**Белорусский национальный технический университет**
Факультет Международный институт дистанционного образования
Кафедра «Информационные системы и технологии»

**ЭЛЕКТРОННЫЙ УЧЕБНО-МЕТОДИЧЕСКИЙ КОМПЛЕКС ПО УЧЕБНОЙ ДИСЦИПЛИНЕ**

# ТЕОРИЯ ИНФОРМАЦИИ

для специальностей:

1-40 01 01 «Программное обеспечение информационных технологий»
1-40 05 01 «Информационные системы и технологии»

Составители:
Бояршинова Оксана Александровна, к.ф.-м.н., зав. кафедрой

Минск БНТУ 2022

к оглавлению

# Теория информации

**Перечень материалов**

Конспект лекций, материалы для лабораторных занятий и контрольных работ, вспомогательный раздел.

**Пояснительная записка**

Цели данного ЭУМК – повышение эффективности организации учебного процесса с использованием дистанционных технологий; предоставление возможности студентам заниматься самообразованием, пользуясь комплектом учебно-методических материалов по дисциплине «Теория информации».

ЭУМК содержит четыре раздела: теоретический, практический, контроля знаний и вспомогательный.

Теоретический раздел представлен конспектом лекций. Лекционный материал подготовлен в соответствии с основными разделами и темами учебной программы.

Практический раздел представлен лабораторными работами, которые помогут освоить теоретический материал дисциплины.

Раздел контроля знаний включает 20 вариантов контрольной работы, содержащие по 10 задач каждый, требования к оформлению контрольной работы, вопросы к зачету.

Вспомогательный раздел представлен учебной программой, приложениями содержащими справочные материалы, списком рекомендуемой литературы.

Данное ЭУМК в первую очередь разработано для студентов МИДО дистанционной (заочной) формы получения образования, однако ЭУМК также может быть полезным студентам дневной формы получения образования и всем кто заинтересован в освоении дисциплины «Теория информации».

# СОДЕРЖАНИЕ

# РАЗДЕЛ 1. ТЕОРЕТИЧЕСКИЙ

## 1. Понятие системы счисления

Представление целых положительных чисел с помощью письменных знаков (символов) называется ***нумерацией***. Письменные знаки (символы), используемые при нумерации, называются ***цифрами***. Необходимо четко понимать различие между числом и символом (группой символов), которым пользуются для его письменного воспроизведения. Например, с одной стороны, для изображения числа «пять» могут использоваться цифра 5 (десятичная система нумерации), цифра V (римская система нумерации) или группа символов 101 (двоичная система нумерации). С другой стороны, группа символов 10 может обозначать число «десять» в десятичной системе или число 2 в двоичной системе. Иными словами, значение символа зависит от системы нумерации и его положения в записи, тогда как с числом всегда связана определенная количественная характеристика.

***Система счисления*** – совокупность правил записи чисел (способ соединения цифр для обозначения числа). Знаки, с помощью которых записываются числа (рис. 1), называются ***цифрами***, а их совокупность – ***алфавитом*** системы счисления.

В любой системе счисления цифры служат для обозначения чисел, называемых узловыми; остальные числа (алгоритмические) получаются в результате каких-либо операций из узловых чисел.

*Пример***.** У вавилонян узловыми являлись числа 1, 10, 60; в римской системе счисления узловые числа – это 1, 5, 10, 50, 100, 500 и 1000, обозначаемые соответственно I, V, X, L, C, D, M.

Системы счисления различаются выбором узловых чисел и способами образования алгоритмических чисел. Можно выделить следующие виды систем счисления:

1) унарная система;
2) непозиционные системы;
3) позиционные системы.

Простейшая и самая древняя система – так называемая ***унарная система*** счисления. В ней для записи любых чисел используется всего один символ – палочка, узелок, зарубка, камушек (рис. 2). Длина записи числа при таком кодировании прямо связана с его величиной, что роднит этот способ с геометрическим представлением чисел в виде отрезков. Именно унарная система лежит в фундаменте арифметики, и именно она до сих пор вводит первоклассников в мир счёта. Унарную систему ещё называют системой бирок.

**ОБОЗНАЧЕНИЯ ЧИСЕЛ**

| Современная | Египетская (иероглифич.) | Египетская (иератическая) | Вавилонская | Греческая (аттическая) | Греческая (ионическая) | Римская | Древнеевр. | Индейцев майя | Древнекитайская (палочк.) | Древнекит. (иероглифическая) | Индийск. (деванагари) | Арабская (алфавит) | Арабская (современная) | Арабская (гобари) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 𓏺 | 𓏺 | 𒁹 | 𐅁 | Α | I | א | • | 一 | 一 | १ | ١ | ١ | ۱ |
| 2 | 𓏺𓏺 | 𓏺𓏺 | 𒁹𒁹 | 𐅂 | Β | II | ב | •• | 二 | 二 | २ | ٢ | ٢ | ۲ |
| 3 | 𓏺𓏺𓏺 | 𓏺𓏺𓏺 | 𒁹𒁹𒁹 | 𐅃 | Γ | III | ג | ••• | 三 | 三 | ३ | ٣ | ٣ | ۳ |
| 4 | 𓏺𓏺𓏺𓏺 | 𓏺𓏺𓏺𓏺 | 𒁹𒁹𒁹𒁹 | 𐅄 | Δ | IIII | ד | •••• | 四 | 四 | ४ | ٤ | ٤ | ۴ |
| 5 | 𓏺𓏺 𓏺𓏺𓏺 | 𓏺𓏺 𓏺𓏺𓏺 | 𒁹𒁹 𒁹𒁹𒁹 | 𐅅 | Ε | V | ה | — | 五 | 五 | ५ | ٥ | ٥ | ۵ |
| 6 | 𓏺𓏺𓏺 𓏺𓏺𓏺 | 𓏺𓏺𓏺 𓏺𓏺𓏺 | 𒁹𒁹𒁹 𒁹𒁹𒁹 | 𐅆 | F | VI | ו | •– | 丅 | 六 | ६ | ٦ | ٦ | ۶ |
| 7 | 𓏺𓏺𓏺𓏺 𓏺𓏺𓏺 | 𓏺𓏺𓏺𓏺 𓏺𓏺𓏺 | 𒁹𒁹𒁹𒁹 𒁹𒁹𒁹 | 𐅇 | Ζ | VII | ז | •• | 𝍠 | 七 | ७ | ٧ | ٧ | ۷ |
| 8 | 𓏺𓏺𓏺𓏺 𓏺𓏺𓏺𓏺 | 𓏺𓏺𓏺𓏺 𓏺𓏺𓏺𓏺 | 𒁹𒁹𒁹𒁹 𒁹𒁹𒁹𒁹 | 𐅈 | Η | VIII | ח | ••• | 𝍡 | 八 | ८ | ٨ | ٨ | ۸ |
| 9 | 𓏺𓏺𓏺𓏺 𓏺𓏺𓏺𓏺𓏺 | 𓏺𓏺𓏺𓏺 𓏺𓏺𓏺𓏺𓏺 | 𒁹𒁹𒁹𒁹 𒁹𒁹𒁹𒁹𒁹 | 𐅉 | Θ | IX | ט | •••• | 𝍢 | 九 | ९ | ٩ | ٩ | ۹ |
| 10 | ∩ | 𓎆 | 𒌋 | 𐅊 | Ι | X | י | = | — | 十 | १० | ١٠ | ١٠ | ۱۰ |
| 20 | ∩∩ | 𓎆𓎆 | 𒌋𒌋 | 𐅋 | Κ | XX | כ | :– | 二 | 二十 | २० | ٢٠ | ٢٠ | ۲۰ |
| 30 | ∩∩∩ | 𓎆𓎆𓎆 | 𒌋𒌋𒌋 | 𐅌 | Λ | XXX | ל | ≡ | 三 | 三十 | ३० | ٣٠ | ٣٠ | ۳۰ |
| 40 | ∩∩∩∩ | 𓎆𓎆𓎆𓎆 | 𒌋𒌋𒌋𒌋 | 𐅍 | Μ | XL | מ | :: | 𝍣 | 四十 | ४० | ٤٠ | ٤٠ | ۴۰ |
| 50 | ∩∩∩∩∩ | 𓎆𓎆𓎆𓎆𓎆 | 𒌋𒌋𒌋𒌋𒌋 | 𐅎 | Ν | L | נ | :– | 𝍤 | 五十 | ५० | ٥٠ | ٥٠ | ۵۰ |
| 60 | ∩∩∩∩∩∩ | 𓎆𓎆𓎆𓎆𓎆𓎆 | 𒌋𒌋𒌋𒌋𒌋𒌋 | 𐅏 | Ξ | LX | ס | ::: | ⊥ | 六十 | ६० | ٦٠ | ٦٠ | ۶۰ |
| 70 | ∩∩∩∩∩∩∩ | 𓎆𓎆𓎆𓎆𓎆𓎆𓎆 | 𒌋𒌋𒌋𒌋𒌋𒌋𒌋 | 𐅐 | Ο | LXX | ע | :::– | 𝍥 | 七十 | ७० | ٧٠ | ٧٠ | ۷۰ |
| 80 | ∩∩∩∩∩∩∩∩ | 𓎆𓎆𓎆𓎆𓎆𓎆𓎆𓎆 | 𒌋𒌋𒌋𒌋𒌋𒌋𒌋𒌋 | 𐅑 | Π | LXXX | פ | :::: | 𝍦 | 八十 | ८० | ٨٠ | ٨٠ | ۸۰ |
| 90 | ∩∩∩∩∩∩∩∩∩ | 𓎆𓎆𓎆𓎆𓎆𓎆𓎆𓎆𓎆 | 𒌋𒌋𒌋𒌋𒌋𒌋𒌋𒌋𒌋 | 𐅒 | Ϙ | XC | צ | ::::– | 𝍧 | 九十 | ९० | ٩٠ | ٩٠ | ۹۰ |
| 100 | 𓍢 | 𓎊 | 𒐕 | Η | Ρ | C | ק | ‖ | 丨 | 百 | १०० | ١٠٠ | ١٠٠ | ۱۰۰ |
| 200 | 𓍢𓍢 | 𓎊𓎊 | 𒐕𒐕 | ΗΗ | Σ | CC | ר | ‖‖ | 𝍨 | 二百 | २०० | ٢٠٠ | ٢٠٠ | ۲۰۰ |
| 300 | 𓍢𓍢𓍢 | 𓎊𓎊𓎊 | 𒐕𒐕𒐕 | ΗΗΗ | Τ | CCC | ש | ‖‖‖ | 𝍩 | 三百 | ३०० | ٣٠٠ | ٣٠٠ | ۳۰۰ |
| 400 | 𓍢𓍢𓍢𓍢 | 𓎊𓎊𓎊𓎊 | 𒐕𒐕𒐕𒐕 | ΗΗΗΗ | Υ | CD | ת | :: | 𝍪 | 四百 | ४०० | ٤٠٠ | ٤٠٠ | ۴۰۰ |
| 500 | 𓍢𓍢𓍢𓍢𓍢 | 𓎊𓎊𓎊𓎊𓎊 | 𒐕𒐕𒐕𒐕𒐕 | Ρ | Φ | D | ך | ‖‖‖‖ | 𝍫 | 五百 | ५०० | ٥٠٠ | ٥٠٠ | ۵۰۰ |
| 600 | 𓍢𓍢𓍢𓍢𓍢𓍢 | 𓎊𓎊𓎊𓎊𓎊𓎊 | 𒐕𒐕𒐕𒐕𒐕𒐕 | ΡΗ | Χ | DC | ם | ‖‖‖‖– | 丅 | 六百 | ६०० | ٦٠٠ | ٦٠٠ | ۶۰۰ |
| 700 | 𓍢𓍢𓍢𓍢𓍢𓍢𓍢 | 𓎊𓎊𓎊𓎊𓎊𓎊𓎊 | 𒐕𒐕𒐕𒐕𒐕𒐕𒐕 | ΡΗΗ | Ψ | DCC | ן | ‖‖‖‖‖ | 𝍬 | 七百 | ७०० | ٧٠٠ | ٧٠٠ | ۷۰۰ |
| 800 | 𓍢𓍢𓍢𓍢𓍢𓍢𓍢𓍢 | 𓎊𓎊𓎊𓎊𓎊𓎊𓎊𓎊 | 𒐕𒐕𒐕𒐕𒐕𒐕𒐕𒐕 | ΡΗΗΗ | Ω | DCCC | ף | ‖‖‖‖‖– | 𝍭 | 八百 | ८०० | ٨٠٠ | ٨٠٠ | ۸۰۰ |
| 900 | 𓍢𓍢𓍢𓍢𓍢𓍢𓍢𓍢𓍢 | 𓎊𓎊𓎊𓎊𓎊𓎊𓎊𓎊𓎊 | 𒐕𒐕𒐕𒐕𒐕𒐕𒐕𒐕𒐕 | ΡΗΗΗΗ | Ϡ | CM | ץ | ‖‖‖‖‖‖ | 𝍮 | 九百 | ९०० | ٩٠٠ | ٩٠٠ | ۹۰۰ |

Рисунок 1 – Знаки, используемые для записи чисел в различных системах счисления


Рисунок 2 – Унарная (единичная) система счисления

## 1.1 Непозиционный системы счисления

В древнейшее время в Греции была распространена т. н. *аттическая* нумерация. Числа 1, 2, 3, 4 обозначались черточками Ⅰ , Ⅱ , Ⅲ , Ⅳ . Число 5 записывалось знаком Γ (древнее начертание буквы «пи», с которой начинается слово «пенте» – пять); числа 6, 7, 8, 9 обозначались ΓⅠ , ΓⅡ , ΓⅢ , ΓⅣ . Число 10 обозначалось Δ (начальной буквой слова «дека» – десять). Числа 100, 1000 и 10000 обозначались Η , Χ , Μ . Числа 50, 500, 5000 обозначались комбинациями знаков 5 и 10, 5 и 100, 5 и 1000. Запись чисел в аттической системе счисления представлена на рис. 3.


Рисунок 3 – Непозиционные системы счисления

В III веке до н.э. аттическая нумерация была вытеснена так называемой *ионийской* системой. В ней числа 1 – 9 обозначались первыми девятью буквами алфавита; числа 10, 20, 30, … , 90 – следующими девятью буквами; числа 100, 200, … , 900 – последними девятью буквами.

Южные и восточные славянские народы для записи чисел пользовались алфавитной нумерацией. У одних славянских народов числовые значения букв установились в порядке славянского алфавита (рис. 4), у других же (в том числе у русских) роль цифр играли не все буквы, а только те, которые имеются в греческом алфавите. Над буквой, обозначавшей цифру, ставился специальный значок: ҃ («титло»).

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1 | А | аз | 10 | І | и* | 100 | Р | рцы |
| 2 | В | веди | 20 | К | како | 200 | С | слово |
| 3 | Г | глаголь | 30 | Л | люди | 300 | Т | твердо |
| 4 | Д | добро | 40 | М | мыслете | 400 | У | ук** |
| 5 | Е | есть** | 50 | Н | наш** | 500 | Ф | ферт |
| 6 | S | зело* | 60 | Ѯ | кси** | 600 | Х | хер |
| 7 | З | земля** | 70 | О | он | 700 | Ѱ | пси* |
| 8 | И | иже** | 80 | П | покой | 800 | Ѡ | омега* |
| 9 | Ѳ | фита* | 90 | Ч | червь | 900 | Ц | цы |

* Буквы, исключенные впоследствии из русского алфавита
** Буквы, у которых изменилось начертание

Рисунок 4 – Система счистления, применявшаяся у славян

*Непозиционные системы* счисления возникли раньше позиционных. Они характеризуются тем, что в них символы, обозначающие то или иное число, не меняют своего значения в зависимости от своего местоположения в записи этого числа. Примерами таких систем являются:

- Единичная система счисления,
- Пятеричная система счисления (Счёт на пятки́),
- Древнеегипетская система счисления,
- Вавилонская система счисления,
- Алфавитные системы счисления,
- Еврейская система счисления,
- Греческая система счисления,
- Римская система счисления,
- Система счисления майя,
- Кипу инков и др.

Наиболее известной непозиционный системой счисления является римская система счисления. В ней для записи чисел используются буквы латинского алфавита. Значения основных цифр римской системы приведены ниже:

I — единица, V — пять, X — десять, L — пятьдесят,
C — сто, D — пятьсот, M — тысяча.

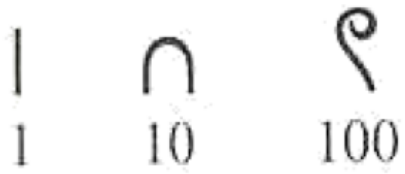
Еще одна особенность – чтобы выразить число и не использовать сотни символов, применяется правило: каждый меньший знак, поставленный справа от большего, прибавляется к его значению, а каждый меньший знак, поставленный слева от большего, вычитается из него.

*Пример*. Написать 475 римскими знаками можно так CCCCXXXXXXXIIIII, но это нерационально. Если отнимать или прибавлять цифры, получится меньшее количество символов – CDLXXV.

**Некоторые числа в римской системе счисления**

| Число в римской системе | Значение в десятичной системе |
|---|---|
| III | 1+1+1=3 |
| IV | 5–1=4 |
| XII | 10+1+1=12 |
| XLV | –10+50+5=45 |
| CDXVIII | –100+500+10+5+1+1+1=418 |
| MMDXCVII | 1000+1000+500–10+100+5+1+1=2597 |

*Древнеегипетская десятичная система счисления.* В Древнем Египте использовали свои символы (цифры) для обозначения чисел 1, 10, 102, 103, 104, 105, 106, 107. Вот некоторые из них:




В Египте – решили группировать по 10, оставив без изменений цифру «1». Здесь, число 10 называется основанием десятичной системы счисления, а все символы – представление числа 10 в определенной степени.

Числа в древнеегипетской системе счисления записывали, в виде комбинаций таких символов, и все они повторялись не больше 9 раз. Результатом было сумма элементов числа. Для примера посмотрите на запись числа 345:

*Вавилонская шестидесятеричная система счисления.* В вавилонской системе счисления использовали только 2 символа (рис. 5): «прямой» клин – для единиц и «лежащий» – для десятков. Для определения значения числа нужно изображение

числа разбить на разряды справа налево. Новый разряд начинается с появления прямого клина после лежачего.


Рисунок 5 – Вавилонская шестидесятеричная система счисления

Числа менее 60 обозначались с помощью двух знаков: прямой клин служил для обозначения единиц, лежачий клин – для обозначения десятков.

Число 60 и все его степени так же обозначаются прямым клином, что и «1». Поэтому вавилонская система счисления получила название шестидесятеричной системы счисления.

Все числа от 1 до 59 вавилоняне записывали в десятичной непозиционной системе, а значения больше 59 — в позиционной с основанием 60. Например, число 92:

Запись числа была не конкретной, так как не было цифры, которая обозначала бы нуль. Представление числа 92 могло обозначать не только 92=60+32, но и, например, 3632=3600+32. Для определения абсолютного значения числа они ввели новый символ для обозначения пропущенного шестидесятеричного разряда, что соответствует появлению цифры 0 в записи десятичного числа:

Значит, число 3632 записывают следующим образом:

Шестидесятеричная вавилонская система – первая система счисления, которая частично основана на позиционном принципе. Эту систему счисления используют и сейчас, например, для определения времени – час состоит из 60 минут, а минута из 60 секунд.

Непозиционные системы счисления имеют *два существенных недостатка:*

- с увеличением изображаемых чисел требуется неограниченное число новых символов;
- процедура выполнения арифметических операций в таких системах счисления чрезвычайно сложна.

Поэтому в настоящее время непозиционные системы счисления практически не используются.

## 1.2 Позиционные системы счисления

Система счисления называется ***позиционной***, если количественный эквивалент цифры зависит от её положения (позиции) в записи числа.

Позиционные системы счисления характеризуются следующими понятиями:

• Для записи любого числа используется ограниченный набор символов. Число используемых символов называется основанием позиционной системы счисления.

• Устанавливается взаимно-однозначное соответствие между набором цифр и числами натурального ряда 0, 1, ..., 0, 1, 2, 3,..., $p-1$, где $p$ – основание системы счисления. Таким образом, численный эквивалент любой цифры меньше основания системы счисления.

• Место каждой цифры в числе называется позицией (отсюда, собственно, название таких систем – позиционные).

• Номер позиции цифры в числе называется разрядом. Нумерация разрядов начинается с нуля и выполняется справа налево. Разряд 0 называется младшим разрядом.

• Каждой цифре, в зависимости от ее позиции, ставится в соответствие количественный эквивалент, определяемый по формуле

$$\alpha_k = a_k p^k, \tag{1.1}$$

где $\alpha_k$ – количественный эквивалент цифры, находящейся в позиции $k$;

$a_k$ – численный эквивалент цифры, находящейся в разряде $k$;

$p$ – основание системы счисления;

$k$ – номер позиции цифры (ее разряд).

Само значение числа (его количественный эквивалент) определяется как сумма вычисленных по формуле (1.1) количественных эквивалентов всех цифр, входящих в запись числа.

Для выполнения операции сложения каждой паре чисел, каждое из которых соотносится с какой-либо одной цифрой, ставится в соответствие число, являющееся результатом их сложения. Аналогично, для выполнения операции умножения каждой такой паре чисел ставится в соответствие число, являющееся результатом их умножения. Эти соответствия оформляются в виде таблицы сложения и таблицы умножения.

Таким образом, любое целое положительное число может быть представлено в виде

$$a_{n-1}a_{n-2}\ldots a_1a_0 = a_{n-1}p^{n-1} + a_{n-2}p^{n-2} + \ldots + a_1p^1 + a_0p^0 \qquad (1.2)$$

где $a_i$ – цифра данной системы счисления ($0 \le a_i \le p$);

$n$ – число разрядов при написании числа;

$p$ – основание системы счисления (некоторое положительное целое число).

В качестве основания системы счисления может быть использовано любое натуральное число $p > 1$. При заданном основании системы счисления $p$ каждому натуральному числу соответствует единственное представление вида (1.2) и каждому представлению вида (1.2) соответствует единственное натуральное число. Естественно, при этом лидирующие нули не учитываются, например 0345 и 345 – это эквивалентные записи одного и того же числа.

Проиллюстрируем сказанное на привычной нам десятичной системе.

Набор цифр для десятичной системы счисления: {0, 1, 2, 3, 4, 5, 6, 7, 8, 9}. Основание системы $p = 10$. Любое число в десятичной системе согласно формуле (1.2) представляется в виде

$$A_{10} = a_{n-1}a_{n-2}\ldots a_1a_0 = a_{n-1}\cdot 10^{n-1} + a_{n-2}\cdot 10^{n-2} + \ldots + a_1\cdot 10^1 + a_0\cdot 10^0,$$

где каждое $a_i$ — одна из цифр множества {1,2,3,4,5,6,7,8,9}.

*Пример*

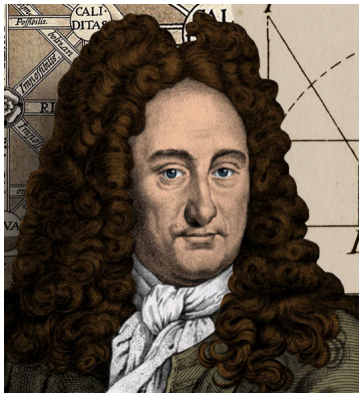
$$473 = 4\cdot 10^2 + 7\cdot 10^1 + 3\cdot 10^0,$$

$$1029 = 1\cdot 10^3 + 0\cdot 10^2 + 2\cdot 10^1 + 9\cdot 10^0.$$

Наиболее применимыми являются: двоичная, десятичная и шестнадцатеричная системы счисления

- компьютер работает только с двоичной информацией;
- человек производит вычисления, используя десятичную систему;
- двоичная информация плохо воспринимается человеком, для ее интерпретации удобнее использовать шестнадцатеричную систему.

