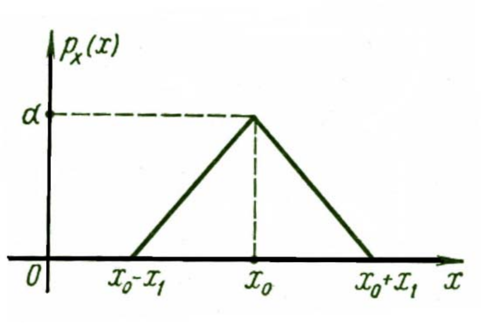
p_k - стационарная вероятность k-го состояния,

$H_k(x)=-\sum_{i=1}^{m}p_{ki}\log p_{ki}$ - энтропия в k-м состоянии.

Энтропия за r шагов равна сумме энтропий за каждый шаг. Так как энтропия за каждый шаг одинакова, то сумма энтропий равна $H(X)^{(r)}=r\cdot H(X)^{(1)}$.

Задача №3

Вычислить энтропию $H_д(x)$ распределения $p_x(x)$ изображенного на рисунке.
Построить график зависимости $H_д(x)$ в функции параметра d.



Решение:

Из условия нормировки $S = d \cdot x_1 = 1 \Rightarrow x_1 = \frac{1}{d}$.

Плотность данного распределения

$$p_x(x) = \begin{cases} d + \dfrac{d(x - x_0)}{x_1}, & x_0 - x_1 \le x \le x_0 \\ d - \dfrac{d(x - x_0)}{x_1}, & x_0 \le x \le x_0 + x_1 \end{cases}$$

Дифференциальная энтропия

$$H_Д(x) = -\int_{-\infty}^{\infty} p_X(x) \log p_X(x)\, dx = -\left[\int_{x_0 - x_1}^{x_0} p_X(x) \ln p_X(x)\, dx + \int_{x_0}^{x_0 + x_1} p_X(x) \ln p_X(x)\, dx\right].$$

$$\int u \ln u\, du = \ln u \cdot \frac{u^2}{2} - \int (\ln u)' \frac{u^2}{2} du = -\frac{u^2}{4} + \frac{1}{2} u^2 \ln u + C = \frac{1}{2} u^2 \left(\ln u - \frac{1}{2}\right) + C$$

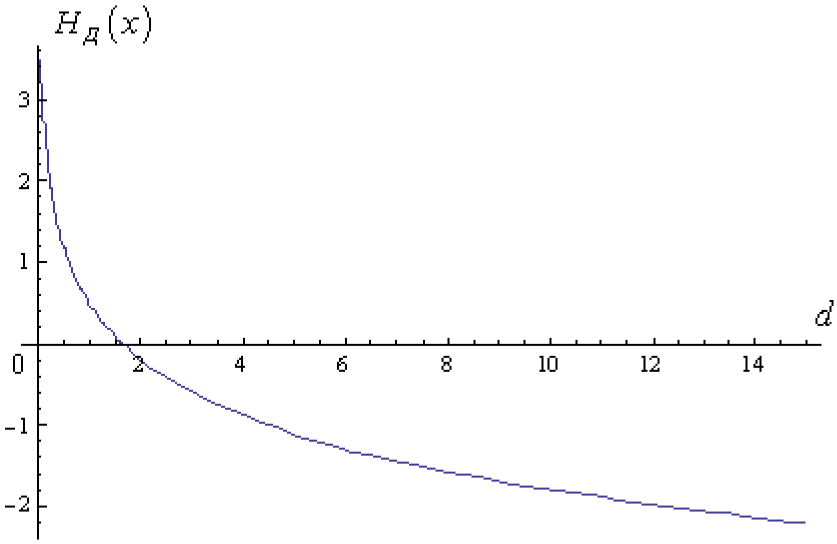
$$\int_{x_0 - x_1}^{x_0} \left(d + \frac{d}{x_1}(x - x_0)\right) \ln\left(d + \frac{d}{x_1}(x - x_0)\right) dx =$$

$$= \frac{x_1}{d} \left(\frac{1}{2}\left(d + \frac{d}{x_1}(x - x_0)\right)^2 \left(\ln\left(d + \frac{d}{x_1}(x - x_0)\right) - \frac{1}{2}\right)\right)\Bigg|_{x_0 - x_1}^{x_0} = \frac{x_1}{d} \cdot \frac{1}{2} d^2 \left(\ln d - \frac{1}{2}\right).$$