## 1.3 Двоичная система счисления



Готфрид Вильгельм Лейбниц (1646-1716)

Систему, на которой основывается работа компьютеров, придумал гениальный немецкий ученый Г.В. Лейбниц (еще до 19 века!). Набор цифр для двоичной системы счисления: {0, 1}. Основание системы $p = 2$. Любое число в двоичной системе представляет собой последовательность нулей и единиц. Для того чтобы подчеркнуть, что это именно двоичная запись, в конце числа можно (но не обязательно) поставить нижний индекс 2 или символ b (от английского binary — «двоичный»). Последнее обозначение является обязательным при задании двоичных констант на языке Assembler. Например

$$5 = 101_2 = 101b$$
$$1025 = 10000000001_2 = 10000000001b$$

Согласно формуле (1.2) число в двоичной системе представляется в виде

$$A_2 = a_{n-1}a_{n-2}\ldots a_1a_0 = a_{n-1}\cdot 2^{n-1} + a_{n-2}\cdot 2^{n-2} + \ldots + a_1\cdot 2^1 + a_0\cdot 2^0, \quad (1.3)$$

где каждое $a_i$ — одна из цифр 0 или 1.

Например,

$$5_{10} = 101_2 = 1\cdot 2^2 + 0\cdot 2^1 + 1\cdot 2^0 = 1\cdot 4 + 1,$$

$$19_{10} = 10011_2 = 1\cdot 2^4 + 0\cdot 2^3 + 0\cdot 2^2 + 1\cdot 2^1 + 1\cdot 2^0 = 2^4 + 2^1 + 2^0,$$

$$0{,}8125_{10} = 0{,}1101_2 = 1\cdot 2^{-1} + 1\cdot 2^{-2} + 0\cdot 2^{-3} + 1\cdot 2^{-4}.$$

В формуле (1.3) разложение двоичного числа по степеням «двойки» выполнено в десятичной системе. То же самое разложение можно записать, используя только цифры двоичной системы

$$A_2 = a_{n-1}a_{n-2}\ldots a_1a_0 = (a_{n-1}\cdot 10^{n-1} + a_{n-2}\cdot 10^{n-2} + \ldots + a_1\cdot 10^1 + a_0\cdot 10^0)_2, \quad (1.4)$$

Те же самые числа при использовании выражения (1.4) будут выглядеть следующим образом:

$$5 = 101_2 = 1\cdot 10^2 + 0\cdot 10^1 + 1\cdot 10^0.$$

Алгоритм перевода чисел из десятичной в двоичную систему представлен на рис. 6:

1. Деление на основание системы до тех пор, пока не останется в остатке значение меньше значения основы.

2. Запись остатков, от последнего к первому.
3. Первый ноль можно не писать.

Этот порядок действия позволят переводить в любую позиционную систему счисления.

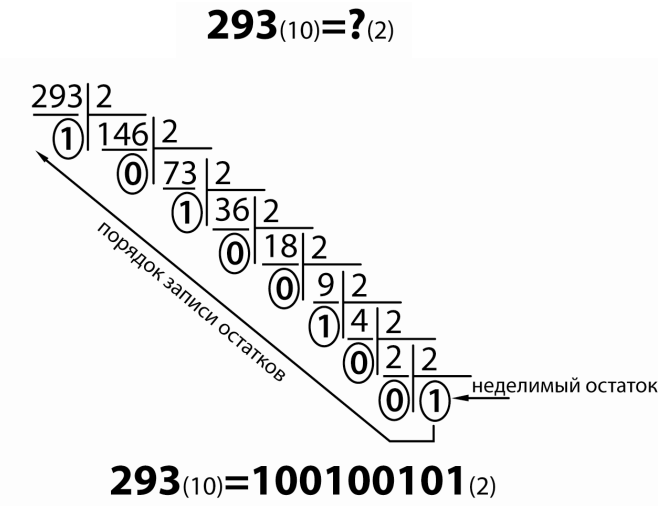



Рисунок 6 – Пример перевода числа из десятичной в двоичную систему

Обратный алгоритм перевода из двоичной в десятичную систему счисления:

Записать число развернуто, то есть, сколько сотен, десятков и единиц в нем, но учитывая основу – 2 (см. табл. 1.1).

Таблица 1.1 – Алгоритм перевода из двоичной в десятичную систему счисления

| Разряды | 3 | 2 | 1 | 0 | -1 | | |
|---|---|---|---|---|---|---|---|
| Число | 1 | 0 | 1 | 1, | $1_2$ | = | $1 \cdot 2^3 + 0 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^0 + 1 \cdot 2^{-1} = 11{,}5_{10}$ |

| Разряды | 5 | 4 | 3 | 2 | 1 | 0 | | |
|---|---|---|---|---|---|---|---|---|
| Число | 1 | 0 | 1 | 1 | 0 | $1_2$ | = | $1 \cdot 2^5 + 0 \cdot 2^4 + 1 \cdot 2^3 + 1 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0 = 45_{10}$ |

Для перевода чисел в десятичную систему счисления удобно пользоваться табл. 1.2.

Таблица 1.2 – Степеней чисел от 2 до 9

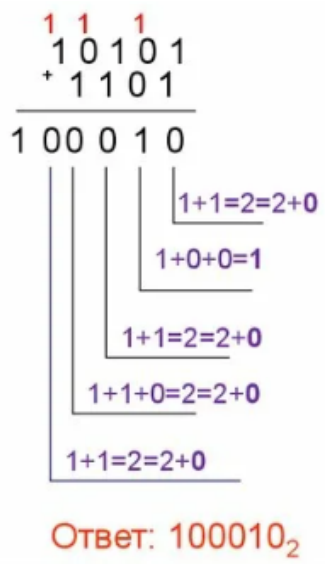
| n | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|
| $2^n$ | 4 | 8 | 16 | 32 | 64 | 128 | 256 | 512 | 1024 |
| $3^n$ | 9 | 27 | 81 | 243 | 729 | 2187 | 6561 | 19683 | 59049 |
| $4^n$ | 16 | 64 | 256 | 1024 | 4096 | 16384 | 65536 | 262144 | |
| $5^n$ | 25 | 125 | 625 | 3125 | 15625 | 78125 | 390625 | | |
| $6^n$ | 36 | 216 | 1296 | 7776 | 46656 | 279936 | | | |
| $7^n$ | 49 | 242 | 2401 | 16807 | 117649 | | | | |
| $8^n$ | 64 | 512 | 4096 | 32768 | | | | | |
| $9^n$ | 81 | 729 | 6561 | 59049 | | | | | |

Таблицы умножения и сложения чисел и арифметические операции в двоичной системе счисления выполняются подобно тому, как это делается в десятичной системе, с той лишь разницей, что при этом используются свои таблицы умножения и сложения (табл. 1.3).

Таблица 1.3 – Сложение и умножение двоичных чисел

| Таблица сложения | | | Таблица умножения | | |
|---|---|---|---|---|---|
| | **0** | **1** | | **0** | **1** |
| **0** | 0 | 1 | **0** | 0 | 0 |
| **1** | 1 | 10 | **1** | 0 | 1 |

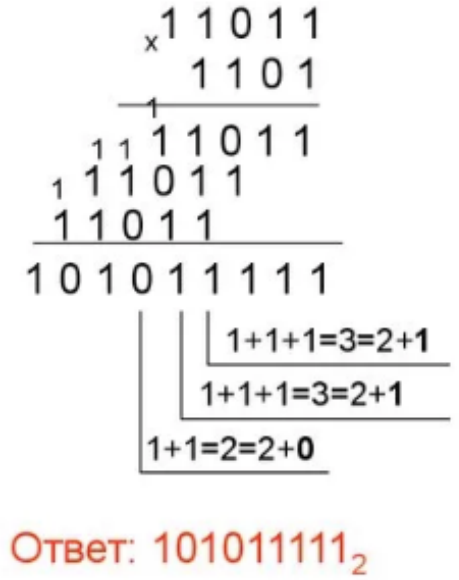
*Пример*: сложение двоичных чисел




$$\begin{array}{rcccccccc} & 1 & 0 & 1 & 1 & 0 & 1 & = & 45 \\ + & 1 & 1 & 1 & 1 & 1 & & = & 31 \\ & - & - & - & - & - & & & - \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & = & 76 \end{array}$$

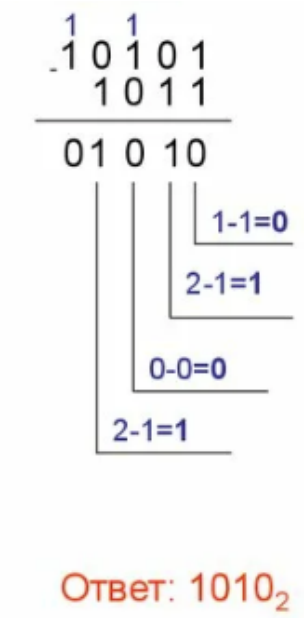
$$\begin{array}{rcccccc} & 1 & 0 & 1 & 1 & 0 & = & 22 \\ + & 1 & 1 & 0 & 1 & 1 & = & 27 \\ & - & - & - & - & - & & - \\ 1 & 1 & 0 & 0 & 0 & 1 & = & 49 \end{array}$$

*Пример*: умножение двоичных чисел




$$\begin{array}{rcccccccc} & & 1 & 1 & 1 & 0 & 1 & 1 & = & 59 \\ X & & & & 1 & 1 & 0 & 1 & = & 13 \\ \hline & & 1 & 1 & 1 & 0 & 1 & 1 & & \\ + & 1 & 1 & 1 & 0 & 1 & 1 & & & \\ + & 1 & 1 & 1 & 0 & 1 & 1 & & & \\ \hline 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & = & 383 \end{array}$$

*Пример*: вычитание двоичных чисел




1 0 0 0 1 0 1 = 69
- 1 1 0 1 1 = 27
- - - - - -
1 0 1 0 1 0 = 42

1 0 1 1 0 0 1 1 1 = 359
- 1 0 0 1 1 0 1 = 77
- - - - - - - -
1 0 0 0 1 1 0 1 0 = 282

Недостатком двоичной системы счисления является необходимость использования большого числа символов при записи даже сравнительно небольших чисел, что существенно затрудняет их восприятие человеком. Например, число 1567 записывается в двоичном виде как 11000011111, а число 8763 – как 10001000111011. Поэтому для интерпретации двоичной информации используется шестнадцатеричная система, запись чисел в которой значительно компактнее.

## 1.4 Шестнадцатеричная система счисления

Основание системы: $p = 16$. Широко используется в низкоуровневом программировании и компьютерной документации, поскольку в современных компьютерах минимально адресуемой единицей памяти является 8-битный байт, значения которого удобно записывать двумя шестнадцатеричными цифрами. Такое использование началось с системы IBM/360, где вся документация использовала шестнадцатеричную систему, в то время как в документации других компьютерных систем того времени (даже с 8-битными символами, как, например, PDP-11 или БЭСМ-6) использовали восьмеричную систему.

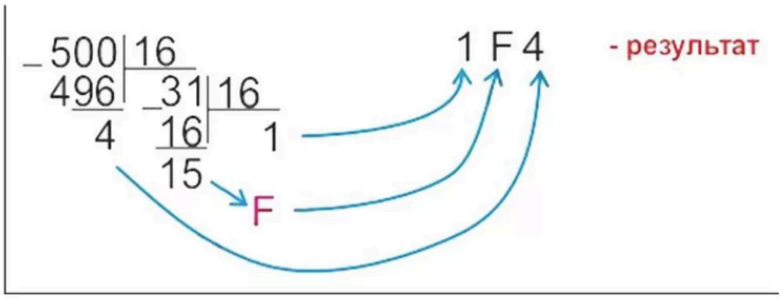
Алфавит: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F.

Здесь только десять цифр из шестнадцати имеют общепринятое обозначение 0,…, 9. Для записи цифр с десятичными количественными эквивалентами 10, 11, 12, 13, 14, 15 обычно используются первые пять букв латинского алфавита (табл. 1.4).

Таблица 1.4 – Соответствия между десятеричной и шестнадцатеричной системами счисления

| **Число десятеричной системы счисления** | **Цифра шестнадцатеричной системы счисления** |
|---|---|
| $10_{10}$ | $A_{16}$ |
| $11_{10}$ | $B_{16}$ |
| $12_{10}$ | $C_{16}$ |
| $13_{10}$ | $D_{16}$ |
| $14_{10}$ | $E_{16}$ |
| $15_{10}$ | $F_{16}$ |

*Пример*. Перевести число $500_{10}$ в шестнадцатеричную систему счисления.




Ответ: 500(10)=1F4(16).

Таблицы сложения и умножения чисел и арифметические операции в шестнадцатиричной системе представлены в таблицах 1.5-1.6 :,

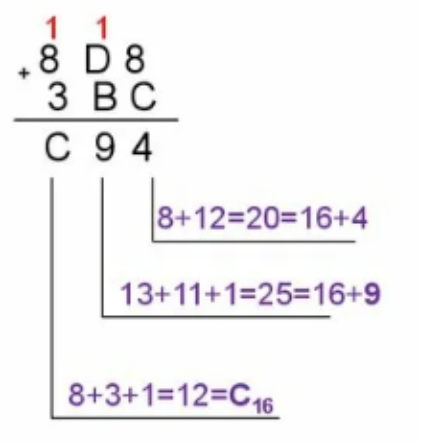
Таблица 1.5 – Сложение шестнадцатеричных чисел

| | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | ***A*** | ***B*** | ***C*** | ***D*** | ***E*** | ***F*** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | *A* | *B* | *C* | *D* | *E* | *F* |
| **1** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | *A* | *B* | *C* | *D* | *E* | *F* | 10 |
| **2** | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | *A* | *B* | *C* | *D* | *E* | *F* | 10 | 11 |
| **3** | 3 | 4 | 5 | 6 | 7 | 8 | 9 | *A* | *B* | *C* | *D* | *E* | *F* | 10 | 11 | 12 |
| **4** | 4 | 5 | 6 | 7 | 8 | 9 | *A* | *B* | *C* | *D* | *E* | *F* | 10 | 11 | 12 | 13 |
| **5** | 5 | 6 | 7 | 8 | 9 | *A* | *B* | *C* | *D* | *E* | *F* | 10 | 11 | 12 | 13 | 14 |
| **6** | 6 | 7 | 8 | 9 | *A* | *B* | *C* | *D* | *E* | *F* | 10 | 11 | 12 | 13 | 14 | 15 |
| **7** | 7 | 8 | 9 | *A* | *B* | *C* | *D* | *E* | *F* | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| **8** | 8 | 9 | *A* | *B* | *C* | *D* | *E* | *F* | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| **9** | 9 | *A* | *B* | *C* | *D* | *E* | *F* | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| ***A*** | *A* | *B* | *C* | *D* | *E* | *F* | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
| ***B*** | *B* | *C* | *D* | *E* | *F* | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 1*A* |
| ***C*** | *C* | *D* | *E* | *F* | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 1*A* | 1*B* |
| ***D*** | *D* | *E* | *F* | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 1*A* | 1*B* | 1*C* |
| ***E*** | *E* | *F* | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 1*A* | 1*B* | 1*C* | 1*D* |
| ***F*** | *F* | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 1*A* | 1*B* | 1*C* | 1*D* | 1*E* |

Таблица 1.6 – Умножение шестнадцатеричных чисел

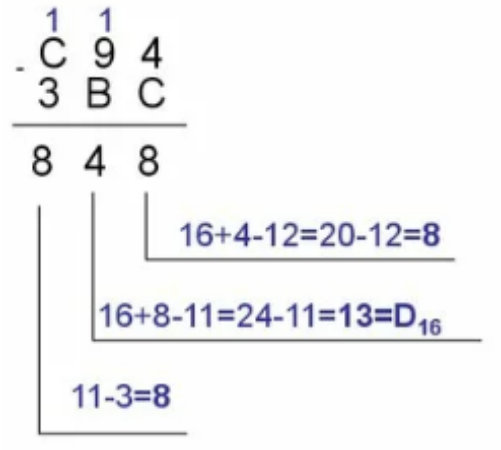
| | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | ***A*** | ***B*** | ***C*** | ***D*** | ***E*** | ***F*** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **1** | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | *A* | *B* | *C* | *D* | *E* | *F* |
| **2** | 0 | 2 | 4 | 6 | 8 | *A* | *C* | *E* | 10 | 12 | 14 | 16 | 18 | 1*A* | 1*C* | 1*E* |
| **3** | 0 | 3 | 6 | 9 | *C* | *F* | 12 | 15 | 18 | 1*B* | 1*E* | 21 | 24 | 27 | 2*A* | 2*D* |
| **4** | 0 | 4 | 8 | *C* | 10 | 14 | 18 | 1*C* | 20 | 24 | 28 | 2*C* | 30 | 34 | 38 | 3*C* |
| **5** | 0 | 5 | *A* | *F* | 14 | 19 | 1*E* | 23 | 28 | 2*D* | 32 | 37 | 3*C* | 41 | 46 | 4*B* |
| **6** | 0 | 6 | *C* | 12 | 18 | 1*E* | 24 | 2*A* | 30 | 36 | 3*C* | 42 | 48 | 4*E* | 54 | 5*A* |
| | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | ***A*** | ***B*** | ***C*** | ***D*** | ***E*** | ***F*** |
| **7** | 0 | 7 | *E* | 15 | 1*C* | 23 | 2*A* | 31 | 38 | 3*F* | 46 | 4*D* | 54 | 5*B* | 62 | 69 |
| **8** | 0 | 8 | 10 | 18 | 20 | 28 | 30 | 38 | 40 | 48 | 50 | 58 | 60 | 68 | 70 | 78 |
| **9** | 0 | 9 | 12 | 1*B* | 24 | 2*D* | 36 | 3*F* | 48 | 51 | 5*A* | 63 | 6*C* | 75 | 7*E* | 87 |
| ***A*** | 0 | *A* | 14 | 1*E* | 28 | 32 | 3*C* | 46 | 50 | 5*A* | 64 | 6*E* | 78 | 82 | 8*C* | 96 |
| ***B*** | 0 | *B* | 16 | 21 | 2*C* | 37 | 42 | 4*D* | 58 | 63 | 6*E* | 79 | 84 | 8*F* | 9*A* | *A*5 |
| ***C*** | 0 | *C* | 18 | 24 | 30 | 3*C* | 48 | 54 | 60 | 6*C* | 78 | 84 | 90 | 9*C* | *A*8 | *B*4 |
| ***D*** | 0 | *D* | 1*A* | 27 | 34 | 41 | 4*E* | 5*B* | 68 | 75 | 82 | 8*F* | 9*C* | *A*9 | *B*6 | *C*3 |
| ***E*** | 0 | *E* | 1*C* | 2*A* | 38 | 46 | 54 | 62 | 70 | 7*E* | 8*C* | 9*A* | *A*8 | *B*6 | *C*4 | *D*2 |
| ***F*** | 0 | *F* | 1*E* | 2*D* | 3*C* | 4*B* | 5*A* | 69 | 78 | 87 | 96 | *A*5 | *B*4 | *C*3 | *D*2 | *E*1 |

*Пример*: сложение в шестнадцатеричный системе счисления




| | | | | | | |
|---|---|---|---|---|---|---|
| | 1 | C | 5 | 2 | = | 7250 |
| + | | 8 | 9 | 1 | = | 2193 |
| | - | - | - | - | | - |
| | 2 | 4 | E | 3 | = | 9443 |

Пример: вычитание в шестнадцатеричный системе счисления




| | | | | | |
|---|---|---|---|---|---|
| | 1 | B | A | = | 442 |
| - | A | 5 | 9 | = | 2649 |
| | - | - | - | | - |
| - | 8 | 9 | F | = | -2207 |

Перевод из двоичной системы счисления в шестнаддатеричную и обратно осуществляется через разбиение на группы по четыре цифры (тетрады) см. табл. 1.7

Таблица 1.7 – Соответствия между десятеричной и шестнадцатеричной системами счисления

| Двоичные тетрады | 0000 | 0001 | 0010 | 0011 | 0100 | 0101 | 0110 | 0111 |
|---|---|---|---|---|---|---|---|---|
| Шестнадцате-ричные цифры | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Двоичные тетрады | 1000 | 1001 | 1010 | 1011 | 1100 | 1101 | 1110 | 1111 |
| Шестнадцате-ричные цифры | 8 | 9 | A | B | C | D | E | F |

## 1.5 Восьмеричная система счисления

***Восьмеричной системой счисления*** называется позиционная система счисления с основанием 8. Для записи чисел в восьмеричной системе счисления используются цифры: 0, 1,2, 3, 4, 5, 6, 7.

На основании формулы (1.2) для целого восьмеричного числа можно записать:

$$A_8 = a_{n-1}a_{n-2}\ldots a_1a_0 = (a_{n-1}\cdot 8^{n-1} + a_{n-2}\cdot 8^{n-2} + \ldots + a_1\cdot 8^1 + a_0\cdot 8^0)_8, \quad (1.5)$$

Те же самые числа при использовании выражения (1.5) будут выглядеть следующим образом:

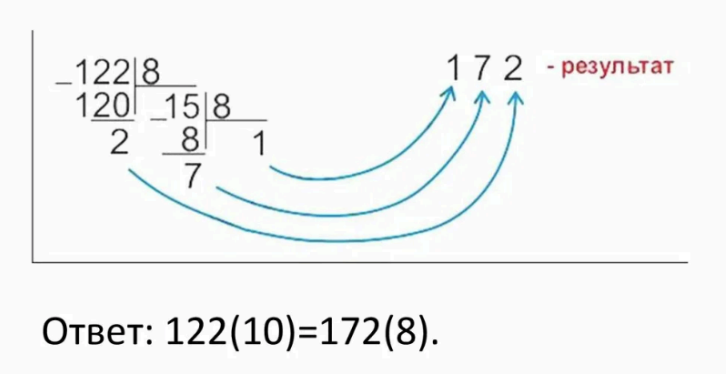
$$1063_8 = 1\cdot 8^3 + 0\cdot 8^2 + 6\cdot 8^1 + 3\cdot 8^0 = 563_{10},$$

$$0{,}04_8 = 4\cdot 8^{-2} = 0{,}0625_{10}.$$

Таким образом, для перевода целого восьмеричного числа в десятичную систему счисления следует перейти к его развёрнутой записи и вычислить значение получившегося выражения.

Для перевода целого десятичного числа в восьмеричную систему счисления следует последовательно выполнять деление данного числа и получаемых целых частных на 8 до тех пор, пока не получим частное, равное нулю. Исходное число в новой системе счисления составляется последовательной записью полученных остатков, начиная с последнего.

*Пример*. Перевести число $122_{10}$ в восьмеричную систему счисления.




Ответ: 122(10)=172(8).