$$\int_{x_0}^{x_0 + x_1} \left(d - \frac{d}{x_1}(x - x_0)\right) \ln\left(d - \frac{d}{x_1}(x - x_0)\right) dx =$$

$$= -\frac{x_1}{d} \left(\frac{1}{2}\left(d - \frac{d}{x_1}(x - x_0)\right)^2 \left(\ln\left(d - \frac{d}{x_1}(x - x_0)\right) - \frac{1}{2}\right)\right)\Bigg|_{x_0}^{x_0 + x_1} = \frac{x_1}{d} \cdot \frac{1}{2} d^2 \left(\ln d - \frac{1}{2}\right).$$

$$H_Д(x) = -2 \cdot \frac{x_1}{d} \cdot \frac{1}{2} d^2 \left(\ln d - \frac{1}{2}\right) = -\frac{x_1}{d} d^2 \left(\ln d - \frac{1}{2}\right) = -\frac{1}{d^2} d^2 \left(\ln d - \frac{1}{2}\right) = -\ln d + \frac{1}{2}.$$



Задача 4

В результате полной дезорганизации управления m самолетов летят произвольными курсами. Управление восстановлено, и все самолеты взяли общий курс со среднеквадратической ошибкой отклонения от курса σ=3^0. Найти изменение энтропии, считая, что в первом случае имело место равномерное распределение вероятностей углов, а во втором случае – нормальное.
Ответ: 4.86 бита
Решение.
Начальное распределение вероятностей углов курсов самолетов равномерное в интервале от $a = 0$ до $b = 360^\circ = 2\pi$ рад с плотностью вероятности $p_{1x}(x) = \dfrac{1}{b-a}, \ a \le x \le b$.
Дифференциальная энтропия этого распределения

$$H_{1д}(x) = -\int_a^b \frac{1}{b-a}\log_2\frac{1}{b-a}dx = -\log_2\frac{1}{b-a} = \log_2(b-a) = \log_2 2\pi = 2{,}65$$

бит.
Конечное распределение вероятностей углов курсов самолетов нормальное с параметрами $a = 0$ рад, $\sigma = 3^\circ = \dfrac{\pi}{60}$ и плотностью вероятности $p_{2x}(x) = \dfrac{1}{\sigma\sqrt{2\pi}}e^{-x^2/(2\sigma^2)}$.
Дифференциальная энтропия этого распределения

$$H_{2д}(x) = -\int_{-\infty}^{\infty} p_{2x}(x)\log_2\left(\frac{1}{\sigma\sqrt{2\pi}}e^{-x^2/(2\sigma^2)}\right)dx =$$

$$= -\log_2\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} p_{2x}(x)dx + \frac{\log_2 e}{2\sigma^2}\int_{-\infty}^{\infty} x^2 p_{2x}(x)dx =$$

$$= -\log_2\frac{1}{\sigma\sqrt{2\pi}} + \frac{\log_2 e}{2\sigma^2}\cdot\sigma^2 = \log_2\sigma\sqrt{2\pi} + \frac{\log_2 e}{2} = \log_2\frac{\pi\sqrt{2\pi e}}{60} = -2{,}21\ бит.$$

Изменение энтропии $\Delta H(x) = H_{2д}(x) - H_{1д}(x) = -2{,}21 - 2{,}65 = -4{,}86$ бит.
Энтропия уменьшилась на 4,86 бит.
Задача 5
Измерительное устройство вырабатывает временные интервалы, распределенные случайным образом в пределах от 100 до 500 мс. Как изменится энтропия случайной величины при изменении точности измерения с 1 мс до 1 мкс?
Ответ: Энтропия увеличивается примерно на 10 бит
Решение.
При точности 1мс дискретная случайная величина X – результат измерения – может равновероятно принимать одно из $n = \dfrac{500-100}{1} = 400$ значений. Энтропия равна $H_1(x) = \log_2 n$.
При точности 1мкс дискретная случайная величина X – результат измерения – может равновероятно принимать одно из $m = \dfrac{500-100}{10^{-3}} = 400\cdot 10^3 = 1000n$ значений. Энтропия равна $H_2(x) = \log_2 m$.
Изменение энтропии

$$\Delta H(x) = H_2(x) - H_1(x) = \log_2 m - \log_2 n = \log_2 1000n - \log_2 n = \log_2 1000 \approx \log_2 1024 = 10$$

бит.
Энтропия увеличилась примерно на 10 бит.
Задача 6
Записать отношения между энтропиями:
H(x), H(y), H(x|y), H(y|x), H(x,y), H(x|y_j), H(y|x_i)
Решение.