Арифметические действия в восьмеричной системе счисления осуществляются согласно таблицам представленным ниже

Таблица сложения

| + | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 10 |
| 2 | 2 | 3 | 4 | 5 | 6 | 7 | 10 | 11 |
| 3 | 3 | 4 | 5 | 6 | 7 | 10 | 11 | 12 |
| 4 | 4 | 5 | 6 | 7 | 10 | 11 | 12 | 13 |
| 5 | 5 | 6 | 7 | 10 | 11 | 12 | 13 | 14 |
| 6 | 6 | 7 | 10 | 11 | 12 | 13 | 14 | 15 |
| 7 | 7 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |

Таблица умножения

| x | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 2 | 0 | 2 | 4 | 6 | 10 | 12 | 14 | 16 |
| 3 | 0 | 3 | 6 | 11 | 14 | 17 | 22 | 25 |
| 4 | 0 | 4 | 10 | 14 | 20 | 24 | 30 | 34 |
| 5 | 0 | 5 | 12 | 17 | 24 | 31 | 36 | 43 |
| 6 | 0 | 6 | 14 | 22 | 30 | 36 | 44 | 52 |
| 7 | 0 | 7 | 16 | 25 | 34 | 43 | 52 | 61 |

Пример применения таблицы сложения в восьмеричной системе счисления представлен ниже

| + | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 10 |
| 2 | 2 | 3 | 4 | 5 | 6 | 7 | 10 | 11 |
| 3 | 3 | 4 | 5 | 6 | 7 | 10 | 11 | 12 |
| 4 | 4 | 5 | 6 | 7 | 10 | 11 | 12 | 13 |
| 5 | 5 | 6 | 7 | 10 | 11 | 12 | 13 | 14 |
| 6 | 6 | 7 | 10 | 11 | 12 | 13 | 14 | 15 |
| 7 | 7 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |

10

6

| | | | |
|---|---|---|---|
| | 1 | | |
| | 4 | 1 | |
| + | | | |
| | 4 | 5 | |
| 1 | 0 | 6 | |

***Замечание***

При вычислении десятичного значения $p$-ичного целого числа по развернутой форме удобно пользоваться *схемой Горнера*, которая позволяет минимизировать арифметические операции и исключить возведение в степень.

Схема Горнера была на самом деле применена англичанином Горнером (а ещё раньше итальянцем Руффини) для вычисления коэффициентов многочлена $p(x + c)$и использовалась для приближённого вычисления корней многочленов. Еще одним применением схемы Горнера является быстрый алгоритм перевода из двоичной системы в десятичную, предложенный Соденом в 1953 году: старшую цифру умножаем на основание, добавляем вторую цифру, результат умножаем на основание, добавляем третью цифру и так до тех пор, пока не прибавим последнюю цифру.

Результатом будет десятичная запись числа. Полученное равенство будет справедливо для любых целых $p$-ичных чисел, а формулу можно записать в общем виде:

$$(a_{n-1}a_{n-2}\dots a_1a_0)_p = (\dots(a_{n-1}\cdot p + a_{n-2)}\cdot p + a_{n-3})\cdot p + \dots) + a_1)\cdot p + a_0\,.$$

*Пример*:

$$10101101_2 = 1\cdot 2^7 + 0\cdot 2^6 + 1\cdot 2^5 + 0\cdot 2^4 + 1\cdot 2^3 + 1\cdot 2^2 + 0\cdot 2^1 + 1\cdot 2^0 =$$
$$= ((((((1\cdot 2 + 0)\cdot 2 + 0)\cdot 2 + 1)\cdot 2 + 0)\cdot 2 + 1 = 173_{10}.$$

# 2. Основные понятия теории информации

## 2.1 Элементы теории вероятностей

Под ***событием*** понимают любой факт, который может произойти в результате какого-либо опыта или эксперимента. Под опытом понимается осуществление определённого комплекса условий. События называются ***совместными,*** если наступление одного из них не исключает наступления другого. В противном случае события называются ***несовместными***. Событие называется ***достоверным***, если оно обязательно произойдет в условиях данного опыта. Событие называется ***невозможным***, если оно не может произойти в условиях данного опыта. Событие называется ***возможным***, или ***случайным***, если в результате опыта оно может появиться, но может и не появиться. События называются ***равновозможными***, если по условиям испытания ни одно из этих событий не является объективно более возможным, чем другие.

Важным понятием является ***полная группа событий***. Несколько событий в данном опыте образуют полную группу, если в результате опыта обязательно появится хотя бы одно из них.

Вероятностью события $X$ называется число, равное отношению числа исходов $m$, благоприятствующих появлению события, к числу всех равновозможных исходов $n$:

$$P(X) = \frac{m}{n}.$$

Вероятность события характеризуется следующими свойствами:

1. $0 \le P(X) \le 1$;
2. $P(X) = 0$ означает, что $X$ - событие невозможное;
3. $P(Y) = 1$ означает, что $Y$ - событие достоверное.

***Теорема сложения вероятностей***. Если два события несовместны, то вероятность наступления каждого из них равна сумме их вероятностей

$$P(X + Y) = P(X) + P(Y).$$

*Следствия:*

1. Для полной группы событий сумма их вероятностей равна 1.
2. Сумма вероятностей двух противоположных событий $X$ и $\overline{X}$ равна 1.

*Условной вероятностью* события $Y$ называется вероятность наступления события $Y$ при условии, что событие $X$ уже наступило. Обозначается $P(Y/X)$ или $Px(Y)$.

***Теорема умножения вероятностей***. Вероятность совместного наступления событий $X$ и $Y$, равна произведению вероятности одного из них на условную вероятность другого

$$P(XY) = P(X)P(Y/X) \text{ или } P(XY) = P(Y)P(X/Y).$$

*Следствие*. Вероятность произведения двух независимых событий равна произведению их вероятностей

$$P(XY) = P(X)P(Y).$$

***Теорема сложения вероятностей для случая, когда события совместны***.

Вероятность наступления хотя бы одного из двух совместных событий равна сумме вероятностей этих событий, минус вероятность их совместного появления,

$$P(X + Y) = P(A) + P(B) - P(AB).$$

Объединение теорем сложения и умножения выражается в *формуле полной вероятности.*

***Теорема.*** Вероятность события $X$, которое может произойти при осуществлении одного из несовместных событий $Y_1, Y_2, Y_3, \ldots Y_n$, образующих полную группу, определяется формулой:

$$P(X) = P(Y_1)P(X/Y_1) + P(Y_2)P(X/Y_2) + \ldots + P(Y_n)P(X/Y_n) = \sum_{i=1}^{n} P(Y_i)P(X/Y_i).$$

События $Y_1, Y_2, Y_3, \ldots Y_n$ называются *гипотезами*.

В случае, если событие $X$, появляющееся совместно с каким-либо из несовместных событий $Y_1, Y_2, Y_3, \ldots Y_n$, образующих полную группу, произошло и требуется произвести оценку вероятностей событий $Y_1, Y_2, Y_3, \ldots Y_n$, применяется ***формула*** Байеса:

$$P(Y_i/X) = \frac{P(Y_i)P(X/Y_i)}{\sum_{i=1}^{n} P(B_i)P(X/Y_i)}.$$

*Пример*. По каналу связи с помехами передается одно из двух сообщений:

1) 11111 с вероятностью равной 0,7;

2) 00000 с вероятностью равной 0,3.

Вероятность правильного приема каждого из символов 0 и 1 равна 0,6. Символы искажаются помехами независимо друг от друга. На выходе канала получают кодовое сообщение 10110.

Определить вероятности передачи первого и второго сообщений.

*Решение*. Пусть событие $X$ состоит в приеме сообщения 10110. Это событие может произойти в совокупности с событием $Y_1$ - передавалось сообщение 11111 и событием $Y_2$ - передавалось сообщение 00000. При этом $P(Y_1) = 0{,}7$ и $P(Y_2) = 0{,}3$.

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

$$P(X/Y_1) = P(1/1)P(0/1)P(1/1)P(1/1)P(0/1),$$

где $P(1/1) = 0{,}6$, $P(0/1) = 1 - P(1/1) = 0{,}4$,

$$P(X/Y_1) = 0{,}6 \cdot 0{,}4 \cdot 0{,}6 \cdot 0{,}6 \cdot 0{,}4 = 0{,}035.$$

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

$$P(X/Y_2) = P(1/0)P(0/0)P(1/0)P(1/0)P(0/0),$$

где $P(0/0) = 0{,}6$, $P(1/0) = 1 - P(0/0) = 0{,}4$,

$$P(X/Y_2) = 0{,}4 \cdot 0{,}6 \cdot 0{,}4 \cdot 0{,}4 \cdot 0{,}6 = 0{,}023.$$

По формуле полной вероятности

$$P(X) = P(Y_1)P(X/Y_1) + P(Y_2)P(X/Y_2) = 0{,}7 \cdot 0{,}035 + 0{,}3 \cdot 0{,}023 = 0{,}0314.$$

По формуле Байеса

$$P(Y_1/X) = \frac{P(Y_1)P(A/B_1)}{P(A)} = \frac{0{,}7 \cdot 0{,}035}{0{,}0314} = 0{,}78,$$
$$P(Y_2/X) = \frac{P(Y_2)P(A/B_2)}{P(A)} = \frac{0{,}3 \cdot 0{,}023}{0{,}0314} = 0{,}22,$$

*Ответ*. Сравнивая вероятности, можно сделать вывод о том, что более вероятна передача сообщения 11111.

## 2.2 Понятие информации

Информация является одним из фундаментальных понятий современной науки наряду с такими понятиями, как «вещество» и «энергия».

Впервые как научное понятие термин «информация» стал применяться в теории журналистики в 30-х годах XX века, хотя в исследованиях по библиотечному делу он появился еще раньше. Под информацией понимались различные сведения, сообщения. Что соответствует переводу с латинского языка informatio – сведение, разъяснение, ознакомление.

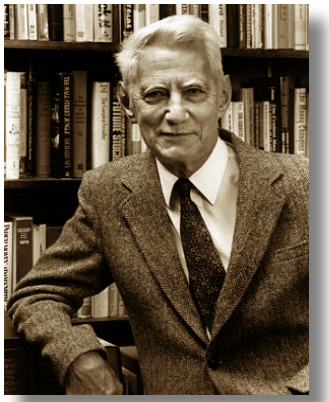
Общее определение этому термину дать невозможно. Однако в различных предметных областях даётся специализированное определение информации, подходящее для данной предметной области.

В физике понятие «информация» рассматривается как антиэнтропия или энтропия[1] с обратным знаком. Поскольку мерой беспорядка термодинамической системы является энтропия системы, то информация (антиэнтропия) является мерой упорядоченности и сложности системы. По мере увеличения сложности системы величина энтропии уменьшается, и величина информации увеличивается. Процесс увеличения информации характерен для открытых, обменивающихся веществом и энергией с окружающей средой, саморазвивающихся систем живой природы (белковых молекул, организмов, популяций животных и т.д.).

В неживой природе понятие «информация» связано с понятием отражения, отображения. В некоторых физических и химических теориях информация определяется как отраженное многообразие. Отражение заключается в таком изменении одного материального объекта под воздействием другого, при котором все особенности отражаемого объекта каким-либо образом воспроизводятся отражающим объектом. В процессе отражения и происходит передача информации. Т.е. информация – это результат отражения. В соответствии с этим взглядом информация существовала и будет существовать вечно, она содержится во всех элементах и системах материального мира.

Под информацией в технике понимают сообщение, передаваемое с помощью знаков и символов. В теории связи, например, под информацией принято понимать любую последовательность символов, не учитывая их смысл.

В основанной американским ученым Клодом Шенноном[2] математической теории информации под информацией понимались не любые сведения, а лишь те, которые снимают полностью или уменьшают существующую до их получения неопределенность (неизвестность). Каждому сигналу в теории Шеннона соответствует вероятность его появления. Например, при передаче текста телеграммы вероятность появления буквы «т» равна 1/33. Чем меньше вероятность появления того или иного сигнала, тем больше информации он несет для потребителя. В обыденном понимании это означает, что, чем неожиданнее новость, тем больше ее информативность (содержательный подход).



Клод Э́лвуд Ше́ннон (1916-2001)

[1] Понятие энтропии впервые было введено в термодинамике в 1865 году для определения меры необратимого рассеивания энергии, меры отклонения реального процесса от идеального. Определённая как сумма приведённых теплот, она является функцией состояния и остаётся постоянной при замкнутых обратимых процессах, тогда как в необратимых – её изменение всегда положительно. Математически энтропия определяется как функция состояния системы, равная в равновесном процессе количеству теплоты, сообщённой системе или отведённой от системы, отнесённому к термодинамической температуре системы.

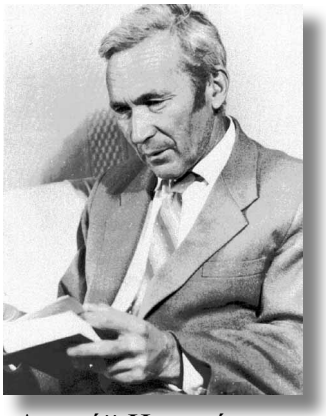
[2] Клод Э́лвуд Ше́ннон (англ. Claude Elwood Shannon; 30 апреля 1916, Мичиган, США – 24 февраля 2001, Медфорд, Массачусетс, США) – американский инженер и математик, его работы являются синтезом математических идей с конкретным анализом чрезвычайно сложных проблем их технической реализации. Является основателем теории информации, нашедшей применение в современных высокотехнологических системах связи. Шеннон внес огромный вклад в теорию вероятностных схем, теорию автоматов и теорию систем управления – области наук, входящие в понятие «кибернетика». В 1948 году предложил использовать слово «бит» для обозначения наименьшей единицы информации.

## ОПРЕДЕЛЕНИЕ$_1$

В содержательном подходе, ***информация*** – это снятая неопределённость. ***Неопределённость*** некоторого события – это количество возможных результатов (исходов) данного события.

Например, если мы подбрасываем вверх монету, то она может упасть двумя различными способами (орлом вверх или решкой вверх). Соответственно, у данного события два возможных исхода. Если же подбрасывать игральный кубик, то исходов будет шесть.

В подходе предложенном математиком А. Н. Колмогоровым[3] – алфавитном, понятие информации определим следующим образом:



Андре́й Никола́евич Колмого́ров (1903-1987)

## ОПРЕДЕЛЕНИЕ$_2$

В алфавитном подходе ***информация*** – это сообщение (последовательность символов некоторого алфавита). Причём существенными являются только размер алфавита и количество символов в сообщении. Конкретное содержание сообщения интереса не представляет. Чаще всего алфавит является двоичным (состоит из 2 символов – «0» и «1»).

Математик Ю.А. Шрейдер оценивал информацию по увеличению объема знаний у человека под воздействием информационного сообщения. Академик А.А. Харкевич измерял содержательность сообщения по увеличению вероятности достижения цели после получения информации человеком или машиной. Таким образом, под информацией в семантической теории понимают сведения, обладающие новизной.

В кибернетике – науке об управлении в живых, неживых и искусственных системах – понятие информации связывают воедино с понятием управления (Норберт Винер). Жизнедеятельность любого организма или нормальное функционирование технического устройства зависит от процессов управления, благодаря которым поддерживаются в необходимых пределах значения их параметров. Процессы управления включают в себя получение, хранение, преобразование и передачу информации. Информация является обозначением содержания, полученного из внешнего мира в процессе приспособления к нему наших чувств. Информацию составляет та часть знания, которая используется для ориентирования, принятия решений, активного действия, управления, т.е. в целях сохранения, совершенствования и развития системы. Данная концепция отрицает

---

[3] Андре́й Никола́евич Колмого́ров (12 (25) апреля 1903, Тамбов – 20 октября 1987, Москва) – советский математик, один из крупнейших математиков XX века. Один из основоположников современной теории вероятностей, им получены фундаментальные результаты в топологии, геометрии, математической логике, классической механике, теории турбулентности, теории сложности алгоритмов, теории информации, теории функций, теории тригонометрических рядов, теории меры, теории приближения функций, теории множеств, теории дифференциальных уравнений, теории динамических систем, функциональном анализе и в ряде других областей математики и её приложений. Автор новаторских работ по философии, истории, методологии и преподаванию математики, известны его работы в статистической физике (в частности, уравнение Джонсона – Мела – Аврами – Колмогорова).

существование информации в неживой природе, не дает ответа на вопросы: являются ли информацией неиспользованные знания, являются ли информацией неосмысленная информация?

Для преодоления этих противоречий академик В.Г. Афанасьев ввел понятие информационных данных.

*Информационные данные – это всякие сведения, сообщения, знания, которые могут храниться, перерабатываться, передаваться, но характер информации они приобретут лишь тогда, когда получат содержание и форму, пригодную для управления, и используются в управлении.*

В биологии, которая изучает живую природу, понятие «информация» связано с целесообразным поведением живых организмов. Такое поведение строится на основе получения и использования организмом информации об окружающей среде.

Понятие информация используется в связи с исследованием механизмов наследственности. В генетике сформулировано понятие генетической информации, которое определяется как программа (код) биосинтеза белков, представленных цепочками ДНК. Реализуется эта информация в ходе развития особи. Последнее обстоятельство позволило проводить научные эксперименты по клонированию, т.е. созданию точных копий организмов их одной клетки.

В социальных науках (социологии, психологии, политологии и др.) под информацией понимают сведения, данные, понятия, отраженные в нашем сознании и изменяющие наши представления о реальном мире. Эту информацию, передающуюся в человеческом обществе и участвующую в формировании общественного сознания, называют социальной информацией.

Под информацией в документалистике понимают все то, что так или иначе зафиксировано в знаковой форме в виде документов.

С точки зрения индивидуального человеческого сознания информация – это то, что поступает в наш мозг из многих источников в разных формах и, взаимодействуя там, образует структуру нашего знания. Под информацией в быту (житейский аспект) понимают сведения об окружающем мире и протекающем в нем процессах, воспринимаемые человеком или специальными устройствами. Информацией для человека являются не только сухие факты, строгие инструкции, но и то, что радует нас, волнует, печалит, заставляет переживать, восторгаться, презирать, негодовать. Более половины общего объема сведений, полученных в процессе разговора, приходится на так называемую несмысловую информацию. Эту информации говорящий по своему желанию, а иногда и непроизвольно, сообщает нам своей тональностью разговора, своей возбужденностью, жестикуляцией, выражением лица, глаз и т.д.

Информация может быть двух видов: *дискретная (цифровая)* и *непрерывная (аналоговая)*. Дискретная информация характеризуется последовательными точными значениями некоторой величины, а непрерывная – непрерывным процессом изменения некоторой величины. Непрерывную информацию может, например, выдавать датчик атмосферного давления или датчик скорости автомашины. Дискретную информацию можно получить от любого цифрового индикатора: электронных часов, счетчика магнитофона и т.п.

Дискретная информация удобнее для обработки человеком, но непрерывная информация часто встречается в практической работе, поэтому необходимо уметь переводить непрерывную информацию в дискретную (дискретизация) и наоборот. Модем (это слово происходит от слов модуляция и демодуляция) представляет собой устройство для такого перевода: он переводит цифровые данные от компьютера в звук или электромагнитные колебания-копии звука и наоборот.

При переводе непрерывной информации в дискретную важна так называемая *частота дискретизации* $\nu$, определяющая период $T = 1/\nu$ между измерениями значений непрерывной величины (см. рис. 7).

Чем выше частота дискретизации, тем точнее происходит перевод непрерывной информации в дискретную. Но с ростом этой частоты растет и размер дискретных данных, получаемых при таком переводе, и, следовательно, сложность их обработки, передачи и хранения. Однако для повышения точности дискретизации необязательно безграничное увеличение ее частоты. Эту частоту разумно увеличивать только до предела, определяемого теоремой о выборках, называемой также теоремой Котельникова или законом Найквиста.

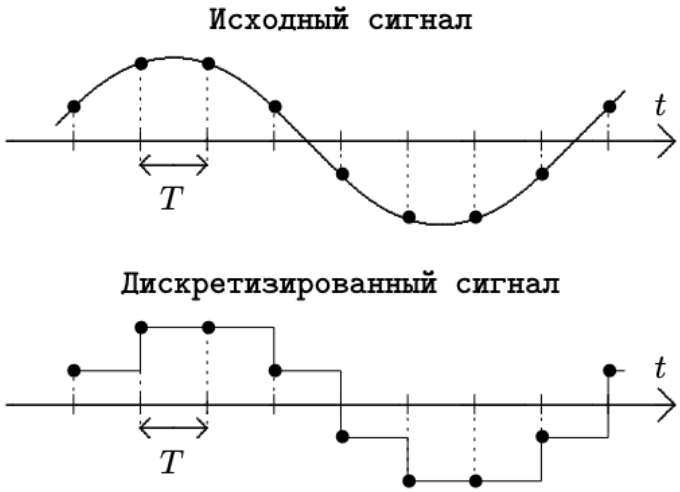



Рисунок 7 – Перевод непрерывного сигнала в дискретный

## 2.3 Информация и данные

Информация и данные – это базовые понятия, которые используются в информатике. Эта наука занимается вопросами систематизации, хранения, обработки и передачи данных и информации средствами вычислительной техники. Эти понятия зачастую используются как синонимы, но между ними существуют и принципиальные различия.

Мы живем в материальном мире. Все, что нас окружает и с чем мы сталкиваемся ежедневно, относится либо к физическим телам, либо к физическим полям. Из курса физики мы знаем, что состояния абсолютного покоя не существует и физические объекты находятся в состоянии непрерывного движения и изменения, которое сопровождается обменом энергией и ее переходом из одной формы в другую.

Все виды энергообмена сопровождаются появлением сигналов (от лат. signum – знак), то есть все сигналы имеют в своей основе материальную энергетическую

основу. При взаимодействии сигналов с физическими телами в последних возникают определенные изменения свойств – это явление называется регистрацией сигналов. Такие изменения можно наблюдать, измерять или фиксировать иными способами – при этом возникают и регистрируются новые сигналы, то есть образуются данные (рис. 8).

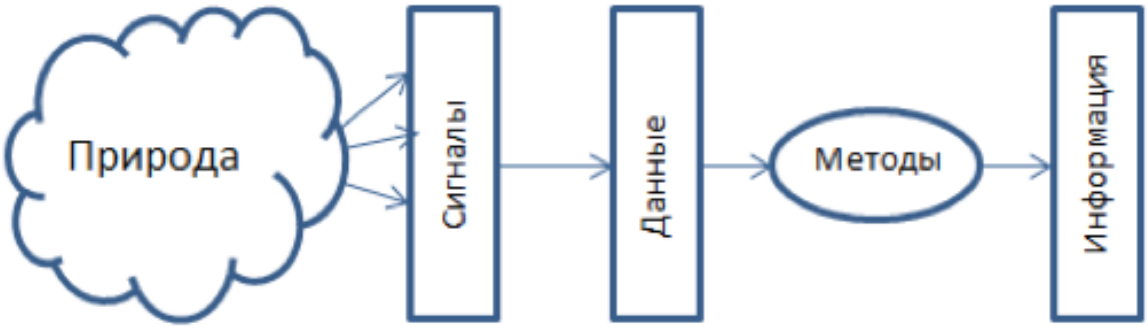



Рисунок 8 – Схема информационного процесса

Отметим, что данные несут в себе информацию о событиях, произошедших в материальном мире, поскольку они являются регистрацией сигналов, возникших в результате этих событий. Однако данные не тождественны информации. Приведем несколько примеров, иллюстрирующих данное утверждение.

Наблюдая за состязанием бегунов, мы с помощью механического секундомера регистрируем начальное и конечное положение стрелки прибора. В итоге мы замеряем величину перемещения за время забега – это регистрация данных. Однако информации о времени преодоления дистанции мы пока не получаем. Для того чтобы данные о перемещении стрелки дали информацию о времени забега, необходимо наличие метода пересчета одной физической величины в другую.

Прослушивая передачу радиостанции на незнакомом языке, мы получаем данные, но не получаем информацию в связи с тем, что не владеем методом преобразования данных в известные нам понятия. Если эти данные записать на лист бумаги или на магнитную ленту, изменится форма представления, произойдет новая регистрация и, соответственно, образуются новые данные. Такое преобразование можно использовать, чтобы все-таки извлечь информацию из данных путем подбора метода, адекватного их новой форме. Для обработки данных, записанных на листе бумаги адекватным может быть метод перевода со словарем, а для обработки данных, записанных на магнитной ленте, можно пригласить переводчика, обладающего своими методами перевода.

Расселом Аккофом была предложена пирамида демонстрирующая соотношение между понятиями данные, информация, знания (рис. 9)

**[ДАННЫЕ — ИНФОРМАЦИЯ — ЗНАНИЯ — ПОНИМАНИЕ — МУДРОСТЬ]**

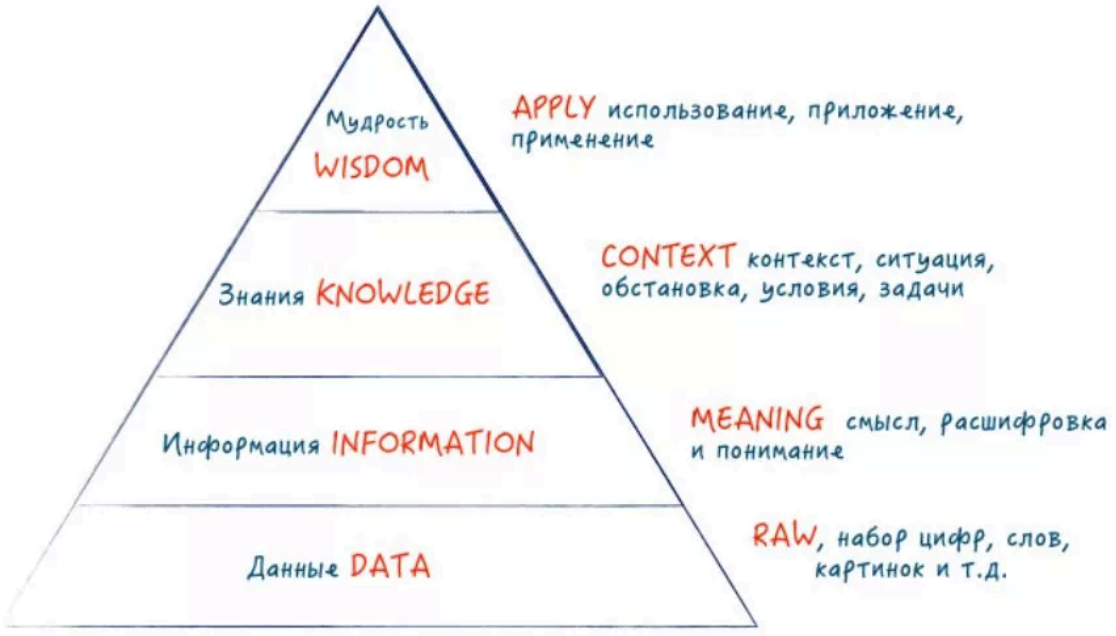



Рисунок 9 – Информационная пирамида Рассела Аккофа

***Данные*** – это некоторые неупорядоченные символы, измерения, сигналы, рассматриваемые безотносительно к какому-либо контексту (данные могут быть цифровыми (факты, результаты измерений), графическими, аудио, видео и т.п. Они могут описываться на различных языках (символьном, математическом, графическом и т.п.).

***Информация*** – обработанные, осмысленные данные. Это упорядоченная часть базы данных, обработанная для использования, и отвечающая на вопрос: «Кто?, Что?, Где?, Когда?» (это данные в определенном контексте, наделенные значимостью и целями)

**ИНФОРМАЦИЯ = ДАННЫЕ + МЕТАДАННЫЕ**

*Преобразование данных в информацию*:

— контекстуализация: известно, для какой цели данные были собраны;

— *категоризация*: известны единицы анализа или ключевые компоненты данных;

— *вычисляемость*: данные могут быть проанализированы математически или статистически;

— *корректировка*: ошибки убраны из данных;

— *сжатие*: данные могут быть обобщены в более сжатую форму.

С точки зрения информатики данные – это совокупность сведений, зафиксированных на определенном носителе в форме, пригодной для постоянного хранения, передачи и обработки. Преобразование и обработка данных позволяет получить информацию. Например, в базах данных хранятся различные данные, а по определенному запросу система управления базой данных выдает требуемую информацию.

Классификация информации может быть выполнена по различным критериям. Например,

1) по форме представления: графическая, текстовая, числовая, звуковая, видео;

2) по способу восприятия: визуальная, аудиальная, тактильная, обонятельная, вкусовая;

3) по стадии обработки: первичная, вторичная, промежуточная, результатная;

4) по стабильности: переменная, постоянная;

5) по функции управления: плановая, нормативно-справочная, учетная, оперативная.

# 3. Общие сведения о передаче информации

*Сбор информации* – это процесс получения информации из внешнего мира и приведение ее к стандарту для данной информационной системы. Обмен информацией между воспринимающей ее системой и окружающей средой осуществляется посредством сигналов. *Сигнал* – средство передачи информации в пространстве и времени или физический процесс, отображающий (несущий) сообщение. В качестве носителя сигнала могут выступать звук, свет, электрический ток, магнитное поле и т.д. Сбор информации, как правило, сопровождается ее регистрацией, т.е. фиксацией информации на материальном носителе (документе или машинном носителе).

*Передача информации* – физический процесс, посредством которого осуществляется перемещение знаков (сведений, способных предоставлять информацию) в пространстве или осуществляется физический доступ субъектов к знакам.

Информация передается в виде некоторых сообщений. *Дискретные сообщения* формируются в результате последовательной выдачи источником сообщений отдельных элементов – *знаков*.

*Непрерывные сообщения* не разделены на элементы. Они описываются непрерывными функциями времени, принимающими непрерывное множество значений (речь, телевизионное изображение).

Преобразование сообщения в сигнал, удобный для передачи по данному каналу связи, называют *кодированием в широком смысле слова*.

В *узком смысле кодирование* представляет собой преобразование дискретного сообщения в последовательность кодовых символов, осуществляемое по определенному правилу. Устройство, выполняющее такую операцию, называют кодирующим или *кодером*. Множество всех кодовых последовательностей (кодовых комбинаций), возможных при данном правиле кодирования, образует *код*. Совокупность символов, из которых составляются кодовые последовательности, называют кодовым *алфавитом*, а их число (объем кодового алфавита) — *основанием кода*. Число символов в кодовой комбинации может быть одинаковым или разным. Соответственно различают равномерные и неравномерные коды. Число символов в кодовой комбинации равномерного кода называется длиной кода.

Операцию восстановления сообщения по принятому сигналу называют *декодированием*.

Для операции сопоставления символов со знаками исходного алфавита используют термин «декодирование». Техническая реализация этой операции осуществляется декодирующим устройством или *декодером*.

*Передающее устройство* осуществляет преобразование непрерывных сообщений или знаков в сигналы, удобные для прохождения по линии связи. При этом один или несколько параметров выбранного сигнала изменяют в соответствии с передаваемой информацией. Такой процесс называют *модуляцией*. Он осуществляется *модулятором*. Обратное преобразование сигналов в символы производится *демодулятором*

Под *линией связи* понимают среду (воздух, металл, магнитную ленту и т.д.), обеспечивающую поступление сигналов от передающего устройства к приемному устройству.

Сигналы на выходе линии связи могут отличаться от сигналов на ее входе (переданных) вследствие затухания, искажения и воздействия помех.

Помехами называют любые мешающие возмущения, как внешние, так и внутренние, вызывающие отклонение принятых сигналов от переданных сигналов. Из смеси сигнала с помехой *приемное устройство* выделяет сигнал и посредством декодера восстанавливает сообщение, которое в общем случае может отличаться от посланного. Меру соответствия принятого сообщения посланному сообщению называют *верностью передачи*.

Принятое сообщение с выхода системы связи поступает к абоненту-получателю, которому была адресована исходная информация.

Совокупность средств, предназначенных для передачи сообщений, называют *каналом связи*.

На рис. 10 представлена структурна схема радиотехнических систем передачи информации (РСПИ)

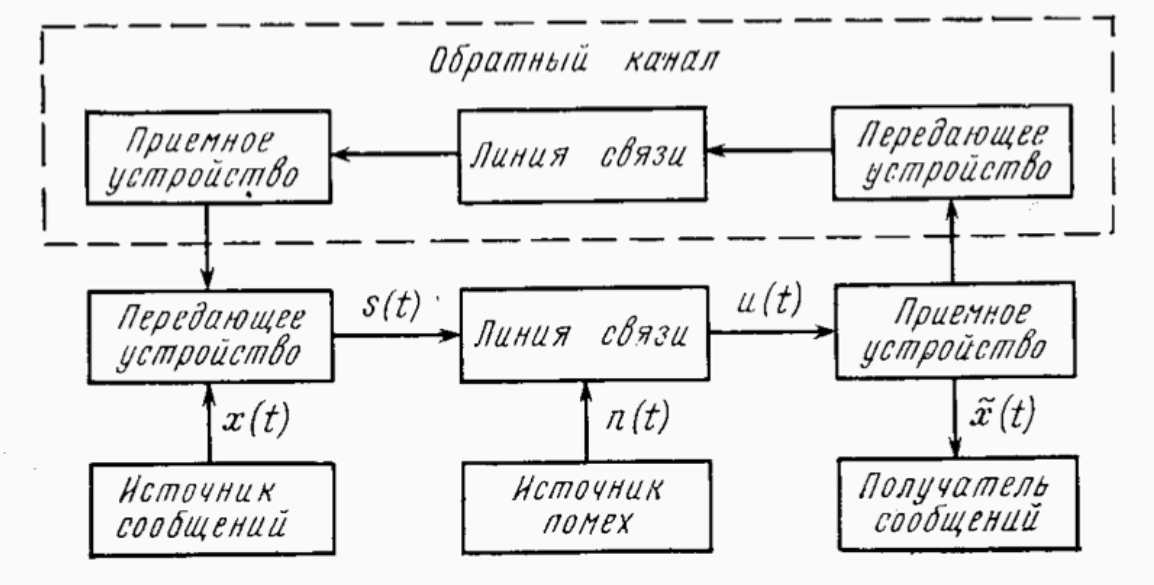



Рисунок 10 – Обобщенная структурная схема РСПИ

## 3.1 Классификация сигналов и их математические модели

Передача информации заключается в переносе ее на расстояние при помощи сигналов различной физической природы. Для того чтобы сделать сигнал объектом теоретического изучения или практических расчетов, необходимо указать способ его математического описания или, другими словами, создать математическую модель сигнала.

*Математической моделью* называют совокупность математических соотношений, описывающих изучаемый процесс или явление. Математические модели позволяют анализировать свойства сигналов и синтезировать сигналы с требуемыми свойствами.

В простейшем случае математическая модель сигнала устанавливает соответствие между любым моментом времени $t \in T$ и величиной сигнала $x \in X$, где $T$ – ограниченный или бесконечный интервал времени, называемый областью определения сигнала, $X$ – множество возможных значений сигнала. Это соответствие может быть задано в форме скалярной функции $x(t)$. В более сложных случаях модель содержит математические соотношения, которые характеризуют некоторые обобщенные свойства сигнала. Создание математической модели – это первый и очень важный этап изучения физического процесса или явления. Во-первых, математическая модель позволяет абстрагироваться от конкретной физической природы носителя сигнала. При этом она приобретает определенную универсальность, то есть способность описывать различные по своей физической природе процессы или по техническому назначению объекты. Одна и та же математическая модель может описывать изменение тока, напряжения, давления, температуры и т.д. Во-вторых, математическая модель создается так, чтобы она описывала именно те свойства сигнала, которые наиболее важны для конкретного исследования. Необходимо учитывать, что математическая модель может хорошо работать в одних условиях и быть совершенно неприемлемой в других. Любая математическая модель имеет свою область применения и эта область, как правило, может быть определена только в результате многократного применения модели. Поэтому при составлении математической модели большое значение приобретают экспериментальные исследования и практический опыт.

Большое значение при выборе математической модели сигнала имеет ее сложность. Математическая модель, с одной стороны, должна быть достаточно сложна, чтобы отображать существенные свойства изучаемого процесса или явления, а, с другой стороны, достаточно проста, чтобы использовать более простые математические методы анализа.

Очевидно, чем полнее математическая модель отражает свойства сигнала, тем шире область ее применения, тем больший круг задач позволяет она решать. В то же самое время, чем полнее математическая модель, тем она сложнее, тем больше трудностей следует ожидать при исследовании из-за необходимости привлекать более сложные математические методы. Поэтому математическая модель сигнала не должна содержать больше подробностей, чем это необходимо для решения данной задачи.

## 3.2 Детерминированные и случайные сигналы

*Детерминированными* или *регулярными* называются такие сигналы, значения которых в любой точке интервала их определения можно рассчитать заранее, имея математическую модель. Математической моделью детерминированного сигнала является детерминированная функция $x(t)$. Такая модель позволяет для любого заданного момента времени $t_i$ однозначно определить значение сигнала $x(t_i)$.

На рис. 11 в качестве примера приведены графики некоторых широко распространенных детерминированных сигналов.

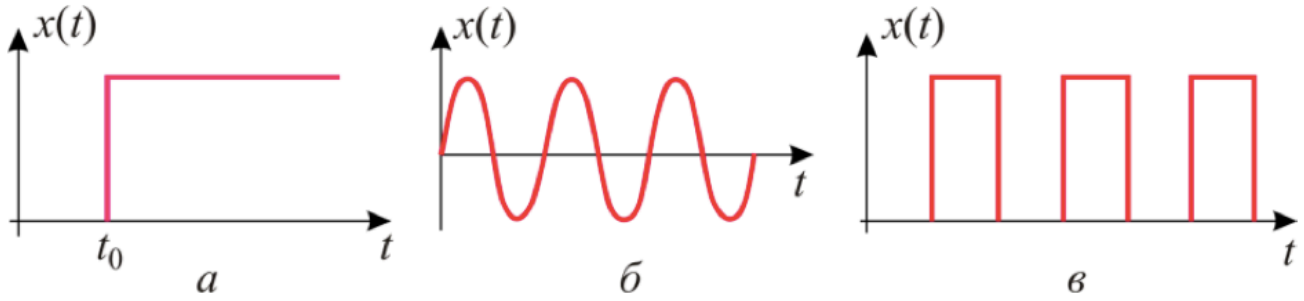



Рисунок 11 – Графики детерминированных сигналов:
*а* – ступенчатого сигнала; *б* – синусоидального сигнала; *в* – импульсного сигнала.

Строго говоря, детерминированные сигналы в чистом виде в природе существовать не могут. Такие сигналы могли бы возникнуть только в изолированных системах. Любая же система находится в некоторой среде, и эта среда влияет на процессы, происходящие в системе. Поэтому детерминированные сигналы являются определенной идеализацией реальных сигналов и не содержат информации. И периодическая последовательность импульсов, и синусоидальный сигнал, а тем более единичный ступенчатый сигнал очень далеки по своим свойствам от реально действующих на системы сигналов. Однако при помощи детерминированных сигналов можно изучить многие существенные особенности установившихся и переходных процессов в линейных и нелинейных системах. Поэтому они широко используются при исследовании систем различного назначения.

*Случайными* или *стохастическими* называются такие сигналы, изменение которых во времени предсказать невозможно. Такие сигналы описываются случайными функциями. Случайной функцией некоторой независимой переменной $x$ называют такую функцию $y(t)$, значение которой при любом заданном $x$ является случайной величиной. Случайные функции, для которых независимой переменной является время $t$, обычно называют случайными (или стохастическими) процессами (сигналами).

Случайность процесса проявляется в том, что вид функции $x(t)$ случайным образом меняется от одного опыта к другому. Функцию, получаемую в результате каждого отдельного опыта, называют реализацией случайного процесса.

Для его количественной оценки вводится понятие случайной функции $X(t)$. Значение $X(t_i)$ случайной функции при фиксированном значении аргумента $t_i$ представляет собой случайную величину. Поэтому можно говорить лишь о вероятности того, что в данный момент времени $t = t_i$ значение $X(t_i)$ случайной функции заключено в интервале между значениями $x$ и $x + dx$. В этом случае при неизменных условиях эксперимента случайная функция $X(t)$ может принимать различные конкретные формы $x_j(t)$ , которые называются реализациями случайного процесса (рис. 12). Множество реализаций образуют ансамбль реализаций, который полностью определяет случайный сигнал $X(t)$.

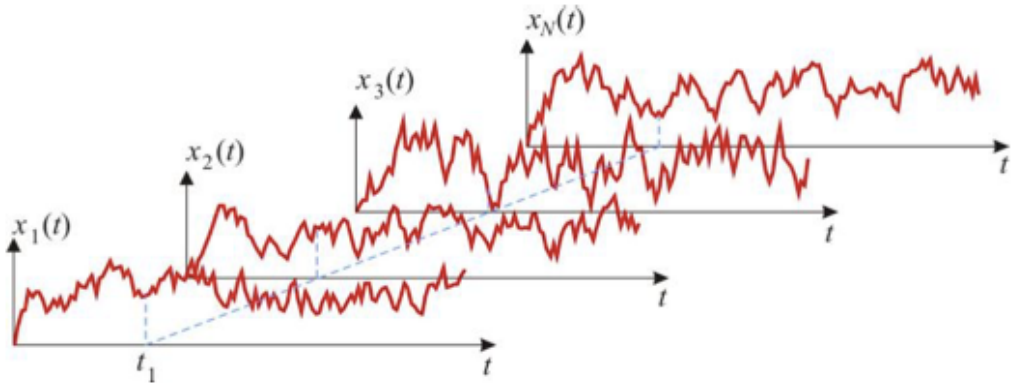



Рисунок 12 – Ансамбль реализаций случайного сигнала

## 3.3 Периодические и непериодические сигналы

На практике часто приходится иметь дело с явлениями, повторяющимися в прежнем виде через определенный промежуток времени $T$. Сигналы, характеризующие периодические явления, возвращаются к своим прежним значениям через указанный интервал времени. Такие явления и сигналы называются ***периодическими***, а промежуток времени $T$ – периодом. Они описываются периодическими функциями.

Функция $x(t)$ называется периодической, если она определена на всей действительной оси и для всякого $t \in (-\infty, +\infty)$ удовлетворяет условию

$$x(t) = x(t + rT), r = \pm 1, \pm 2, .... \qquad (3.1)$$

Основная особенность периодического сигнала состоит в том, что его значения периодически повторяются и что периодичность эта существует вечно. Очевидно, что периодических явлений и сигналов в строгом смысле соотношения (3.1) в действительности нет и быть не может из-за физического ограничения времени наблюдения сигналов. Однако периодическая функция – это удобная математическая абстракция, облегчающая теоретический анализ сигналов.

Если для периодической функции $x(t)$ существует интеграл (собственный или несобственный), то при любом действительном числе $a$ выполняется условие

$$\int_a^{a+T} x(t)dt = \int_0^T x(t)dt.$$

## 3.4 Импульсные сигналы

***Импульсными*** называют сигналы, которые существуют на ограниченном отрезке времени или на некоторой совокупности ограниченных чередующихся отрезков времени, называемых областями существования.

Импульсный сигнал, имеющий единственную область существования, называется импульсом (например, на рис. 13,*а* показан прямоугольный импульс). В общем случае импульсный сигнал представляет собой последовательность

импульсов, продолжающихся на конечном или бесконечном интервале времени (рис. 13,*б*).

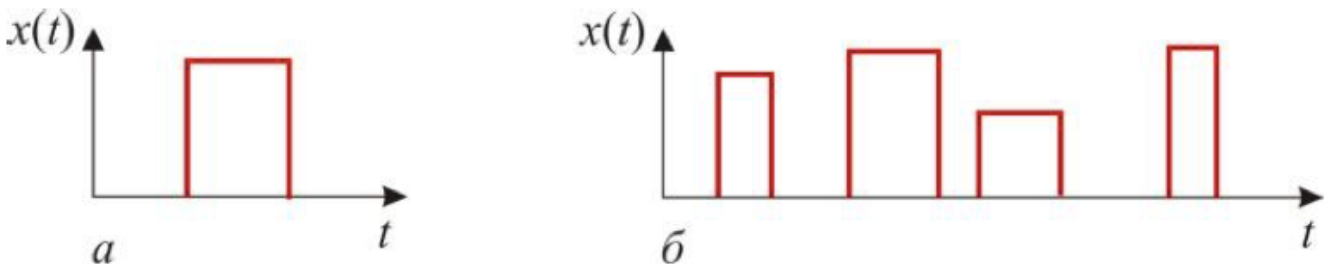



Рисунок 13 – Графики импульсных сигналов:
*а* – одиночного импульса; *б* – последовательности импульсов.

Сигналы, в любой форме материального представления, содержат определенную полезную информацию. Главная цель обработки физических сигналов заключается в необходимости получения содержащейся в них информации. Эта информация обычно присутствует в амплитуде сигнала (абсолютной или относительной), в частоте или в спектральном составе, в фазе или в относительных временных зависимостях нескольких сигналов.

Приведем некоторые примеры сигналов и необходимости их обработки с целью извлечения содержащейся в них информации.

- Параметры перехваченного радиосигнала (частоты излучения, типа модуляции, начала и конца посылок) позволяет идентифицировать объект наблюдения.
- Анализ сейсмических сигналов помогает получить представление о конфигурации геологических объектов и предсказать землетрясения.
- Изучение отклика геодезического зонда позволяет обнаружить полезные ископаемые и определить характеристики залежей.
- Анализ электрокардиограммы способствует диагностированию болезней сердца, в том числе на ранних стадиях их возникновения.

Как правило, весь процесс обработки сигнала делится на первичную обработку и вторичную – интерпретацию полученных результатов. Например, при анализе электрокардиограмм первичная обработка заключается в усилении сигналов датчиков, фильтрации помех и аналого-цифровом преобразовании. Вторичная обработка может состоять в определении длительностей *RR*-интервалов и построении *RR*-интервалограммы, которая характеризует не только функциональное состояние сердца, но и общее состояние систем организма.

Одну из самых распространенных операций первичной обработки сигналов представляет фильтрация. Цель фильтрации состоит в подавлении помех, содержащихся в сигнале, или в выделении отдельных составляющих сигнала. Другой широко распространенной задачей обработки сигналов является оценка спектра, которая позволяет получить представление о внутренней структуре наблюдаемого явления.

До недавнего времени обработка сигналов, как правило, выполнялась при помощи аналоговых методов и устройств. Цифровая обработка сигналов заявила о

себе в начале второй половины прошлого века и вызвала оживленные дискуссии, вплоть до полного отрицания со стороны некоторых ученых.

Основной предметной областью теории и практики во второй половине XX века стали цифровая фильтрация и спектральный анализ, причем оба направления рассматривались с общей позиции частотных представлений. В это время появились алгоритмы быстрого преобразования Фурье (БПФ), прибавившие оптимизма сторонникам методов цифровой обработки. За последние десятилетия благодаря интенсивному развитию микроэлектроники цифровая обработка сигналов вышла на передовые позиции и во многих прикладных областях вытеснила аналоговую.