Связь между энтропией совместного распределения и полными энтропиями и средними условными энтропиями

$$H(x,y)=H(x)+H(y/x)=H(y)+H(x/y)$$

.

Если *x* и *y* независимы, то $H(y/x)=H(y),\ H(x/y)=H(x)$ и энтропия совместного распределения $H(x,y)=H(x)+H(y)$ максимальна, так как $H(y/x)\le H(y),\ H(x/y)\le H(x)$

.

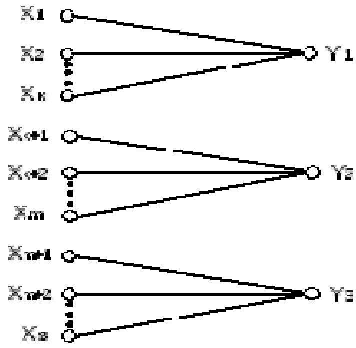
Связь между частными и средними условными энтропиями

$$H(y/x)=\sum_{i=1}^{n}p(x_i)H(y/x_i),\ H(x/y)=\sum_{j=1}^{m}p(y_j)H(x/y_j)$$

.

Задача 7

На рисунке представлена диаграмма канала со слабым разрешением. Определить количество информации, передаваемое по каналу.



Ответ: I (x,y)=H(x)

Решение:

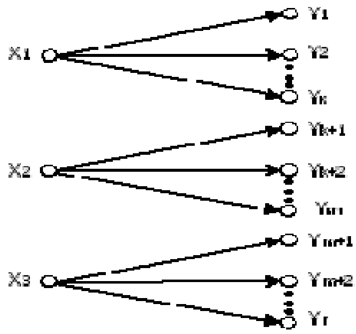
Количество переданной по каналу информации равно $I=H-H_{апост.}$.

$H=H(y)$ – энтропия полученных элементов *y* до отправления элемента *x*; $H_{апост.}=H(y/x)$ - энтропия полученных элементов *y* после отправления элемента *x*. Для данного канала со слабым разрешением $H(y/x)=0$, так как полученный элемент *y* однозначно определяется по отправленному элементу *x*.

Получаем $I=H(y)$.

Задача 2

На рисунке представлена диаграмма канала с неоднозначностью. Определить количество информации, передаваемое по каналу.



ответ: I (x,y)=H(x)

Решение:

Количество переданной по каналу информации равно $I = H - H_{апост.}$.

$H = H(x)$ – энтропия отправленных элементов *x* до получения элемента *y*; $H_{апост.} = H(x/y)$ - энтропия отправленных элементов *x* после получения элемента *y*. Для данного канала с неоднозначностью $H(x/y) = 0$, так как полученный элемент *y* однозначно определяет отправленный элемент *x*.

Получаем $I = H(x)$.

Задачи по вычислению энтропии

1. Найдите энтропию для числа белых шаров при извлечении двух шаров из урны, содержащей два белых и один черный шар.
2. Найдите энтропию для числа козырных карт при извлечении двух карт из колоды в 36 карт.
3. Какую степень неопределенности содержит опыт угадывания суммы очков на извлеченной кости из полного набора домино?
4. Найдите энтропию для числа тузов при извлечении трех карт из карт с картинками.
5. Найдите дифференциальную энтропию для равномерного распределения.
6. Найдите дифференциальную энтропию для показательного закона распределения, если известно, что случайная величина *x* принимает значение меньше единицы с вероятностью 0,5.

Отчет

Отчет должен быть оформлен в текстовом редакторе и содержать:

- наименование работы;
- цель работы;
- задание;
- последовательность выполнения работы;
- ответы на контрольные вопросы;
- вывод о проделанной работе.

Контрольные вопросы

1. Как определяется энтропия дискретных случайных величин?
2. Приведите примеры энтропий для классических законов распределения.

Практическая работа
«ПУ кодирование»

Цель: Познакомиться с различными кодировками символов, используя текстовые редакторы, выполнить задания в различных текстовых приложениях.

Оборудование: ПК.

Программное обеспечение: операционная система, текстовые редакторы.