***Цифровая обработка сигналов*** – одна из самых динамичных и быстро развивающихся технологий в мире. Сегодня обработка аналоговых сигналов с использованием цифровых методов все шире используется для решения множества прикладных задач в связи, радиолокации, звуковой локации, акустике, измерительной технике, медицине, ядерной энергетике и других областях науки и техники, в которых прежде доминировали аналоговые системы.

Процесс цифровой обработки сигналов состоит из трех этапов:

- преобразование аналогового сигнала $x(t)$ в дискретную последовательность $x(n)$ , каждый отсчет которой представлен в виде двоичного числа;
- преобразование дискретной последовательности $x(n)$ по заданному алгоритму в дискретную последовательность $y(n)$;
- преобразование дискретной последовательности $y(n)$ в аналоговый сигнал $y(t)$ .

На практике цифровая обработка сигналов реализуется в виде программы на компьютере, либо в виде специализированного устройства, содержащего микропроцессор и электронные схемы ввода-вывода.

# 4. Задачи и постулаты прикладной теории информации

К теории информации относят результаты решения ряда фундаментальных теоретических вопросов:

- анализ сигналов как средства передачи сообщений, включающий вопросы оценки переносимого ими «количества информации»;
- анализ информационных характеристик источников сообщений и каналов связи и обоснование принципиальной возможности кодирования и декодирования сообщений, обеспечивающих предельно допустимую скорость передачи сообщений по каналу связи, как при отсутствии, так и при наличии помех.

В теории информации исследуются информационные системы при четко сформулированных *условиях (постулатах)*:

*1. Источник сообщения осуществляет выбор сообщения из некоторого множества с определенной вероятностью.*

*2. Сообщения могут передаваться по каналу связи в закодированном виде. Кодированные сообщения образуют множество, являющееся взаимно однозначным отображением множества сообщений. Правило декодирования известно декодеру (записано в его программе).*

*3. Сообщения следуют друг за другом, причем число сообщений может быть сколь угодно большим.*

*4. Сообщение считается принятым верно, если в результате декодирования оно может быть в точности восстановлено. При этом не учитывается, сколько времени прошло с момента передачи сообщения до момента окончания декодирования, и какова сложность операций кодирования и декодирования.*

*5. Количество информации не зависит от смыслового содержания сообщения, от его эмоционального воздействия, полезности и даже от его отношения к реальной действительности.*

Хранение и накопление информации вызвано многократным ее использованием, применением постоянной информации, необходимостью комплектации первичных данных до их обработки. Хранение осуществляется на машинных носителях в виде информационных массивов, где данные располагаются по установленному в процессе проектирования группировочному признаку.

***Качество информации*** – положительная характеристика информации, отражающая степень ее полезности для пользователя.

К ***потребительским показателям*** *качества* информации можно отнести следующие:

- Репрезентативность – правильность отбора и формирования информации в целях адекватного отражения свойств объекта.

- Содержательность – семантическая емкость. С увеличением содержательности информации растет семантическая пропускная способность информационных систем, так как для получения одних и тех же сведений требуется преобразовать меньший объем данных.
- Достаточность (полнота) – свойство информации исчерпывающе (для данного потребителя) характеризовать отображаемый объект и/или процесс.
- Доступность – свойство информации, характеризующее возможность ее получения данным потребителем.
- Актуальность – определяется степенью сохранения ценности информации для управления в момент ее использования и зависит от динамики изменения ее характеристик и от интервала времени, прошедшего со времени решения поставленной задачи.
- Своевременность – означает ее поступление не позже заранее назначенного момента времени, согласованного со временем решения поставленной задачи.
- Точность – определяется степенью близости получаемой информации к реальному состоянию объекта, процесса, явления и т. п.
- Достоверность – определяется ее свойством отражать реально существующие объекты с необходимой точностью.
- Устойчивость – отражает способность информации реагировать на изменение исходных данных без нарушения необходимой точности.
- Защищенность – свойство, характеризующее невозможность несанкционированного использования или изменения.
- Эргономичность – свойство, характеризующее удобство формы или объем информации с точки зрения данного потребителя.

## 4.1 Энтропия. Количество информации. Единицы измерения информации

Ральф Винтон Лайон Хартли (1888-1970)

Количеством информации называют числовую характеристику сигнала, независящую от его формы и содержания, и характеризующую неопределенность, которая исчезнет после получения сообщения в виде данного сигнала.

Научный подход к оценке сообщений был предложен в 1928 году Ральфом Хартли. Проиллюстрируем его следующим образом.

Пусть имеется алфавит, из букв которого составляется сообщение. Количество букв в алфавите равно $m$. В этом случае количество возможных равновероятных вариантов разных сообщений длины $n$ будет равно $N = m^n$ .

*Пример*. Пусть алфавит состоит из двух букв «А» и «В», длина сообщения 3 буквы. Тогда $m = 2$ , $n = 3$ . С таким алфавитом и длине сообщения можно составить $N = m^n = 2^3 = 8$ разных сообщений («ААА», «ААВ», «АВА», «АВВ», «ВАА», «ВАВ», «ВВА», «ВВВ»).

Формула Хартли определяет количество информации, содержащейся в сообщении длины $n = 3$:

$$H = log_2 N = n log_2 m,$$

где $H$ – количество информации. Для приведенного выше примера $H = 3$ бит.

***Бит*** – минимальная единица измерения информации, которая представляет собой двоичный знак двоичного алфавита {0;1}.

Если $N$ – это количество равновероятных событий, то количество информации в сообщении о том, что произошло одно из N событий, равно $H = log_2 N$ бит.

*Пример*. В поезде 16 вагонов. Сообщение о том, что ваш знакомый приехал во 2-м вагоне, несет $H = log_2 16 = 4$ бита информации.

В соответствии с объемным методом совокупность 100 букв – фраза из 100 букв из газеты, пьесы Шекспира или теории относительности Эйнштейна – имеет в точности одинаковое количество информации.

Рассмотрим ряд примеров, прежде чем рассмотреть другой подход к измерению информации.

*Пример.* Книга лежит на одной из двух полок – верхней или нижней. Сообщение о том, что книга лежит на верхней полке, уменьшает неопределенность ровно вдвое и несет 1 бит информации.

*Сообщение о том, что произошло одно событие из двух равновероятных, несет 1 бит информации.*

*Пример.* Нестеров живет на Ленинградской улице. Мы получили сообщение, что номер его дома есть число четное, которое уменьшило неопределенность. После получения такой информации, мы стали знать больше, но информационная неопределенность осталась, хотя и уменьшилась в два раза.

*Пример.* Ваш друг живет в 16-ти этажном доме. Сколько информации содержит сообщение о том, что друг живет на 7 этаже.

*Решение*: Информационная неопределенность (количество возможных результатов события) равна 16. Будем задавать вопросы, на которые можно ответить только «да» или «нет». Вопрос будем ставить так, чтобы каждый ответ приносил 1 бит информации, т.е. уменьшал информационную неопределенность в два раза.

Задаем вопросы: - Друг живет выше 8-го этажа?
- Нет.

После этого ответа число вариантов уменьшилось в два раза, следовательно, информационная неопределенность уменьшилась в два раза. Получен 1 бит информации.

- Друг живет выше 4-го этажа?
- Да.

Число вариантов уменьшилось еще в два раза, получен еще 1 бит информации.

- Друг живет выше 6-го этажа?
- Да.

После данного ответа осталось два варианта: друг живет или на 7 этаже, или на 8 этаже. Получен еще 1 бит информации.

- Друг живет на 8-м этаже?
- Нет.
- Все ясно. Друг живет на 7-м этаже.

Каждый ответ уменьшал информационную неопределенность в два раза.

Всего было задано 4 вопроса. Получено 4 бита информации. Сообщение о том, что друг живет на 7-м этаже 16-ти этажного дома несет 4 бита информации.

В теории информации принят и другой подход к измерению информации, называемый энтропийным. ***Энтропийный подход*** исходит из следующей модели. Получатель сообщения имеет определенные представления о возможных наступлениях некоторых событий. Эти представления в общем случае недостоверны и выражаются вероятностями, с которыми он ожидает то или иное событие. Общая мера неопределенности – энтропия, характеризуется некоторой математической зависимостью от совокупности этих вероятностей. Количество информации в сообщении определяется тем, насколько уменьшается эта мера после получения сообщения. Разберемся как это работает.

Любое сообщение в теории информации является сведением о некоторой физической системе. Если состояние системы известно заранее, то нет смысла передавать сообщение[4].

В качестве объекта, о котором передается информация, будем рассматривать физическую систему $X$, которая случайным образом может оказаться в том или ином состоянии, т.е. систему, которой присуща какая-то степень неопределенности.

Сведения, полученные о системе, будут тем ценнее, чем больше была неопределенность системы до получения сведений. Степень неопределенности физической системы определяется числом ее возможных состояний и вероятностями состояний.

Рассмотрим систему $X$, которая может принимать конечное множество состояний $x_1, x_2, \dots x_n$ с вероятностями $p_1, p_2, \dots p_n$, $\sum_i p_i = 1$, где

$$p_i = P(X \sim x_i)$$

или

| $x_i$ | $x_1$ | $x_2$ | $\cdots$ | $x_n$ |
|---|---|---|---|---|
| $p_i$ | $p_1$ | $p_2$ | $\cdots$ | $p_n$ |

[4] Блинова И.В., Попов И.Ю. Теория информации. Учебное пособие. – СПб: Университет ИТМО, 2018. – 84 с

В качестве меры неопределенности системы применяется специальная характеристика, называемая ***энтропией***.

Энтропией называется величина, вычисляемая по формуле:

$$H(X) = -\sum_{i=1}^{n} p_i log_a p_i.$$

Основание логарифма можно взять любым $a > 1$. Выбор основания равносилен выбору единицы измерения энтропии. Если за основание выбрать число 10, то говорят о «десятичных единицах» энтропии (дитах). На практике в качестве основания удобнее использовать число 2. При выполнении вычислений будем считать $a = 2$. В этом случае за единицу измерения энтропии принимается энтропия простейшей системы, которая имеет два равновозможных состояния, а сама энтропия измеряется в «двоичных единицах» или битах (binary digit). В вычислительной технике вся обрабатываемая и хранимая информация вне зависимости от ее природы представлена в двоичной форме (с использованием алфавита, состоящего всего из двух символов 0 и 1).

| $x_i$ | $x_1$ | $x_2$ |
|---|---|---|
| $p_i$ | 1/2 | 1/2 |

Здесь $H(X) = -\frac{1}{2} log \frac{1}{2} - \frac{1}{2} log \frac{1}{2} = 1$ бит. Бит – это очень маленькая единица, поэтому используют другую величину в 8 раз большую – байт, и кратные ей величины (табл).

***Байт*** – единица количества информации в СИ, представляющая собой восьмиразрядный двоичный код, с помощью которого можно представить один символ.

Рассмотрим систему, которая имеет $n$ равновероятных состояний:

| $x_i$ | $x_1$ | $x_2$ | $\cdots$ | $x_n$ |
|---|---|---|---|---|
| $p_i$ | $1/n$ | $1/n$ | $\cdots$ | $1/n$ |

Здесь имеем: $H(X) = -n \frac{1}{n} log \frac{1}{n} = log n$. Следовательно, энтропия системы с равновозможными состояниями равна логарифму числа состояний.

**Информационный объем сообщения** (информационная емкость сообщения) - количество информации в сообщении, измеренное в стандартных единицах или производных от них (Кбайтах, Мбайтах и т. д.)[14].

| | | |
|---|---|---|
| | 1 байт = 8 бит | B |
| Кило- | 1 Кбайт = $2^{10}$ байт = 1024 байт | KB |
| Мега- | 1 Мбайт = $2^{10}$ Кбайт = $2^{20}$ байт | MB |
| Гига- | 1 Гбайт = $2^{10}$ Мбайт = $2^{20}$ Кбайт = $2^{30}$ байт | GB |
| Тера- | 1 Тбайт = $2^{10}$ Гбайт = $2^{20}$ Мбайт = $2^{30}$ Кбайт = $2^{40}$ байт | TB |
| Пета- | 1 Пбайт = $2^{50}$ байт | PB |
| Экса- | 1 Эбайт = $2^{60}$ байт | EB |
| Зетта- | 1 Збайт = $2^{70}$ байт | ZB |
| Йотта- | 1 Йбайт = $2^{80}$ байт | YB |

## 4.2 Свойства энтропии

1. Если состояние системы в точности известно заранее, то ее энтропия равна нулю. В этом случае все вероятности $p_1, p_2, p_3, \ldots, p_n$ в формуле для энтропии обращаются в ноль, кроме одной, которая равна 1.

Слагаемое $p_k log p_k = 0$, т.к. $log 1 = 0$. Остальные слагаемые обращаются в ноль, т. к.

$$\lim_{p \to 0} p log p = 0.$$

2. Энтропия системы с конечным множеством состояний достигает максимума, когда все состояния равновероятны.

$$\Delta H(X) = -\sum_{i=1}^{n} p_i log p_i \text{ - энтропия системы.}$$

Рассмотрим функцию $H(p_1, \ldots, p_n) = -p_1 log p_1 - \ldots - p_n log p_n$. Найдем условный экстремум этой функции при условии $p_1 + \ldots + p_n = 1$.

Чтобы найти условный экстремум функции $H(p_1, \ldots, p_n)$, надо определить обычный экстремум функции

$$L(p_1, \ldots, p_n, \lambda) = -p_1 log p_1 - \ldots - p_n log p_n + \lambda(p_1 + \ldots + p_n - 1).$$

(метод неопределенных множителей Лагранжа).

Необходимое условие экстремума: $L'_{p_i} = 0$, $(i = 1,...n)$ и $L'_{\lambda} = 0$.

Дифференцируем функцию $L(p_1, \ldots, p_n, \lambda)$ и приравниваем производные к нулю. Получим систему уравнений:

$$log p_1 = \lambda - log\, e$$
$$\ldots$$
$$log p_n = \lambda - log\, e$$

$$p_1 + \ldots + p_n = 1.$$

Система имеет единственное решение $p_1 = \ldots = p_n = \frac{1}{n}$ .

Достаточное условие экстремума:

$$d^2L = -\frac{1}{p_1}dp_1^2 - \ldots - \frac{1}{p_n}dp_n^2 < 0.$$

Значит функция $H(p_1, \ldots, p_n)$ имеет условный максимум в точке $(\frac{1}{n}, \ldots, \frac{1}{n})$.

Максимальная энтропия системы $H_{max}(X) = log\ n$.

Следовательно, максимальное значение энтропии системы с конечным множеством состояний равно логарифму числа состояний и достигается, когда все состояния равновероятны.

Введем специальную функцию:

$$\eta(p) = -plogp$$

Тогда формула для энтропии примет вид:

$$H(X) = -\sum_{i=1}^{n} \eta(p_i).$$

Представим формулу для энтропии в виде математического ожидания. Рассмотрим $logP(X)$ как случайную величину. Когда система $X$ принимает значения $x_1, \ldots, x_n$, случайная величина $logP(X)$ принимает значения $logp(x_1), \ldots, logp(x_n)$ с вероятностями$p(x_1), \ldots p(x_n)$:

| $logp(x_i)$ | $logp(x_1)$ | $logp(x_2)$ | ... | $logp(x_n)$ |
|---|---|---|---|---|
| $p(x_i)$ | $p(x_1)$ | $p(x_2)$ | ... | $p(x_n)$ |

Тогда, $M[-logP(x)] = -p(x_1)logp(x_1) - \ldots - p(x_n)logp(x_n)$.

Следовательно, $H(X) = M[-logP(X)]$.

*Пример*. Найти энтропию системы $X$, вероятности состояний которой заданы законом распределения:

| $x_i$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ |
|---|---|---|---|---|---|---|
| $p_i$ | 0,2 | 0,3 | 0,1 | 0,05 | 0,15 | 0,2 |

*Решение*.

Для расчетов воспользуемся таблицами в ПРИЛОЖЕНИИ 1:

$$H(X) = \sum_{i=1}^{6} \eta(p_i) = 0{,}4644 + 0{,}5211 + 0{,}3322 + 0{,}2161 + 0{,}4105 + 0{,}4644 = 2{,}4087 \text{ бит.}$$

*Пример*. Определить максимально возможную энтропию технического устройства, состоящего из четырех элементов (устройство выходит из строя при отказе любого из элементов).

*Решение*. Рассмотрим техническое устройство как систему $X$, которая может находиться в $2^4 = 16$ состояниях. Энтропия системы достигает максимума, когда все состояния равновероятны и равна $H(X) = log16 = 4$ бит.

*Пример*. Написать функцию $H(p_1, p_2)$, определяющую энтропию системы с тремя состояниями. Чему равно наибольшее значение этой функции?

*Решение*. Рассмотрим систему $X$ с тремя состояниями $x_1, x_2, x_3$. Вероятности состояний:

| $x_i$ | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|
| $p_i$ | $p_1$ | $p_2$ | $1 - p_1 - p_2$ |

Энтропия такой системы равна:

$$H(p_1, p_2) = -p_1 log p_1 - p_2 log p_2 - (1 - p_1 - p_2) log(1 - p_1 - p_2).$$

Энтропия системы максимальна, если состояния системы равновероятны:

| $x_i$ | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|
| $p_i$ | 1/3 | 1/3 | 1/3 |

и равна логарифму числа состояний. Следовательно, функция $H(p_1, p_2)$ достигает максимума в точке (1/3, 1/3) и наибольшее значение функции равно $H_{max}(p_1, p_2) = log3$.

*Пример*. Имеются две урны. В первой – 5 красных, 7 белых шаров. Во второй – 8 красных и 6 белых шаров. Из каждой урны берут по два шара. Исход какого из двух опытов следует считать более неопределенным?

*Решение*. Рассмотрим первый опыт, извлечение двух шаров из первой урны, как систему $X$ с тремя исходами:

| Исходы $x_i$ | $x_1$ (2 кр. ш.) | $x_2$ (1 кр. ш.+1 бел. ш.) | $x_3$ (2 кр. ш.) |
|---|---|---|---|
| Вероятности исходов | $\frac{5}{12}\cdot\frac{4}{11}=\frac{20}{132}$ | $\frac{5}{12}\cdot\frac{7}{11}+\frac{7}{12}\cdot\frac{5}{11}=\frac{70}{132}$ | $\frac{7}{12}\cdot\frac{6}{11}=\frac{42}{132}$ |

Найдем энтропию системы $X$:

$$H(X)=\eta(\frac{20}{132})+\eta(\frac{70}{132})+\eta(\frac{42}{132})=0{,}4118+0{,}4854+0{,}5256=1{,}4228 \text{ бит.}$$

Рассмотрим второй опыт, извлечение двух шаров из второй урны, как систему $Y$ с тремя исходами:

| Исходы $y_i$ | $y_1$ (2 кр. ш.) | $y_2$ (1 кр. ш.+1 бел. ш.) | $y_3$ (2 кр. ш.) |
|---|---|---|---|
| Вероятности исходов | $\frac{8}{14}\cdot\frac{7}{13}=\frac{56}{182}$ | $\frac{8}{14}\cdot\frac{6}{13}+\frac{6}{14}\cdot\frac{8}{13}=\frac{96}{182}$ | $\frac{6}{14}\cdot\frac{5}{13}=\frac{30}{182}$ |

$$H(Y)=\eta(\frac{56}{182})+\eta(\frac{96}{182})+\eta(\frac{30}{182})=0{,}5230+0{,}4870+0{,}4277=1{,}4377 \text{ бит.}$$

Т.к. $H(X)<H(Y)$, то исход второго опыта является более неопределенным.

## 4.3 Энтропия сложной системы

Рассмотрим сложную систему, полученную объединением двух или более простых систем.

Под объединением двух систем $X$ и $Y$ с возможными состояниями $x_1,\dots x_n$ и $y_1,\dots y_n$ понимается сложная система $(X,Y)$, состояния которой $(x_i,y_j)$ представляют собой все всевозможные комбинации состояний $x_i,y_j\ (i=1,...n,\ j=1...m)$ систем $X$ и $Y$.

Пусть $p_{ij}$ – вероятность того, что система $(X,Y)$ будет в состоянии $(x_i,y_j)$:

$$p_{ij}=P((X\sim x_i)(Y\sim y_j))$$

| $(x_i,y_j)$ | $x_1$ | $x_2$ | … | $x_n$ |
|---|---|---|---|---|
| $y_1$ | $p_{11}$ | $p_{21}$ | … | $p_{n1}$ |
| $y_2$ | $p_{12}$ | $p_{22}$ | … | $p_{n2}$ |
| … | … | … | … | … |
| $y_n$ | $p_{1m}$ | $p_{2m}$ | … | $p_{nm}$ |

Энтропия сложной системы равна сумме произведений вероятностей всех возможных ее состояний на их логарифмы с обратным знаком:

$$H(X,Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} log p_{ij}$$

или

$$H(X,Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} \eta(p_{ij}).$$

Энтропия сложной системы в форме математического ожидания:

$$H(X,Y) = M[-logP(X,Y)].$$

Рассмотрим случай, когда системы $X$ и $Y$ независимы, т.е. принимают свои состояния независимо одна от другой. Найдем энтропию такой системы.

По теореме умножения вероятностей для независимых систем:

$$P(X,Y) = P(X)\cdot P(Y) \Rightarrow logP(X,Y) = logP(X) + logP(Y).$$

Так как

$$H(X,Y) = M[-logP(X,Y)],$$

то

$$H(X,Y) = M[-logP(X) - logP(Y)] = M[-logP(X)] - M[-logP(Y)].$$

Следовательно

$$H(X,Y) = H(X) + H(Y).$$

Итак, при объединении независимых систем их энтропии складываются (теорема сложения энтропий).

Для произвольного числа независимых систем $X_1, X_2, \ldots, X_n$:

$$H(X_1, X_2, \ldots X_n) = \sum_{i=1}^{n} H(X_i).$$

Если системы зависимы, то энтропия сложной системы меньше, чем сумма энтропий ее составных частей.

*Пример*. Вероятности состояний независимых систем $X$ и $Y$ заданы таблицами
Найти энтропию объединенной системы $(X, Y)$.

*Решение*. Т.к. системы $X$ и $Y$ независимы, то $H(X,Y) = H(X) + H(Y)$. Тогда,

$$H(X) = \eta(0{,}6) + \eta(0{,}4) = 0{,}4422 + 0{,}5288 = 0{,}971 \text{ бит},$$
$$H(Y) = \eta(0{,}5) + \eta(0{,}5) = 0{,}5 + 0{,}5 = 1 \text{ бит},$$
$$H(X, Y) = 0{,}971 + 1 \approx 1{,}971 \text{ бит}.$$

## 4.4 Условная энтропия

Пусть имеются две зависимые системы $X$ и $Y$. Пусть система $X$ приняла состояние $x_i$.

$p(y_i/x_i)$ – условная вероятность того, что система $Y$ примет состояние $y_j$ при условии, что система $X$ находится в состоянии $x_i$:

$$p(y_i/x_i) = P(Y \sim y_i/X \sim x_i).$$

Тогда,

$$H(Y, x_i) = -\sum_{j=1}^{m} p(i_j/x_i) log p(y_j/x_i) \tag{4.1}$$

или

$$H(Y, x_i) = -\sum_{j=1}^{m} \eta(p(y_i/x_i))$$

где $H(Y, x_i)$ – ***частная условная энтропия*** системы $Y$ при условии, что система $X$ находится в состоянии $x_i$.

Частная условная энтропия зависит от того, какое состояние $x_i$ приняла система $X$.

Определим ***среднюю (полную) условную энтропию*** системы $Y$ с учетом того, что система $X$ может принимать разные состояния:

$$H(Y, X) = \sum_{i=1}^{n} p_i H(Y/x_i). \tag{4.2}$$

Используя выражение (4.1), получим

$$H(Y/X) = -\sum_{i=1}^{n} \sum_{j=1}^{m} p_i p_j (y_i/x_i) log p(y_j/x_i) \tag{4.3}$$

или

$$H(Y/X) = -\sum_{i=1}^{n} \sum_{j=1}^{m} p_i \eta(p(y_i/x_i)).$$

С другой стороны, по теореме умножения вероятностей $p_{ij} = p_i \cdot p(y_j/x_i)$. Используя выражение (4.3), получим

$$H(Y/X) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} log p(y_j/x_i). \qquad (4.4)$$

Придадим выражению (4.4) форму математического ожидания:

$$H(Y/X) = -M[-logP(Y/X)].$$

Величина $H(Y/X)$ характеризует степень неопределенности системы $Y$, остающуюся после того, как состояние системы $X$ полностью определилось.

$H(Y/X)$ – полная условная энтропия системы $Y$ относительно системы $X$.

Аналогично определим полную условную энтропию системы $X$ относительно системы $Y$. Пусть $r_j = P(Y \sim y_j)$, тогда

$$H(Y/X) = \sum_{j=1}^{m} r_j H(X/y_j) \qquad (4.5)$$

или

$$H(Y/X) = \sum_{i=1}^{n}\sum_{j=1}^{m} r_j \eta(p(x_i/y_j)).$$

*Пример*. Сложная система $(X, Y)$, задана таблицей

| $(x_i, y_j)$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ |
|---|---|---|---|---|
| $y_1$ | 0,5 | 0,3 | 0 | 0,01 |
| $y_2$ | 0,1 | 0,01 | 0,06 | 0,02 |

Найти полные условные энтропии $H(Y/X)$ и $H(X/Y)$, и частные энтропии $H(Y/x_i), i = 1,2,3,4$ и $H(X/y_j), j = 1,2$.

*Решение*. Напишем таблицы вероятностей для систем $X$ и $Y$:

| $x_i$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ |
|---|---|---|---|---|
| $p_i$ | 0,6 | 0,31 | 0,06 | 0,03 |

| $y_j$ | $x_1$ | $x_2$ |
|---|---|---|
| $r_j$ | 0,81 | 0,19 |

1. Т. к. $p_{ij} = p_i \cdot p(y_j/x_i)$, то $p(y_j/x_i) = \dfrac{p_{ij}}{p_i}$. Найдем условные вероятности $p(y_j/x_i)$:

$$p(y_1/x_1) = \frac{0{,}5}{0{,}6} \quad p(y_1/x_2) = \frac{0{,}3}{0{,}31} \quad p(y_1/x_3) = 0 \quad p(y_1/x_4) = \frac{0{,}01}{0{,}03}$$

$$p(y_2/x_1) = \frac{0{,}1}{0{,}6} \quad p(y_2/x_2) = \frac{0{,}01}{0{,}31} \quad p(y_2/x_3) = 1 \quad p(y_2/x_4) = \frac{0{,}02}{0{,}03}$$

$$H(Y/x_1) = \sum_{j=1}^{2} \eta(p(y_j/x_1)) = \eta(p(y_1/x_1)) + \eta(p(y_2/x_1)) =$$

$$= \eta(\frac{0{,}5}{0{,}6}) + \eta(\frac{0{,}1}{0{,}6}) = \eta(0{,}833) + \eta(0{,}166) = 0{,}2196 + 0{,}4301 = 0{,}6497,$$

$$H(Y/x_2) = \eta(\frac{0{,}3}{0{,}31}) + \eta(\frac{0{,}01}{0{,}31}) = 0{,}0468 + 0{,}1589 = 0{,}2057,$$

$$H(Y/x_3) = 0,$$

$$H(Y/x_4) = \eta(\frac{0{,}01}{0{,}3}) + \eta(\frac{0{,}02}{0{,}3}) = 0{,}5283 + 0{,}3905 = 0{,}9188.$$

$$H(Y/X) = \sum_{i=1}^{4} p_i H(Y/x_i) = p_1 H(Y/x_1) + p_2 H(Y/x_2) + p_3 H(Y/x_3) + p_4 H(Y/x_4) =$$

$$= 0{,}6 \cdot 0{,}6497 + 0{,}31 \cdot 0{,}2057 + 0{,}03 \cdot 0{,}9188 = 0{,}4811 \text{ бит.}$$

2. Т. к. $p_{ij} = r_j \cdot p(x_i/y_j)$, то $p(x_i/y_j) = \dfrac{p_{ij}}{r_j}$. Найдем условные вероятности $p(x_i/y_j)$:

$$p(x_1/y_1) = \frac{0{,}5}{0{,}81} \quad p(x_2/y_1) = \frac{0{,}3}{0{,}81} \quad p(x_3/y_1) = 0 \quad p(x_4/y_1) = \frac{0{,}01}{0{,}81}$$

$$p(x_1/y_2) = \frac{0{,}1}{0{,}19} \quad p(x_2/y_2) = \frac{0{,}01}{0{,}19} \quad p(x_3/y_2) = \frac{0{,}06}{0{,}19} \quad p(x_4/y_2) = \frac{0{,}02}{0{,}19}$$

$$H(X/y_1) = \sum_{i=1}^{4} \eta(p(x_i/y_1)) = \eta(p(x_1/y_1)) + \eta(p(x_2/y_1)) + \eta(p(x_3/y_1)) + \eta(p(x_4/y_1)) =$$

$$= \eta(\frac{0{,}5}{0{,}81}) + \eta(\frac{0{,}3}{0{,}81}) + \eta(\frac{0{,}01}{0{,}81}) = 1{,}0371,$$

$$H(X/y_2) = \eta(\frac{0{,}1}{0{,}19}) + \eta(\frac{0{,}01}{0{,}19}) + \eta(\frac{0{,}06}{0{,}19}) + \eta(\frac{0{,}2}{0{,}19}) = 1{,}5757,$$

$$H(X/Y) = \sum_{j=1}^{2} r_j H(X/y_j) = r_1 H(X/y_1) + r_2 H(X/y_2) = 0{,}81 \cdot 1{,}0371 + 0{,}19 \cdot 1{,}5757 = 1{,}1394 \text{ бит}.$$

Выше было дано определение энтропии, как меры неопределенности состояния физической системы. После получения сведений неопределенность может быть уменьшена. Чем больше получено сведений, тем больше будет информация о системе, менее неопределенным будет состояние системы. Поэтому количество информации о системе измеряют уменьшением энтропии этой системы.

Рассмотрим систему $X$. Пусть в результате наблюдений над системой состояние системы становится полностью известным. Энтропия системы до наблюдений $H(X)$. После получения сведений состояние системы полностью определилось и энтропия системы стала равна нулю. При этом информация, получаемая в результате выяснения состояния системы $X$, равна уменьшению энтропии:

$$I(X) = H(X) - 0,$$

т. е.

$$I(X) = H(X) = -\sum_{i=1}^{n} p_i log p_i.$$

Тогда, величина $I(X)$ – среднее значение случайной величины $-log p_i$ $(i = 1,2,...n)$:

| $-log p_i$ | $-log p_1$ | $-log p_2$ | … | $-log p_n$ |
|---|---|---|---|---|
| $p_i$ | $p_1$ | $p_2$ | … | $p_n$ |

При этом каждое отдельное слагаемое $-log p_i$ рассматривается, как частная информация, получаемая от отдельного сообщения, состоящего в том, что система $X$ находится в состоянии $x_i$. Введем величину:

$$I(x_i) = -log p_i.$$

Тогда информация $I(X)$ представляется как средняя, или полная, информация, получаемая от всех возможных отдельных сообщений с учетом их вероятностей.

***Свойства информации:***

1. Полная и частная информация не могут быть отрицательными:

$$I(X) \geq 0, I(x_i) \geq 0.$$

2. Если все всевозможные состояния системы одинаково вероятны $(p_1 = p_2 = \ldots p_n = \frac{1}{n})$, то частная информация $I(x_i)$ от каждого отдельного сообщения,

$$I(x_i) = -log p_i = log\, n,$$

равна средней (полной) информации

$$I(X) = -n\frac{1}{n}log\frac{1}{n} = logn.$$

3. Если состояния системы обладают различными вероятностями, то информации, получаемые от разных сообщений неодинаковы. Наибольшую информацию несут сообщения о тех событиях которые изначально были наименее вероятны.

Рассмотрим систему $X$, вероятности состояний которой связаны с законом распределения:

| $x_i$ | $x_1$ | $x_2$ |
|---|---|---|
| $p_i$ | 0,99 | 0,01 |

Частная информация $I(x_1)$, получаемая от сообщения, что система $X$ приняла состояние $x_1$, равна:

$$I(x_1) = -log0{,}99 \approx 0{,}0144 \text{ бит},$$

а частная информация в случае $X \sim x_2$:

$$I(x_2) = -log0{,}01 \approx 6{,}6438 \text{ бит}.$$

*Пример*. Лыжник съезжает с годы без падения с вероятностью 0,95. Какое количество информации мы получим, узнав, что лыжник упал на склоне?

*Решение*. Рассмотрим лыжника, как систему $X$ с двумя состояниями:

| $x_i$ | $x_1$ | $x_2$ |
|---|---|---|
| $p_i$ | 0,95 | 0,05 |

Сообщение «лыжник упал» несет информацию, равную:

$$I(x_2) = -log0{,}05 \approx 4{,}3219 \text{ бит}.$$

*Пример*. Игрок наудачу бросает два игральных кубика. Какое количество информации при этом получает игрок?

*Решение*. Рассмотрим два кубика, как систему $X$ с $6 \cdot 6 = 36$ равновероятными состояниями. Тогда,

$$I(X) = -log36 \approx 5{,}1699 \text{ бит}.$$

*Пример*. Из колоды в 36 карт наудачу берут три. Найти частную информацию из сообщения: «выпали шестерка, семерка, туз».

*Решение*. Вероятность события «выпали шестерка, семерка, туз» равна $p_i = 6 \cdot \frac{4 \cdot 4 \cdot 4}{36 \cdot 35 \cdot 34}$. Частная информация в этом случае равна:

$$I(x_i) = -\log p_i = 6{,}8017 \text{ бит.}$$

В предыдущих примерах предполагалась, что наблюдение ведется непосредственно за самой системой $X$. На практике часто система $X$ оказывается недоступной. При этом выясняется состояние другой системы $Y$, связанной с $X$. Например, вместо непосредственного наблюдения за космическим кораблем ведется наблюдение за системой сигналов, передаваемых его аппаратурой.

Различия между интересующей нас системой $X$ и поддающейся наблюдению системой $Y$ бывают двух типов:

1. Различия за счет того, что некоторые состояния системы $X$ не находят отражение в системе $Y$. Например, за счет округления численных данных.

2. Различия за счет ошибок: неточностей измерения параметров системы $X$ и ошибок при передаче сообщений. Например, искажение сигнала в следствии помех.

В случае, когда интересующая система $X$ и наблюдаемая система $Y$ различны, то количество информации о системе $X$ дающее наблюдение системы $Y$ определяется как уменьшение энтропии системы $X$ в результате получения сведений о состоянии системы $Y$:

$$I(Y, X) = H(X) - H(X/Y).$$

Величина $I(Y, X)$ – средняя, или полная, информация о системе $X$, содержащаяся в системе $Y$.

***Теорема***. Докажем, что $I(Y, X) = I(X, Y)$, т.е. из двух систем каждая содержит относительно другой одну и ту же полную информацию.

Запишем энтропию объединенной системы двумя равносильными формулами:

$$H(X, Y) = H(X) + H(Y/X) \text{ и } H(X, Y) = H(Y) + H(X/Y).$$

Следовательно,

$$H(X) + H(Y/X) = H(Y) + H(X/Y)$$
$$H(X) - H(X/Y) = H(Y) - H(Y/X)$$
$$I(Y/X) = I(X/Y)$$

$I(X/Y)$ — полная взаимная информация, содержащаяся в системах $X$ и $Y$.

Посмотрим, во что обращается информация в крайних случаях полной зависимости и полной независимости систем.

1. Если $X$ и $Y$ независимы, то $H(Y/X) = H(Y)$ и $I(X, Y) = 0$, т.е. полная взаимная информация, содержащаяся в независимых системах равна нулю (нельзя получить сведения о системе, наблюдая вместо нее другую, никак с ней не связанную).

2. Если состояние системы $X$ полностью определяет состояние системы $Y$ и наоборот, т.е. системы эквивалентны, то $H(X) = H(Y)$ и $H(X/Y) = H(Y/X) = 0$. Следовательно,

$$I(X, Y) = I(X) = I(Y) = H(X) = H(Y).$$

3. Рассмотрим случай, когда между системами $X$ и $Y$ наблюдается строгая односторонняя зависимость: состояние одной системы $Y$ полностью определяет состояние другой $X$ (подчиненной системы), но не наоборот.
$X$ — подчиненная система, тогда $H(X, Y) = 0$ и $I(X, Y) = H(X)$, т.е. полная взаимная информация, содержащаяся в системах, из которых одна является подчиненной, равна энтропии подчиненной системы.

Получим выражение для информации $I(X, Y)$ через энтропию объединенной системы $H(X, Y)$ и энтропию ее составных частей $H(X)$ и $H(Y)$. Т.к.

$$H(X, Y) = H(Y) + H(X/Y), \text{ то } H(X/Y) = H(X, Y) - H(Y).$$

Тогда,

$$I(X, Y) = H(Y) - H(X, Y) = H(X) + H(Y) - H(X, Y).$$

Значит, полная взаимная информация, содержащаяся в двух системах, равна сумме энтропий этих систем минус энтропия объединенной системы.

Получим выражение для информации $I(X, Y)$ через вероятности состояний систем. Имеем равенство

$$I(X, Y) = H(X) + H(Y) - H(X, Y),$$

где

$$H(X) = M[-logP(X)], H(Y) = M[-logP(Y)],$$

$$H(X, Y) = M[-logP(X, Y)],$$

тогда

$$I(X, Y) = M[-logP(X) - logP(Y) + logP(X, Y)],$$

$$I(X, Y) = M[log\frac{P(X, Y)}{P(X)P(Y)}].$$

Запишем последнюю формулу в виде:

$$I(X, Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} log\frac{p_{ij}}{p_i r_j},$$

где $p_{ij} = P((X \sim x_i)(Y \sim y_j)), p_i = P(X \sim x_i), r_j = P(Y \sim y_j)$.

*Пример*. Вероятности состояний зависимых систем X и Y заданы таблицей:

| $(x_i, y_j)$ | $x_1$ | $x_2$ |
|---|---|---|
| $y_1$ | 0,45 | 0,15 |
| $y_2$ | 0,05 | 0,35 |

Найти полную взаимную информацию, содержащуюся в этих системах.

*Решение*. Найдем вероятности состояний каждой из систем:

| $x_i$ | $x_1$ | $x_2$ |
|---|---|---|
| $p_i$ | 0,6 | 0,4 |

| $y_j$ | $y_1$ | $y_2$ |
|---|---|---|
| $p_j$ | 0,5 | 0,5 |

$H(X) = 1,$
$H(Y) = \eta(0,6) + \eta(0,4) = 0,4422 + 0,5288 = 0,971$
$H(X, Y) = \eta(0,45) + \eta(0,15) + \eta(0,05) + \eta(0,35) = 0,5184 + 0,4105 + 0,2161 + 0,5301 = 1,6751$
$I(X, Y) = H(X) + H(Y) - H(X, Y) = 0,2959$ бит.

## 4.5 Частная информация о системе

I. Найдем частную информацию о системе $X$, содержащуюся в отдельном сообщении, указывающем, что система $Y$ находится в конкретном состоянии $y_j$. Обозначим эту информацию $I(y_j, X)$. При этом должно быть выполнено:

$$I(Y, X) = \sum_{j=1}^{m} r_j I(y_j, X).$$

Имеем $p_{ij} = r_j p(x_i / y_j)$, тогда

$$I(Y, X) = \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} log \frac{p_{ij}}{p_i r_j} = \sum_{i=1}^{n}\sum_{j=1}^{m} r_j p(x_i/y_j) log \frac{r_j p(x_i/y_j)}{p_i r_j} = \sum_{j=1}^{m}\sum_{i=1}^{n} p(x_i/y_j) log \frac{p(x_i/y_j)}{p_i}.$$

$$\Rightarrow I(y_j, X) = \sum_{i=1}^{n} p(x_i/y_j) log \frac{p(x_i/y_j)}{p_i}.$$

Примем это выражение за определение частной информации о системе, содержащейся в сообщении о событии. Можно доказать, что

Частная информация о системе $X$, заключенная в сообщении о состоянии $y_j$ системы $Y$, не может быть отрицательной.

Значит, неотрицательна и полная взаимная информация, как математическое ожидание неотрицательной случайной величины: $I(X,Y)>0$. Имеем $I(X,Y)=H(Y)-H(Y/X)$, тогда $H(Y)-H(Y/X)\geq 0$ и следовательно, $H(Y)\geq H(Y/X)$. Полная условная энтропия системы не превосходит ее безусловной энтропии.

Преобразуем формулу для частной информации, введя вместо условных $p(x_i/y_j)$ вероятностей безусловные. Т.к. $p(x_i/y_j)=\dfrac{p_{ij}}{r_j}$, то

$$I(y_j,X)=\sum_{i=1}^{n}\frac{p_{ij}}{r_j}log\frac{p_{ij}}{p_i r_j}.$$

Таким образом, определили частную информацию о системе $X$, содержащуюся в сообщении «система $Y$ находится в состоянии $y_j$».

II. Определим частную информацию о событии $X\sim x_i$, содержащуюся в событии $Y\sim y_j$, т.е. получим информацию «от события к событию». Введем информацию «от события к событию» следующим образом:

$$I(y_j,x_i)=log\frac{p(x_i/y_j)}{p_i}. \qquad (4.6)$$

Частная информация о событии, получаемая в результате сообщения о другом событии, равна логарифму отношения вероятности первого сообщения после сообщения о другом событии к его же вероятности до сообщения.

Из формулы (4.6) видно, что если вероятность события $X\sim x_i$ в результате сообщения $Y\sim y_j$ увеличивается, т.е.

$$p(x_i/y_j)>p_i,$$

то информация $I(y_j,x_i)$ положительна. В противном случае она отрицательна.

В частном случае, когда появление события $Y\sim y_j$ полностью исключает возможность появления события $X\sim x_i$ (т.е. когда эти события несовместны), то

$$I(y_j,x_i)=-\infty.$$

Т.к. $I(y_j,x_i)=log\dfrac{p(x_i/y_j)}{p_i}=log\dfrac{p_{ij}}{p_i r_j}$, то частная информация симметрична относительно $x_i$ и $y_j$. Следовательно,

$$I(x_i, y_j) = I(y_j, x_i).$$

Таким образом, ввели три вида информации:

1. $I(Y, X)$ — полная информация о системе $X$, содержащаяся в системе $Y$.
2. $I(y_j, X)$ — частная информация о системе $X$, содержащаяся в событии (сообщении) $Y \sim y_j$.
3. $I(y_j, x_i)$ — частная информация о событии $X \sim x_i$, содержащаяся в событии $Y \sim y_j$.

Таким образом, информацию можно измерять длиной сообщения в битах. Такой способ ничего не говорит об информативности сообщения, но характеризует объём работы системы связи при передаче. Если же в задаче необходимо учитывать информативность, то следует пользоваться энтропийным подходом к измерению информации. При этом нужно уточнить, о каком множестве событий будет сообщаться, каковы их вероятности.

## 4.6 Информационные характеристики каналов связи

Основными характеристиками информационных систем являются такие понятия, как энтропия, скорость передачи информации, избыточность, пропускная способность канала связи. Для организации передачи информации по каналам связи надо учитывать не только количество информации, но и обеспечение передачи его в более короткий срок, не только хранение определенного количества, но и хранение с помощью минимального объема аппаратуры и т.д.

Предположим, что по каналу связи передали за время $T$ количество информации, которое равно

$$I_T = H_T - H_T(Y/X).$$

Можно вычислить скорость передачи данных

$$\vartheta = \frac{I_T}{T} = \frac{1}{T}(H_T(X) - H_T(X/Y)) = H(X) - H(X/Y).$$

Скорость передачи данных представляет собой количество информации, приходящееся на одно сообщение. Если количество сообщений равно $n$, то скорость передачи $n$ сообщений в секунду будет определяться

$$\vartheta = n(H(X) - H(X/Y)).$$

В этом случае максимальная пропускная способность данного канала представляет собой $c$

$$c = max\vartheta = n(H(X) - H(X/Y))_{max} = n \cdot I(X, Y)_{max}.$$

Различают техническую скорость передачи и информационную скорость передачи информации.

*Технической* скоростью передачи информации по каналу связи называется число символов, передаваемых в единицу времени

$$\vartheta_T = \frac{1}{\tau} \text{ бод}^5.$$

Информационной скоростью передачи информации по каналу связи называется среднее количество информации, которое передается в единицу времени

$$\vartheta = nH \text{ бит/с}.$$

Если сообщения являются равновероятными, составленные из равновероятных взаимно независимых символов, то их информационная скорость равна

$$\vartheta = \frac{1}{\tau} log m.$$

Если символы в сообщении не равновероятны, то скорость определяется

$$\vartheta = -\frac{1}{\tau} \sum_i p_i log_2 p_i.$$

В случае, если символы имеют разную длительность

$$\vartheta = -\frac{\sum_i p_i log_2 p_i}{\sum_i \tau_i p_i}.$$

Пропускная способность канала передачи информации характеризуется максимальной энтропией

$$c_{max} = \frac{H_{max}}{\tau} \text{ бит/с}.$$

Для двоичного кода

$$c_{max} = \frac{log_2 2}{\tau} = \frac{1}{\tau} \text{ бит/с}.$$

---

5 бод (англ. *baud*) в связи и электронике — единица измерения символьной скорости, количество изменений информационного параметра несущего периодического сигнала в секунду. Названа по имени Эмиля Бодо, изобретателя кода Бодо — кодировки символов для телетайпов.

***1 Теорема Шеннона (для канала с помехами)***

Пусть есть источник информации с энтропией $H(X)$ и канал связи с пропускной способностью $c$, то если $c > H(X)$, то всегда можно закодировать достаточно длинное сообщение таким образом, что оно будет передано без задержек.

Если $c < H(X)$, то передача информации без задержек невозможна.

Для всех практических каналов характерно наличие помех. Но есть случаи, когда помехи малы, то вероятность искажения передаваемого сообщения равна нулю и можно считать, что все сигналы передаются верно. В этом случае среднее количество информации, переносимое одним символом

$$I(X,Y) = I(Y,X) = H(X),\ H_{max} = log_2 m.$$

Следовательно, пропускная способность канала без помех за единицу времени

$$c = n log_2 m.$$

| $x_i$ | $x_1$ | $x_2$ |
|---|---|---|
| $p_i$ | 0,5 | 0,5 |

| $y_j$ | $y_1$ | $y_2$ |
|---|---|---|
| $p_j$ | 0,6 | 0,4 |

Реальные каналы характеризуются тем, что в них всегда есть помехи. Пропускная способность дискретного канала с помехами вычисляется по формуле

$$c = n(H(Y) - H(Y/X))_{max}.$$

***2 Теорема Шеннона***

Если источник сообщений имеет энтропию $H(X)$, а канал связи – пропускную способность $c$, то можно закодировать сообщения таким образом, чтобы передавать информацию по каналу со средней скоростью, сколь угодно близкой к величине $c$, но не превзойти её.