Теоретические основы

Правило цифрового представления символов следующее: каждому символу ставится в соответствие некоторое целое число, то есть каждый символ нумеруется.

Пример:

Рассмотрим последовательность строчных букв русского алфавита: а, б, в, г, д, е, ё, ж, з, и, й. к, л, м. н. о, п, р, с, т, у, ф, х, ц, ч, ш, щ, ъ, ы, в, э, ю, я. Присвоив каждой букве номер от 0 до 33. получим простейший способ представления символов. Последнее число - 32 в двоичной форме имеет вид 100000, то есть для хранения символа в памяти понадобится 6 бит.Так как с помощью шести бит можно представить число $2^6 - 1 = 63$, то шести бит будет достаточно для представления 64 букв.

Имеются разные стандарты для представления, символов, которые отличаются лишь порядком нумерации символов. Наиболее распространён американский стандартный код для информационного обмена - ASCII [American Standard-Code for Information Interchange] введён в США в 1963г. В 1977 году в несколько модифицированном виде он был принят в качестве всемирного стандарта Международной организации стандартов [International Standards Organization -. ISO] под названием ISO-646.
Согласно этому стандарту каждому символу поставлено в соответствие число от 0 до 255. Символы от 0 до 127 - латинские буквы, цифры и знаки препинания - составляют постоянную часть таблицы. Остальные символы используются для представления национальных алфавитов. Конкретный состав этих символов определяется кодовой страницей.
В русской версии ОС Windows95 используется кодовая, страница 866. В ОС Linux для представления русских букв более употребительна кодировка КОИ-8. Недостатки такого способа кодировки национального, алфавита очевидны.
Во-первых, невозможно одновременное представление русских и ,например, французских букв. Во-вторых, такая кодировка совершенно непригодна для представления, китайских иероглифов. В 1991 году была создана некоммерческая организация Unicode, в которую входят представители ряда фирм (Borland. IBM, Noyell, Sun и др) и которая занимается развитием и внедрением нового стандарта.
Кодировка Unicode использует 16 разрядов ,и может содержать 65536 символов. Это символы большинства народов мира, элементы иероглифов, спецсимволы, 5000 – мест для частного использования, резерв из 30000 мест.

Пример:

ASCII-код символа A= 65_{10} =41_{16}= 01000111_2;
Unicode-код символа C= 67_{10}=0000000001100111_2

Задания

1. Закодируйте свое имя, фамилию и отчество с помощью одной из таблиц (win-1251, KOI-8)
2. Раскодируйте ФИО соседа
3. Закодируйте следующие слова, используя таблицы ASCII-кодов: ИНФОРМАТИЗАЦИЯ, МИКРОПРОЦЕССОР, МОДЕЛИРОВАНИЕ
4. Раскодируйте следующие слова, используя таблицы ASCII-кодов:

88 AD E4 AE E0 AC A0 E2 A8 AA A0
50 72 6F 67 72 61 6D
43 6F 6D 70 75 74 65 72 20 49 42 4D 20 50 43

5. Текстовый редактор Блокнот

Открыть блокнот.
а) Используя клавишу Alt и малую цифровую клавиатуру раскодировать фразу:
145 170 174 224 174 255 170 160 173 168 170 227 171 235;
Технология выполнения задания: При удерживаемой клавише Alt, набрать на малой цифровой клавиатуре указанные цифры. Отпустить клавишу Alt, после чего в тексте появится буква, закодированная набранным кодом.
б) Используя ключ к кодированию, закодировать слово – зима;
Технология выполнения задания: Из предыдущего задания выяснить, каким кодом записана буква а. Учитывая, что буквы кодируются в алфавитном порядке, выяснить коды остальных букв. Что вы заметили при выполнении этого задания во время раскодировки? Запишите свои наблюдения.

6. Текстовый процессор.

Технология выполнения задания: рассмотрим на примере: представить в различных кодировках слово <u>Кодировка</u>

Решение:

- Создать новый текстовый документ в текстовом редакторе;
- Выбрать – Команда – Вставка – Символ.

В открывшемся окне «Символ» установить из: Юникод (шестн.),

- В наборе символов находим букву **К** и щелкнем на ней левой кнопкой мыши (ЩЛКМ).
- В строке код знака появится код выбранной буквы 041А (незначащие нули тоже записываем).
- У буквы **о** код – 043Е и так далее: д – 0434, и – 0438, р – 0440, о – 043Е, в – 0432, к – 043А, а – 0430.
- Установить Кириллица (дес.)
- К – 0202, о – 0238, д – 0228, и – 0232, р – 0240, о – 0238, в –0226, к – 0202, а –0224.