К. Шеннон предложил и метод такого кодирования, который получил название статистического или оптимального кодирования. В дальнейшем идея такого кодирования была развита в работах Фано и Хаффмана и в настоящее время широко используется на практике для «сжатия сообщений».

# 5. Код, кодировка

## 5.1 Общие понятия теории кодирования информации

В вычислительных машинах символы не могут храниться иначе, как в виде последовательностей бит (как и числа). Для передачи символа и его корректного отображения ему должна соответствовать уникальная последовательность нулей и единиц. Для этого были разработаны таблицы кодировок.

Преобразование информации из одной формы представления в другую называют ***кодированием***. Средством кодирования служит таблица соответствия знаковых систем, которая устанавливает взаимно-однозначное соответствие между знаками или группой знаков двух различных знаковых систем.

В процессе обмена информацией часто приходится производить операции кодирования и декодирования информации. Например, при вводе знака алфавита в компьютер путем нажатия соответствующей клавиши на клавиатуре происходит кодирование знака, т.е. преобразование его в компьютерный код. При выводе знака на экран монитора или принтер происходит обратный процесс – декодирование, когда из компьютерного кода знак преобразуется в его графическое изображение.

*Кодирование* – это операция преобразования знаков или групп знаков одной знаковой системы в знаки или группы знаков другой знаковой системы.

При кодировании информации ставятся следующие цели:

1) удобство физической реализации;
2) удобство восприятия;
3) высокая скорость передачи и обработки;
4) экономичность, т.е. уменьшение избыточности сообщения;
5) надежность, т.е. защита от случайных искажений;
6) сохранность, т.е. защита от нежелательного доступа к информации.

Эти цели часто противоречат друг другу. Например, стремясь к экономным сообщениям, мы тем самым уменьшаем их надежность и удобство восприятия.

На разных этапах обработки информации достигаются разные цели и поэтому информация неоднократно перекодируется, преобразуется из вида, удобного для восприятия человеком, к виду, удобному для обработки автоматическими системами, и наоборот.

В компьютере для представления информации используется двоичное кодирование, т.к. используются технические устройства, которые могут сохранять и распознавать не более двух различных состояний (цифр):

- электромагнитные реле (замкнуто/разомкнуто), широко использовались в конструкциях первых ЭВМ;
- участок поверхности магнитного носителя информации (намагничен/размагничен);
- участок поверхности лазерного диска (отражает/не отражает) и т.д.

Все виды информации в компьютере кодируются на машинном языке, в виде логических последовательностей нулей и единиц. Как кодируются числа в двоичной системе, было рассмотрено при рассмотрении систем счисления.

Число элементов символа кода, используемое для представления одного символа алфавита исходного источника сообщений, называют ***значностью кода***.

Если значность кода одинакова для всех символов алфавита исходного сообщения, то код называют равномерным, в противном случае – неравномерным.

Число элементов входящих в кодовый символ иногда называют длиной кодового символа. С точки зрения избыточности все коды можно разделить на ***неизбыточные коды*** и ***избыточные***. В избыточных кодах число элементов кодовых символов может быть сокращено за счет более эффективного использования оставшихся элементов, в неизбыточных же кодах сокращение числа элементов в кодовых символах невозможно.

Задачи кодирования при отсутствии помех и при их наличии существенно различны. Поэтому различают

- эффективное (статистическое) кодирование,
- и корректирующее (помехоустойчивое) кодирование.

При эффективном кодировании ставится задача добиться представления символов алфавита источника сообщений минимальным числом элементов кодовых символов в среднем на один символ алфавита источника сообщений за счет уменьшения избыточности кода, что ведет к повышению скорости передачи сообщения. А при корректирующем (помехоустойчивом) кодировании ставится задача снижения вероятности ошибок в передаче символов исходного алфавита путем обнаружения и исправления ошибок за счет введения дополнительной избыточности кода.

Отдельно стоящей задачей кодирования является защита сообщений от несанкционированного доступа, искажения и уничтожения их. При этом виде кодирования кодирование сообщений осуществляется таким образом, чтобы даже получив их, злоумышленник не смог бы их раскодировать. Процесс такого вида кодирования сообщений называется ***шифрованием*** (или зашифровкой), а процесс декодирования – *расшифрованием* (или расшифровкой). Само кодированное сообщение называют шифрованным (или просто шифровкой), а применяемый метод кодирования – шифром.

Довольно часто в отдельный класс выделяют методы кодирования, которые позволяют построить (без потери информации) коды сообщений, имеющие меньшую длину по сравнению с исходным сообщением. Такие методы кодирования называют методами сжатия или упаковки данных. Качество сжатия определяется коэффициентом сжатия, который обычно измеряется в процентах и который показывает на сколько процентов кодированное сообщение короче исходного.

При автоматической обработке информации с использованием ЭВМ как правило используют числовое (цифровое) кодирование, при этом, естественно, возникает вопрос обоснования используемой системы счисления. Действительно, при уменьшении основания системы счисления упрощается алфавит элементов символов кода, но происходит удлинение символов кода. С другой стороны, чем больше основание системы счисления, тем меньшее число разрядов требуется для представления одного символа кода, а, следовательно, и меньшее время для его передачи, но с ростом основания системы счисления существенно повышаются требования к каналам связи и техническим средствам распознавания элементарных сигналов, соответствующих различным элементам символов кода. В частности, код числа, записанного в двоичной системе счисления в среднем приблизительно в 3,5

раза длиннее десятичного кода. Так как во всех системах обработки информации приходится хранить большие информационные массивы в виде числовой информации, то одним из существенных критериев выбора алфавита элементов символов числового кода (т.е. основания используемой системы счисления) является минимизация количества электронных элементов в устройствах хранения, а также их простота и надежность.

***Кодирование текстовой информации***

Если каждому символу алфавита сопоставить определенное целое число (например, порядковый номер), то с помощью двоичного кода можно кодировать и текстовую информацию.

Если для кодирования одного символа использовать 1 байт = 8 бит информации, то с помощью одного байта можно закодировать $N = 2^N = 256, N = 8$ символов.

Такое количество символов вполне достаточно для представления текстовой информации, включая прописные и строчные буквы русского и латинского алфавита, цифры, знаки основных арифметических операций и некоторые общепринятые специальные символы, например, символ «§». При этом кодирование заключается в установлении однозначного соответствия символа и уникального десятичного кода от 0 до 255 или двоичного кода от 00000000 до 11111111. Если человек различает символы по начертаниям, то компьютер по их кодам.

Присвоение символу конкретного кода – это вопрос соглашения, которое фиксируется в кодовой таблице.

Институтом стандартизации США (ANSI – American National Standard Institute) в 1963 году была введена в действие система кодирования ASCII (American Standard Code for Information Interchange – стандартный код информационного обмена США). В системе ASCII закреплены две таблицы кодирования – базовая (рис. 14) и расширенная (рис. 15). Базовая таблица закрепляет значения кодов от 0 до 127, а расширенная относится к символам с номерами от 128 до 255.

| Dec | Hex | Char | Cmd | Dec | Hex | Char | Cmd | Dec | Hex | Char | Cmd | Dec | Hex | Char | Cmd |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 00 | | NUL | 32 | 20 | | (sp) | 64 | 40 | @ | | 96 | 60 | ` | |
| 1 | 01 | ☺ | SOH | 33 | 21 | ! | | 65 | 41 | A | | 97 | 61 | a | |
| 2 | 02 | ☻ | STX | 34 | 22 | " | | 66 | 42 | B | | 98 | 62 | b | |
| 3 | 03 | ♥ | ETX | 35 | 23 | # | | 67 | 43 | C | | 99 | 63 | c | |
| 4 | 04 | ♦ | EOT | 36 | 24 | $ | | 68 | 44 | D | | 100 | 64 | d | |
| 5 | 05 | ♣ | ENQ | 37 | 25 | % | | 69 | 45 | E | | 101 | 65 | e | |
| 6 | 06 | ♠ | ACK | 38 | 26 | & | | 70 | 46 | F | | 102 | 66 | f | |
| 7 | 07 | • | BEL | 39 | 27 | ' | | 71 | 47 | G | | 103 | 67 | g | |
| 8 | 08 | ◘ | BS | 40 | 28 | ( | | 72 | 48 | H | | 104 | 68 | h | |
| 9 | 09 | ○ | TAB | 41 | 29 | ) | | 73 | 49 | I | | 105 | 69 | i | |
| 10 | 0A | ◙ | LF | 42 | 2A | * | | 74 | 4A | J | | 106 | 6A | j | |
| 11 | 0B | ♂ | VT | 43 | 2B | + | | 75 | 4B | K | | 107 | 6B | k | |
| 12 | 0C | ♀ | FF | 44 | 2C | , | | 76 | 4C | L | | 108 | 6C | l | |
| 13 | 0D | ♪ | CR | 45 | 2D | - | | 77 | 4D | M | | 109 | 6D | m | |
| 14 | 0E | ♫ | SO | 46 | 2E | . | | 78 | 4E | N | | 110 | 6E | n | |
| 15 | 0F | ☼ | SI | 47 | 2F | / | | 79 | 4F | O | | 111 | 6F | o | |
| 16 | 10 | ► | DLE | 48 | 30 | 0 | | 80 | 50 | P | | 112 | 70 | p | |
| 17 | 11 | ◄ | DC1 | 49 | 31 | 1 | | 81 | 51 | Q | | 113 | 71 | q | |
| 18 | 12 | ↕ | DC2 | 50 | 32 | 2 | | 82 | 52 | R | | 114 | 72 | r | |
| 19 | 13 | ‼ | DC3 | 51 | 33 | 3 | | 83 | 53 | S | | 115 | 73 | s | |
| 20 | 14 | ¶ | DC4 | 52 | 34 | 4 | | 84 | 54 | T | | 116 | 74 | t | |
| 21 | 15 | § | NAK | 53 | 35 | 5 | | 85 | 55 | U | | 117 | 75 | u | |
| 22 | 16 | ▬ | SYN | 54 | 36 | 6 | | 86 | 56 | V | | 118 | 76 | v | |
| 23 | 17 | ↨ | ETB | 55 | 37 | 7 | | 87 | 57 | W | | 119 | 77 | w | |
| 24 | 18 | ↑ | CAN | 56 | 38 | 8 | | 88 | 58 | X | | 120 | 78 | x | |
| 25 | 19 | ↓ | EM | 57 | 39 | 9 | | 89 | 59 | Y | | 121 | 79 | y | |
| 26 | 1A | → | SUB | 58 | 3A | : | | 90 | 5A | Z | | 122 | 7A | z | |
| 27 | 1B | ← | ESC | 59 | 3B | ; | | 91 | 5B | [ | | 123 | 7B | { | |
| 28 | 1C | ∟ | FS | 60 | 3C | < | | 92 | 5C | \ | | 124 | 7C | \| | |
| 29 | 1D | ↔ | GS | 61 | 3D | = | | 93 | 5D | ] | | 125 | 7D | } | |
| 30 | 1E | ▲ | RS | 62 | 3E | > | | 94 | 5E | ^ | | 126 | 7E | ~ | |
| 31 | 1F | ▼ | US | 63 | 3F | ? | | 95 | 5F | _ | | 127 | 7F | ⌂ | DEL |

Рисунок 14 – Базовая таблица кодировки ASCII[6]

В таблице приведены ASCII-символы (Char) и их коды в десятичной (Dec) и шестнадцатиричной (Hex) системах счисления. Некоторые коды (99-32h, 7Fh) могут использоваться и в качестве команд (Cmd)[7]

Первые 32 кода (0-31) базовой таблицы отданы производителям аппаратных средств (компьютеров и печатающих устройств). В этой области размещаются так называемые управляющие коды, которым не соответствуют никакие символы языков, и, соответственно, эти коды не выводятся ни на экран, ни на устройства печати, но ими можно управлять тем, как производится вывод других данных или действия с другими данными (например, код 8 соответствует возврату на один символ назад (BACKSPACE)).

Начиная с кода 32 по код 127 размещены коды символов английского алфавита, знаков препинания, цифр, арифметических действий и некоторых вспомогательных символов.

Поддержка производителей оборудования и программ вывела американский код ASCII на уровень международного стандарта, и национальным системам кодирования пришлось «отступить» во вторую, расширенную часть системы кодирования, определяющую значения кодов со 128 по 255 (рис. 15).

[6] https://ru.wikipedia.org/wiki/ASCII

[7] https://www.industrialnets.ru/files/misc/ascii.pdf

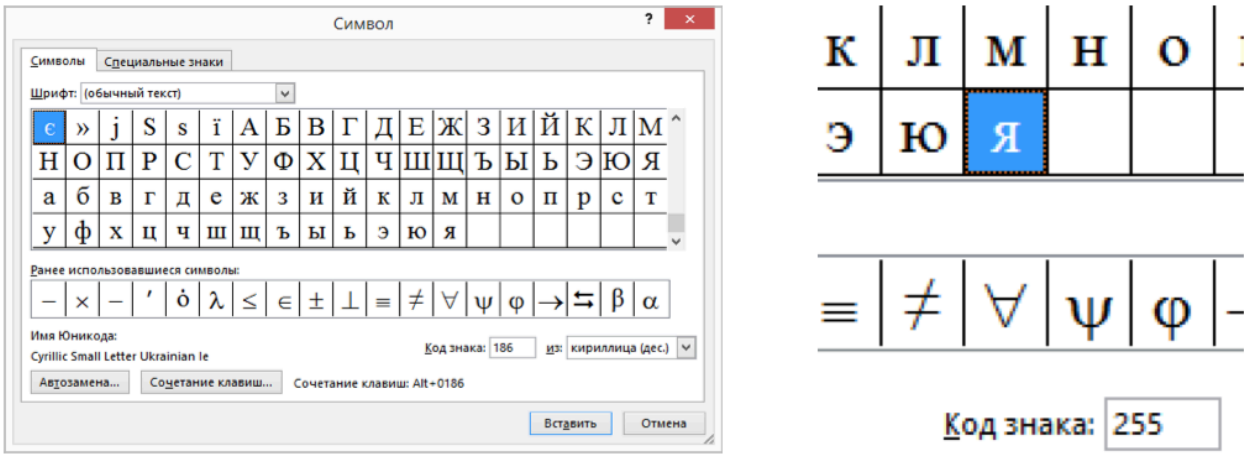



Рисунок 15 – Расширенная таблица кодировки ASCII (кириллица)

| Стандартная часть таблицы ASCII | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| № п/п | символ | двоичный код | № п/п | символ | двоичный код | № п/п | символ | двоичный код | № п/п | символ | двоичный код |
| 32 | пробел | 00100000 | 56 | 8 | 00111000 | 80 | P | 01010000 | 104 | h | 01101000 |
| 33 | ! | 00100001 | 57 | 9 | 00111001 | 81 | Q | 01010001 | 105 | i | 01101001 |
| 34 | " | 00100010 | 58 | : | 00111010 | 82 | R | 01010010 | 106 | j | 01101010 |
| 35 | # | 00100011 | 59 | ; | 00111011 | 83 | S | 01010011 | 107 | k | 01101011 |
| 36 | $ | 00100100 | 60 | < | 00111100 | 84 | T | 01010100 | 108 | l | 01101100 |
| 37 | % | 00100101 | 61 | | 00111101 | 85 | U | 01010101 | 109 | m | 01101101 |
| 38 | & | 00100110 | 62 | > | 00111110 | 86 | V | 01010110 | 110 | n | 01101110 |
| 39 | ' | 00100111 | 63 | ? | 00111111 | 87 | W | 01010111 | 111 | o | 01101111 |
| 40 | ( | 00101000 | 64 | @ | 01000000 | 88 | X | 01011000 | 112 | p | 01110000 |
| 41 | ) | 00101001 | 65 | A | 01000001 | 89 | Y | 01011001 | 113 | q | 01110001 |
| 42 | * | 00101010 | 66 | B | 01000010 | 90 | Z | 01011010 | 114 | r | 01110010 |
| 43 | + | 00101011 | 67 | C | 01000011 | 91 | [ | 01011011 | 115 | s | 01110011 |
| 44 | , | 00101100 | 68 | D | 01000100 | 92 | \ | 01011100 | 116 | t | 01110100 |
| 45 | - | 00101101 | 69 | E | 01000101 | 93 | ] | 01011101 | 117 | u | 01110101 |
| 46 | . | 00101110 | 70 | F | 01000110 | 94 | ^ | 01011110 | 118 | v | 01110110 |
| 47 | / | 00101111 | 71 | G | 01000111 | 95 | _ | 01011111 | 119 | w | 01110111 |
| 48 | 0 | 00110000 | 72 | H | 01001000 | 96 | ' | 01100000 | 120 | x | 01111000 |
| 49 | 1 | 00110001 | 73 | I | 01001001 | 97 | a | 01100001 | 121 | y | 01111001 |
| 50 | 2 | 00110010 | 74 | J | 01001010 | 98 | b | 01100010 | 122 | z | 01111010 |
| 51 | 3 | 00110011 | 75 | K | 01001011 | 99 | c | 01100011 | 123 | { | 01111011 |
| 52 | 4 | 00110100 | 76 | L | 01001100 | 100 | d | 01100100 | 124 | \| | 01111100 |
| 53 | 5 | 00110101 | 77 | M | 01001101 | 101 | e | 01100101 | 125 | } | 01111101 |
| 54 | 6 | 00110110 | 78 | N | 01001110 | 102 | f | 01100110 | 126 | ~ | 01111110 |
| 55 | 7 | 00110111 | 79 | O | 01001111 | 103 | g | 01100111 | 127 | | 01111111 |

Отсутствие единого стандарта в этой области привело к множественности одновременно действующих кодировок. Только в России можно указать три действующих стандарта кодировки (Windows-1251[8], КОИ-8[9], ISO[10]).

Если проанализировать организационные трудности, связанные с созданием единой системы кодирования текстовых данных, то можно прийти к выводу, что они вызваны ограниченным набором кодов (всего 256). В то же время очевидно, что если, например, кодировать символы не восьмиразрядными двоичными числами, а числами с большим количеством разрядов, то и диапазон возможных значений кодов станет намного большим. Такая система, основанная на 16-разрядном кодировании символов, получила название универсальной – UNICODE. Шестнадцать разрядов позволяют обеспечить уникальные коды для $N = 2^{16} = 65536$ различных символов – этого количества кодов достаточно для размещения в одной таблице символов большинства языков планеты.

Несмотря на тривиальную очевидность такого подхода, простой механический переход на данную систему долгое время сдерживался из-за недостаточных ресурсов средств вычислительной техники (в системе кодирования UNICODE все текстовые документы автоматически становятся вдвое длиннее). Во второй половине 90-х годов прошлого века технические средства достигли необходимого уровня обеспеченности ресурсами, и сегодня идет постепенный перевод документов и программных средств на универсальную систему кодирования. Эту кодировку поддерживают, начиная с 1997 года, Microsoft Windows&Office.

***Кодирование графической информации***

Существует несколько способов кодирования графической информации.

При рассмотрении графического изображения с помощью увеличительного стекла заметно, что в его состав входит несколько мельчайших точек, образующих характерный узор (или растр) (рис. 16).


Рисунок 16 – Растровое изображение

[8] https://ru.wikipedia.org/wiki/Windows-1251

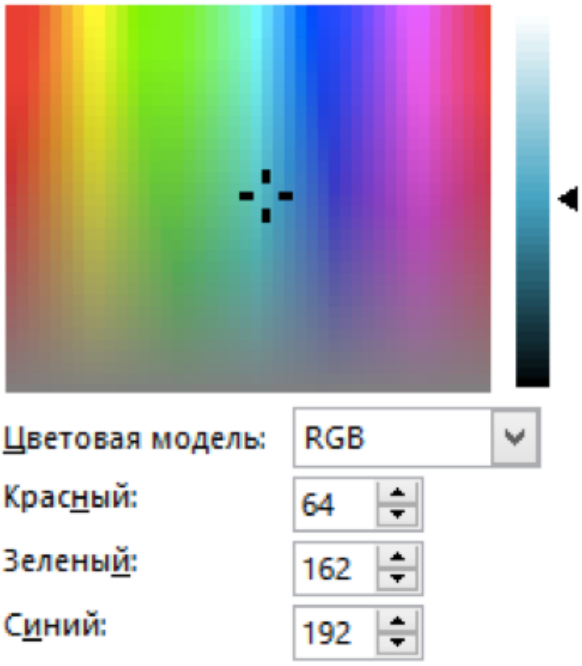
[9] https://ru.wikipedia.org/wiki/КОИ-8

[10] ISO (International Standard Organization – Международный институт стандартизации) - международный стандарт, в котором предусмотрена кодировка символов русского алфавита.

Растровое компьютерное изображение состоит из пикселей. ***Пиксель***, пиксел (англ. pixel – сокращение от pix element) – физическая точка или наименьший логический элемент двумерного цифрового изображения.

Линейные координаты и индивидуальные свойства каждой из точек изображения можно выразить с помощью целых чисел, поэтому способ растрового кодирования базируется на использовании двоичного кода представления графических данных. Общеизвестным стандартом считается приведение черно-белых иллюстраций в форме комбинации точек с 256 градациями серого цвета, т. е. для кодирования яркости любой точки необходимы 8-разрядные двоичные числа.

В основу кодирования цветных графических изображений положен принцип разложения произвольного цвета на основные составляющие (принцип декомпозиции), в качестве которых применяются три основных цвета: красный (Red), зеленый (Green) и синий (Blue). На практике принимается, что любой цвет, который воспринимает человеческий глаз, можно получить с помощью механической комбинации этих трех цветов. Такая система кодирования называется RGB (по первым буквам основных цветов).




Если для кодирования яркости каждой из основных составляющих использовать по 256 значений (это 8 бит), то на кодирование цвета одной точки требуется 24 бита. С помощью 24 бит можно закодировать более 16,5 млн. различных цветов ($2^{24} = 16777216$).

Режим представления цветной графики с использованием 24 двоичных разрядов называется ***полноцветным (True Color)***.

Каждому из основных цветов можно сопоставить дополнительный цвет, т.е. цвет, дополняющий основной цвет до белого. Для любого из основных цветов дополнительным будет являться цвет, который образован суммой пары остальных основных цветов. Соответственно среди дополнительных цветов можно выделить голубой (Cyan, C), пурпурный (Magenta, M) и желтый (Yellow, K). Принцип разложения произвольного цвета на составляющие компоненты используется не только для основных цветов, но и для дополнительных, т. е. любой цвет можно представить в виде суммы голубой, пурпурной и желтой составляющей. Этот метод кодирования цвета применяется в полиграфии, но там используется еще и четвертая краска – черная (Black, K), поэтому эта система кодирования обозначается четырьмя буквами – CMYK. Для представления цветной графики в этой системе применяется 32 двоичных разряда. Данный режим также носит название полноцветного.

При уменьшении количества двоичных разрядов, применяемых для кодирования цвета каждой точки, сокращается объем данных, но заметно уменьшается диапазон кодируемых цветов. Кодирование цветной графики 16-разрядными двоичными числами носит название режима High Color.

При кодировании графической цветной информации с применением 8 бит данных можно передать только 256 оттенков. Данный метод кодирования цвета называется индексным. Код каждой точки растра выражает не цвет сам по себе, а только его номер (индекс) в некоторой справочной таблице, называемой палитрой. Эта таблица должна прикладываться к графическим данным, так как без нее нельзя адекватно воспроизвести информацию на экране или принтере. [Видео]

## 5.2 Оптимальное кодирование информации

Преобразование информации из одной формы представления (знаковой системы) в другую называется кодированием. Все виды информации в вычислительных системах кодируются на машинном языке в виде последовательностей логических нуля и единицы.