7. Открыть Текстовый редактор
Используя окно «Вставка символа» выполнить задания: Закодировать слово **Forest**
а) Выбрать шрифт Courier New, кодировку ASCII(дес.) Ответ: **70 111 114 101 115 116**
б) Выбрать шрифт Courier New, кодировку Юникод(шест.) Ответ: **0046 006F 0072 0665 0073 0074**
в) Выбрать шрифт Times New Roman, кодировку Кирилица(дес.) Ответ: **70 111 114 101 115 116**
г) Выбрать шрифт Times New Roman, кодировку ASCII(дес.) Ответ: **70 111 114 101 115 116**
Вывод: ____
Выполнение лабораторной работы оформить в виде таблицы.
8. Буква Z имеет десятичный код 90, а z – 122. Записать слово «sport» в десятичном коде.
9. С помощью десятичных кодов зашифровано слово «info» 105 110 102 111. Записать последовательность десятичных кодов для этого же слова, но записанного заглавными буквами.
10. Буква Z имеет десятичный код 90, а z – 122. Записать слово «forma» в десятичном коде.
11. С помощью десятичных кодов зашифровано слово «port» 112 111 114 116. Записать последовательность десятичных кодов для этого же слова, но записанного заглавными буквами. Ответ: **80 79 82 84**

Отчет

Отчет должен быть оформлен в текстовом редакторе и содержать:
- наименование работы;
- цель работы;
- задание;
- последовательность выполнения работы;
- ответы на контрольные вопросы;
- вывод о проделанной работе.

Контрольные вопросы

1. Какие возможности предоставляет текстовые редакторы по работе с символами?
2. Какие вы знаете алгоритмы кодирования информации?
3. Где применяется алгоритм кодирования информации?

Практические задания

Вариант №1

1. Измерьте информационный объем сообщения «Ура! Скоро Новый год!» в битах, байтах, килобайтах (Кб), мегабайтах (Мб).
2. Измерьте примерную информационную емкость одной страницы любого своего учебника, всего учебника.
3. Сколько таких учебников может поместиться на дискете 1,44 Мб, на винчестере в 1 Гб.
4. В детской игре «Угадай число» первый участник загадывает целое число от 1 до 32. Второй участник задает вопросы: «Загаданное число больше числа ___?». Какое количество вопросов при правильной стратегии гарантирует угадывание?

Указание: Вопрос задавайте таким образом, чтобы информационная неопределенность (число вариантов) уменьшалась в два раза.

5. Яд находится в одном из 16 бокалов. Сколько единиц информации будет содержать сообщение о бокале с ядом?
6. Подсчитайте объем информации, содержащейся в романе А. Дюма "Три мушкетера", и определите, сколько близких по объему произведений можно разместить на одном лазерном диске? (590 стр., 48 строк на одной странице, 53 символа в строке).
7. Информационное сообщение объемом 1.5 Кбайта содержит 3072 символа. Сколько символов содержит алфавит, при помощи которого было записано это сообщение? (Объяснение решения задачи на доске).
8. Подсчитать в килобайтах количество информации в тексте, если текст состоит из 600 символов, а мощность используемого алфавита – 128 символов.
9. Скорость информационного потока – 20 бит/сек. Сколько времени потребуется для передачи информации объемом в 10 килобайт.
10. Книга, набранная с помощью компьютера, содержит 240 страниц; на каждой странице — 42 строки, в каждой строке — 64 символа. Каков объем информации в книге?

Вариант №2

1. Сравните (поставьте знак отношения)
   - 200 байт и 0,25 Кбайт.
   - 3 байта и 24 бита.
   - 1536 бит и 1,5 Кбайта.
   - 1000 бит и 1 Кбайт.
   - 8192 байта и 1 Кбайт.
2. В барабане для розыгрыша лотереи находится 32 шара. Сколько информации содержит сообщение о первом выпавшем номере (например, выпал номер 15)?
3. При игре в кости используется кубик с шестью гранями. Сколько бит информации получает игрок при каждом бросании кубика?
4. Книга, набранная с помощью компьютера, содержит 150 страниц; на каждой странице — 40 строк, в каждой строке — 60 символов. Каков объем информации в книге?
5. Проводят две лотереи: «4 из 32» и «5 из 64» Сообщение о результатах какой из лотерей несет больше информации?
6. На диске объемом 100 Мбайт подготовлена к выдаче на экран дисплея информация: 24 строчки по 80 символов, эта информация заполняет экран целиком. Какую часть диска она занимает?
7. В школьной библиотеке 16 стеллажей с книгами. На каждом стеллаже 8 полок. Библиотекарь сообщил Пете, что нужная ему книга находится на пятом стеллаже на третьей сверху полке. Какое количество информации библиотекарь передал Пете?
8. В коробке лежат 7 цветных карандашей. Какое количество информации содержит сообщение, что из коробки достали красный карандаш?
9. Какое количество информации несет сообщение: “Встреча назначена на сентябрь”.
10. Сообщение занимает 3 страницы по 25 строк. В каждой строке записано по 60 символов. Сколько символов в использованном алфавите, если все сообщение содержит 1125 байтов?

Вариант № 3

1. Какое количество информации будет содержать зрительное сообщение о цвете вынутого шарика, если в непрозрачном мешочке находится 50 белых, 25 красных, 25 синих шариков?

2. В корзине лежит 16 шаров разного цвета. Сколько информации несет сообщение, что достали белый шар?

3. В корзине лежат черные и белые шары. Среди них18 черных шаров. Сообщение о том, что достали белый шар, несет 2 бита информации. Сколько всего шаров в корзине?

4. В некоторой стране автомобильный номер длиной 5 символов составляется из заглавных букв (всего используется 30 букв) и десятичных цифр в любом порядке. Каждый символ кодируется одинаковым и минимально возможным количеством бит, а каждый номер – одинаковым и минимально возможным количеством байт. Определите объем памяти, необходимый для хранения 50 автомобильных номеров.

5. Два текста содержат одинаковое количество символов. Первый текст записан на русском языке, а второй на языке племени нагури, алфавит которого состоит из 16 символов. Чей текст несет большее количество информации?

6. Объем сообщения, содержащего 2048 символов, составил 1/512 часть Мбайта. Определить мощность алфавита

7. Известно, что видеопамять компьютера имеет объем 512 Кбайт. Разрешающая способность экрана 640 на 200. Сколько страниц экрана одновременно разместится в видеопамяти при палитре

а) из 8 цветов; б) 16
цветов;

в) 256 цветов?

8. Подсчитать, сколько места будет занимать одна минута цифрового звука на жестком диске или любом другом цифровом носителе, записанного с частотой

а) 44.1 кГц;

б) 11 кГц;

в) 22 кГц;

г) 32 кГц и разрядностью 16 бит

9. Для кодирования нотной записи используется 7 значков-нот. Каждая нота кодируется одними тем же минимально возможным количеством бит. Чему равен информационный объем сообщения, состоящего из 180 нот?

10. Рассчитать время звучания моноаудиофайла, если при 16-битном кодировании и частоте дискретизации 32кГц его объем равен 6300 Кбайт.

11. Музыкальная запись выполнена в формате CDDA (частота дискретизации 44100 Гц, 16 бит, стерео) и имеет продолжительность 19 мин 20 сек. Сколько секунд займет передача этой записи по каналу с пропускной способностью 16000 байт/сек?

12. После изменения свойств Рабочего стола монитор приобрёл разрешение 1024*768 точек и получил возможность отображать 65 536 цветов. Какой объём видеопамяти необходим для текущего изображения Рабочего стола?

13. В процессе преобразования растрового графического изображения количество цветов в палитре уменьшилось с 16 777 216 до 256. Во сколько раз при этом уменьшился информационный объём изображения?

Время на выполнение: 60 минут

Бальная оценка заданий

Один правильный ответ - 10 баллов Итого: 10 ответов

- 100 баллов

Таблица перевода баллов в оценки:

Баллы	Оценка
менее 35	2 (неудовлетворительно)
35-49	3 (удовлетворительно)
50-74	4 (хорошо)
75-100	5 (отлично)

Лист регистрации изменений и дополнений ФОС по дисциплине ОП.12 Основы теории информации

№ изме-нени я	Номера листов		Основания для внесения изменений	Дата	№ протокола педсовета	Дата введения изменений	Подпись	Расшифровка подписи
	Замененных	Новых						