Двоичный код можно представить как два равновероятных состояния: «0» и «1». Количество информации любого двоичного символа равно одному биту. Два символа двоичного кода несут информацию в 2 бита, три цифры – в 3 бита и т.д. Количество информации в битах равно количеству цифр двоичного машинного кода. Восемь последовательных бит равно одному байту. В одном байте можно закодировать значение одного символа из 256 возможных.

Одной из основных проблем кодирования при передаче информации по каналам связи является оптимальный способ кодирования, когда на передачу сообщения уходит минимальное время.

Предположим, что на передачу каждого элементарного символа (0 или 1) затрачивается одно и тоже время, то оптимальным будет такой код, при котором на передачу сообщения заданной длины будет затрачено минимальное количество символов.

Например, пусть имеются буквы русского алфавита а, б, в, г,...+ промежуток между словами (-). Если не различать ь и ъ (как принято в телеграфии), то получим 32 буквы. Требуется закодировать двоичным кодом буквы так, чтобы каждой букве соответствовала определенная комбинация символов 0 и 1 и, чтобы среднее число этих символов на букву текста было минимальным.

*Вариант 1*. Не меняя порядка букв, пронумеровав их от 0 до 31 и перевести их в двоичную систему счисления, получим следующий код:

а ~ 00000
б ~ 00001
в ~ 00010
г ~ 00011
.....
я ~ 11110
- ~ 11111

В этом коде на каждую букву тратится ровно пять элементарных символов. Является ли этот код оптимальным? Можно ли составить другой код, при котором на одну букву в среднем приходится меньше элементарных символов?

*Вариант 2*. Так как одни буквы встречаются часто (а, о, е), а другие (щ, э, ф) редко, то часто встречающиеся буквы целесообразно закодировать меньшим числом символов, а реже встречающиеся – большим. Чтобы составить такой код нужно знать частоты букв русского алфавита (таблица ) . Пользуясь такой таблицей, можно составить наиболее экономичный код на основе соображений, связанных с количеством информации. Код будет самым экономичным, когда каждый символ будет передавать максимальную информацию.

Таблица 5.1 - Статистические данные русского алфавита

| буква | частота | буква | частота | буква | частота |
|---|---|---|---|---|---|
| - | 0,145 | к | 0,029 | ч | 0,013 |
| 0 | 0,095 | м | 0,026 | й | 0,01 |
| е | 0,074 | д | 0,026 | х | 0,009 |
| а | 0,064 | п | 0,024 | ж | 0,008 |
| и | 0,064 | у | 0,021 | ю | 0,007 |
| т | 0,056 | я | 0,019 | ш | 0,006 |
| н | 0,056 | ы | 0,016 | щ | 0,003 |
| с | 0,047 | з | 0,015 | э | 0,003 |
| р | 0,041 | ъ,ь | 0,015 | ф | 0,002 |
| в | 0,039 | б | 0,015 | | |
| л | 0,036 | г | 0,014 | | |

### 5.2.1 Метод Шеннона-Фано

Метод Шеннона-Фано является методом оптимального кодирования. Алгоритм построения кода Шеннона-Фано состоит в том, что кодируемые символы (буквы) разделяются на две равновероятные подгруппы: для символов 1-й подгруппы на втором месте ставится 0, а для 2-й подгруппы – 1 и т.д. Возможен другой вариант, когда первая подгруппа соответствует «1», а вторая – «0». Главное, определить правило изначально и следовать ему на протяжении всего процесса кодирования.

Необходимо взять первые шесть букв (от – до т). Сумма их вероятностей равна 0,498, на все остальные (от н до ф) приходится 0,502. Первые шесть букв будут иметь на первом месте 0, остальные 1. Далее снова первая группа делится на две приблизительные равновероятные подгруппы: (от – до щ) и (от е до т) и т.д.

Для всех букв первой подгруппы на втором месте ставится 0, а второй подгруппы – 1. Процесс продолжается до тех пор, пока в каждой группе не останется ровно одна буква, которая и будет закодирована определенным двоичным кодом.

*Пример.* Имеется алфавит символов и их вероятности, с которыми они встречаются в тексте. Построить таблицу кодов символов методом Шеннона-Фано. Закодировать сообщение «вилка» и раскодировать заданную последовательность кодов.

| а | в | л | и | е | с | к |
|---|---|---|---|---|---|---|
| 0,3 | 0,2 | 0,15 | 0,1 | 0,1 | 0,08 | 0,07 |

*Решение.* Составим таблицу кодов для символов алфавита

| буква | вероятности | символы кода | | | | код |
|---|---|---|---|---|---|---|
| а | 0,3 | 0 | 0 | | | 00 |
| в | 0,2 | | 1 | | | 01 |
| л | 0,15 | 1 | 0 | 0 | | 100 |
| и | 0,1 | | | 1 | | 101 |
| е | 0,1 | | 1 | 0 | | 110 |
| с | 0,08 | | | 1 | 0 | 1110 |
| к | 0.07 | | | | 1 | 1111 |

Сообщению «вилка» соответствует выходная последовательность кодов 01101100111100. Выходной последовательности кодов 100101111000 соответствует сообщение «лиса».

### 5.2.2 Метод Хаффмана

*Метод Хаффмана* – замена кода равной длины для символов на коды неравной длины в соответствии с частотой появления символов в данных, коды минимальной длины присваиваются наиболее часто встречающимся символам. Если частоты появления символов являются степенью двойки ($2n$), то этот метод достигает теоретической энтропийной границы эффективности сжатия для методов такого типа. [Видео]

*Пример*. Создадим дерево Хаффмана для предложения «*How much wood could a woodchuck chuck?»* Вычислим количество появлений символов этого предложения и представили их на рис. 18. Выберем два узла с наименьшими значениями. Существует несколько узлов, из которых можно выбрать, но мы выберем узлы «*m*» и «*?*». Для обоих этих узлов число появлений символов равно 1. Создадим родительский узел, значение счетчика которого равно 2, и присоединим к нему два выбранных узла в качестве дочерних. Поместим родительский узел обратно в пул. Повторим цикл с самого начала. На этот раз мы выбираем узлы «*a*» и «*l*», объединяем их в мини-дерево и помещаем родительский узел (значение счетчика которого снова равно 2) обратно в пул. Снова повторим цикл. На этот раз в нашем распоряжении имеется единственный узел, значение счетчика которого равно 1 (узел «*h*») и три узла со значениями счетчиков, равными 2 (узел «*k*» и два родительских узла, которые были добавлены перед этим). Выберем узел «*k*», присоединим его к узлу «*h*» и снова добавим в пул родительский узел, значение счетчика которого равно 3. Затем выберем два родительских узла со значениями счетчиков, равными 2, присоединим их к новому родительскому узлу со значением счетчика, равным 4, и добавим этот родительский узел в пул. Несколько первых шагов построения дерева Хаффмана и результирующее дерево показаны на рис. 17.

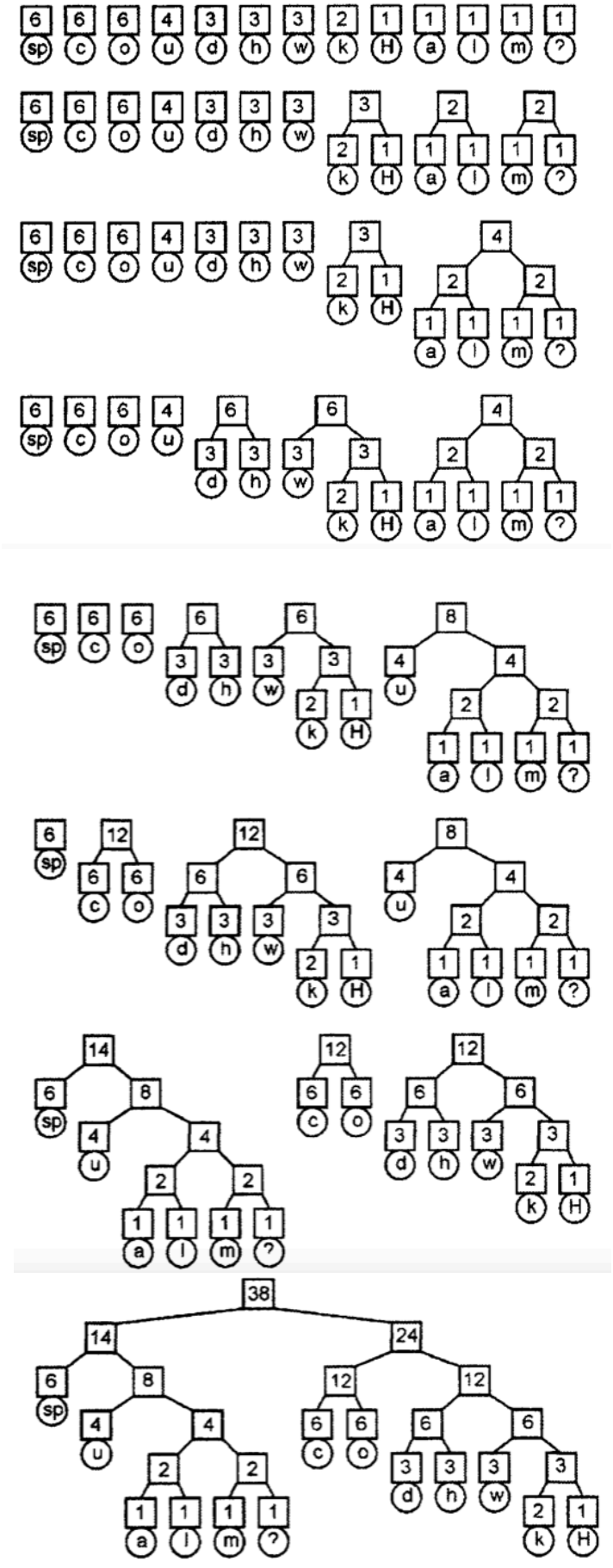
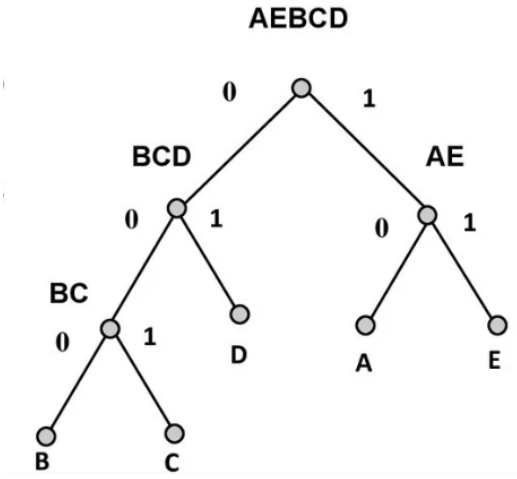



Рисунок 17 – Построение дерева Хаффмана

***Декодирование кода Хаффмана***

При приеме символа кодового слова декодирование начинается с начальной точки отсчета дерева (с корня дерева). Если входному символу соответсвует значение 1, следует двигаться по ветви с присвоенным значением 1. Если принимается 0, следует идти по ветви, соответствующей значению 0. При попадании на конечны узел дерева принимается решение о принятом символе. При попадании в узел, из которого выходят две ветви, следующий принятый символ (0 или 1) указывает, по какой ветви следует двигаться. Движение по дереву продолжается до достижения конечного узла. Для наглядности процесса декодирования кодового слова (001) изобразим кодовое дерево декодера источника $X = \{A, E, B, C, D\}$. Очевидно, декодирование кодовой последовательности (001) приводит к символу $C$.




| Таблица кодирования | | | | |
|---|---|---|---|---|
| **A** | **B** | **C** | **D** | **E** |
| **10** | **000** | **001** | **01** | **11** |

Рисунок 18 – Кодовое дерево декодера источника $X = \{A, E, B, C, D\}$

Передача информации посредством эффективного кодирования требует использования каналов, характеризующихся повышенной надежностью. Вероятность ошибки в таком канале должна быть сравнительно малой. Даже одиночная ошибка в потоке кодированной информации приводит к неправильному декодированию сообщения источника.

Адаптивный алгоритм Хаффмана. Адаптивный алгоритм эффективного кодирования реализуется с помощью двух операций:

1. Вначале выполняется кодирование источника в предположении, что все символы имеют равные вероятности появления.

2. По мере накопления знаний о статистических характеристиках источника выполняется кодирование по алгоритму Хаффмана[11].

[11] Митюхин, А.И. Прикладная теория информации : учеб.-метод. пособие / А. И. Митюхин. - Минск : БГУИР, 2018. - 168 с. : ил. ISBN 978-985-543-356-0.

### 5.2.3 Избыточность и оптимальное кодирование

Если энтропия источника сообщений не равна максимальной энтропии для алфавита с заданным количеством качественных признаков, то это означает, что сообщения данного источника могли нести большее количество информации. Абсолютная недогруженность на символ такого источника

$$\Delta D = (H_{max} - H) \text{ бит/символ.}$$

Для определения количества «лишней» информации, которая заложена в структуре алфавита либо в природе кода, вводится понятие ***избыточности.*** *Информационная избыточность* показывает относительную недогруженность на символ алфавита и является безразмерной величиной

$$\Delta D = \frac{(H_{max} - H)}{H_{max}} = 1 - \frac{H}{H_{max}},$$

где $\frac{H}{H_{max}} = \mu$ – коэффициент сжатия (относительная энтропия).

Кроме общего понятия избыточности есть частные виды избыточности. Избыточность, обусловленная неравновероятным распределением символов в сообщении

$$D_p = 1 - (\frac{-\sum_i p_i log p_i}{log m}),$$

где – $m$ число знаков в алфавите.

Избыточность, вызванная статистической связью между символами сообщения

$$D_s = 1 - (\frac{-\sum_i \sum_j p(x_i) p(y_j/x_i) log p(y_j/x_i)}{-\sum_i p_i log p_i}).$$

Полная информационная избыточность

$$D = D_s + D_p - D_s D_p.$$

Для данного способа кодирования характерна избыточность. Это объясняется тем, что в результате неравномерного распределения качественных признаков этого кода не может быть задана одной цифрой на основании статистических испытаний. При передаче десятичных цифр двоичным кодом максимально загруженными бывают только те символы вторичного алфавита, которые передают значения, являющиеся целочисленными степенями двойки. В остальных случаях тем же количеством символов может быть передано большее количество цифр (сообщений). Фактически для передачи сообщения достаточно иметь длину кодовой комбинации

$$L \geq \frac{logN}{logm},$$

где $N$ – общее количество передаваемых сообщений. $L$ можно представить как

$$L \geq \frac{logm_1}{logm_2},$$

где $m_1$ и $m_2$ – соответственные качественные признаки первичного и вторичного алфавитов. Поэтому для цифры 5 в двоичном коде $L \geq log5/log2 = 2,32$ дв. символа. Однако эту цифру надо округлить в большую сторону до ближайшего целого числа, так как длина кода не может быть выражена дробным числом. В общем случае избыточность от округления

$$D_0 = \frac{k - \varphi}{k},$$

где $\varphi = \frac{logm_1}{logm_2}$, $k$ – округленное до ближайшего целого числа значение.

В приведенном примере $D_0 = \frac{3 - 2,32}{3} = 0,227.$

Таким образом, избыточность может быть заложена как в первичном алфавите, так и в природе кода, составленного во вторичном алфавите.

Избыточность – не всегда нежелательное явление. Для повышения помехоустойчивости кодов избыточность необходима и ее вводят искусственно в виде добавочных символов. Наиболее эффективным способом уменьшения избыточности является построение оптимальных кодов. Оптимальными называются коды, которые имеют практически нулевую избыточностью. Оптимальные коды имеют минимальную среднюю длину кодовых слов – $L$. Верхняя и нижняя границы $L$ определяются из неравенства

$$\frac{H}{logm} \leq L \leq \frac{H}{logm} + 1,$$

где $H$ – энтропия первичного алфавита, $m$ – число качественных признаков вторичного алфавита.

Под термином «оптимальный код» понимают коды с практически нулевой избыточностью, так как сравнивают длину кодовой комбинации с энтропией источника сообщений, не учитывая взаимозависимость символов. С учетом взаимозависимости символов эффективность кодирования никогда не будет 100%.

При построении оптимальных кодов наибольшее распространение получили методы Шеннона-Фано и Хаффмана[12].

[12] Бурькова Е.В. Теория информации: методические указания / Е.В. Бурькова; Оренбургский гос. ун-т. – Оренбург: ОГУ, 2018. – 50 с.

### 5.2.4 Префиксные коды[13]

Рассмотрев методики построения эффективных кодов, нетрудно убедиться в том, что эффект достигается благодаря присвоению более коротких кодовых комбинаций более вероятным сообщениям и более длинных менее вероятным сообщениям. Таким образом, эффект связан с различием в числе символов кодовых комбинаций. А это приводит к трудностям при декодировании. Конечно, для различения кодовых комбинаций можно ставить специальный разделительный символ, но при этом значительно снижается эффект, которого мы добивались, так как средняя длина кодовой комбинации по существу увеличивается на символ.

Более целесообразно обеспечить однозначное декодирование без введения дополнительных символов. Для этого эффективный код необходимо строить так, чтобы ни одна комбинация кода не совпадала с началом более длинной комбинации. Коды, удовлетворяющие этому условию, называют ***префиксными кодами***. Последовательность 100000110110110100 комбинаций префиксного кода, например,

| $X_1$ | $X_2$ | $X_3$ | $X_4$ |
|---|---|---|---|
| 00 | 01 | 101 | 100 |

декодируется однозначно:

| 100 | 00 | 01 | 101 | 101 | 101 | 00 |
|---|---|---|---|---|---|---|
| $X_4$ | $X_1$ | $X_2$ | $X_3$ | $X_3$ | $X_3$ | $X_1$ |

Последовательность 000101010101 комбинаций непрефиксного кода, например,

| $X_1$ | $X_2$ | $X_3$ | $X_4$ |
|---|---|---|---|
| 00 | 01 | 101 | 010 |

(комбинация 01 является началом комбинации 010), может быть декодирована по-разному:

| 00 | 01 | 01 | 01 | 010 | 101 |
|---|---|---|---|---|---|
| $X_1$ | $X_2$ | $X_2$ | $X_2$ | $X_4$ | $X_3$ |

| 00 | 010 | 101 | 010 | 101 |
|---|---|---|---|---|
| $X_1$ | $X_4$ | $X_3$ | $X_4$ | $X_3$ |

| 00 | 01 | 010 | 101 | 01 | 01 |
|---|---|---|---|---|---|
| $X_1$ | $X_2$ | $X_4$ | $X_3$ | $X_2$ | $X_2$ |

Нетрудно убедиться, что коды, получаемые в результате применения методики Шеннона-Фано или Хаффмана, являются префиксными.

---

[13] Сорока Н. И., Кривинченко Г. А. ТЕОРИЯ ПЕРЕДАЧИ ИНФОРМАЦИИ / Н. И. Сорока Конспект лекций для студентов специальности 1-53 01 07 «Информационные технологии и управление в технических системах»

### 5.2.5 Недостатки системы эффективного кодирования

Причиной одного из недостатков является различие в длине кодовых комбинаций. Если моменты снятия информации с источника неуправляемы, кодирующее устройство через равные промежутки времени выдает комбинации различной длины. Так как линия связи используется эффективно только в том случае, когда символы поступают в нее с постоянной скоростью, то на выходе кодирующего устройства должно быть предусмотрено буферное устройство. Оно запасает символы по мере поступления и выдает их в линию связи с постоянной скоростью. Аналогичное устройство необходимо и на приемной стороне.

Второй недостаток связан с возникновением задержки в передаче информации. Наибольший эффект достигается при кодировании длинными блоками, а это приводит к необходимости накапливать знаки, прежде чем поставить им в соответствие определенную последовательность символов. При декодировании задержка возникает снова. Общее время задержки может быть велико, особенно при появлении блока, вероятность которого мала. Это следует учитывать при выборе длины кодируемого блока.

Еще один недостаток заключается в специфическом влиянии помех на достоверность приема. Одиночная ошибка может перевести передаваемую кодовую комбинацию в другую, не равную ей по длительности. Это повлечет за собой неправильное декодирование ряда последующих комбинаций, которые называют треком ошибки.

Специальными методами построения эффективного кода трек ошибки стараются свести к минимуму.

Следует отметить относительную сложность технической реализации систем эффективного кодирования.

Методы эффективного кодирования Шеннона-Фано и Хаффмана, рассмотренные выше, позволяют производить кодирование, если известна статистика входных сообщений, т.е. известна вероятность их появления.

# 6. Простейшие алгоритмы сжатия информации

***Цель сжатия*** – уменьшение количества бит, необходимых для хранения или передачи заданной информации, что дает возможность передавать сообщения более быстро и хранить более экономно и оперативно (последнее означает, что операция извлечения данной информации с устройства ее хранения будет проходить быстрее, что возможно, если скорость распаковки данных выше скорости считывания данных с носителя информации)[14].

Методы сжатия информации были разработаны как математическая теория, которая долгое время (до первой половины 80-х годов), мало использовалась в компьютерах на практике. Пусть у нас имеется файл размером 1 (один) мегабайт. Нам необходимо получить из него файл меньшего размера. Ничего сложного – запускаем архиватор, к примеру, WinZip, и получаем в результате, допустим, файл размером 600 килобайт. Куда же делись остальные 424 килобайта?

Сжатие информации является одним из способов ее кодирования. Вообще коды делятся на три большие группы – коды сжатия (эффективные коды), помехоустойчивые коды и криптографические коды. Коды, предназначенные для сжатия информации, делятся, в свою очередь, на коды без потерь и коды с потерями. Кодирование без потерь подразумевает абсолютно точное восстановление данных после декодирования и может применяться для сжатия любой информации. Кодирование с потерями имеет обычно гораздо более высокую степень сжатия, чем кодирование без потерь, но допускает некоторые отклонения декодированных данных от исходных.

Все методы сжатия информации можно условно разделить на два больших непересекающихся класса (рис. 19):

***Обратимое сжатие (сжатие без потерь)***. Обратимое сжатие всегда приводит к снижению объема выходного потока информации без изменения его информативности, т.е. без потери информационной структуры. Более того, из выходного потока, при помощи восстанавливающего алгоритма, можно получить входной[15].

Методы сжатия этого класса не могут допустить утрату информации, поэтому они основаны только на устранении ее избыточности, а информация имеет избыточность почти всегда (правда, если до этого кто-то ее уже не уплотнил). Если бы избыточности не было, нечего было бы и сжимать.

*Пример 1*. В русском языке 33 буквы, десять цифр и еще примерно полтора десятка знаков препинания и прочих специальных символов. Для текста, который записан только прописными русскими буквами (как в телеграммах и радиограммах) вполне хватило бы шестидесяти разных значений. Тем не менее, каждый символ обычно кодируется байтом, который содержит 8 битов и может выражать 256

---

[14] Лидовский В. В. Теория информации: Учебное пособие. — М.: Компания Спутник+, 2004. — 111 с. — ISBN 5-93406-661-7.

[15] Тропченко А.Ю., Тропченко А.А. Методы сжатия изображений, аудиосигналов и видео: Учебное пособие – СПб: СПбГУ ИТМО, 2009. – 108 с.

различных кодов. Это первое основание для избыточности. Для нашего «телеграфного» текста вполне хватило бы шести битов на символ.

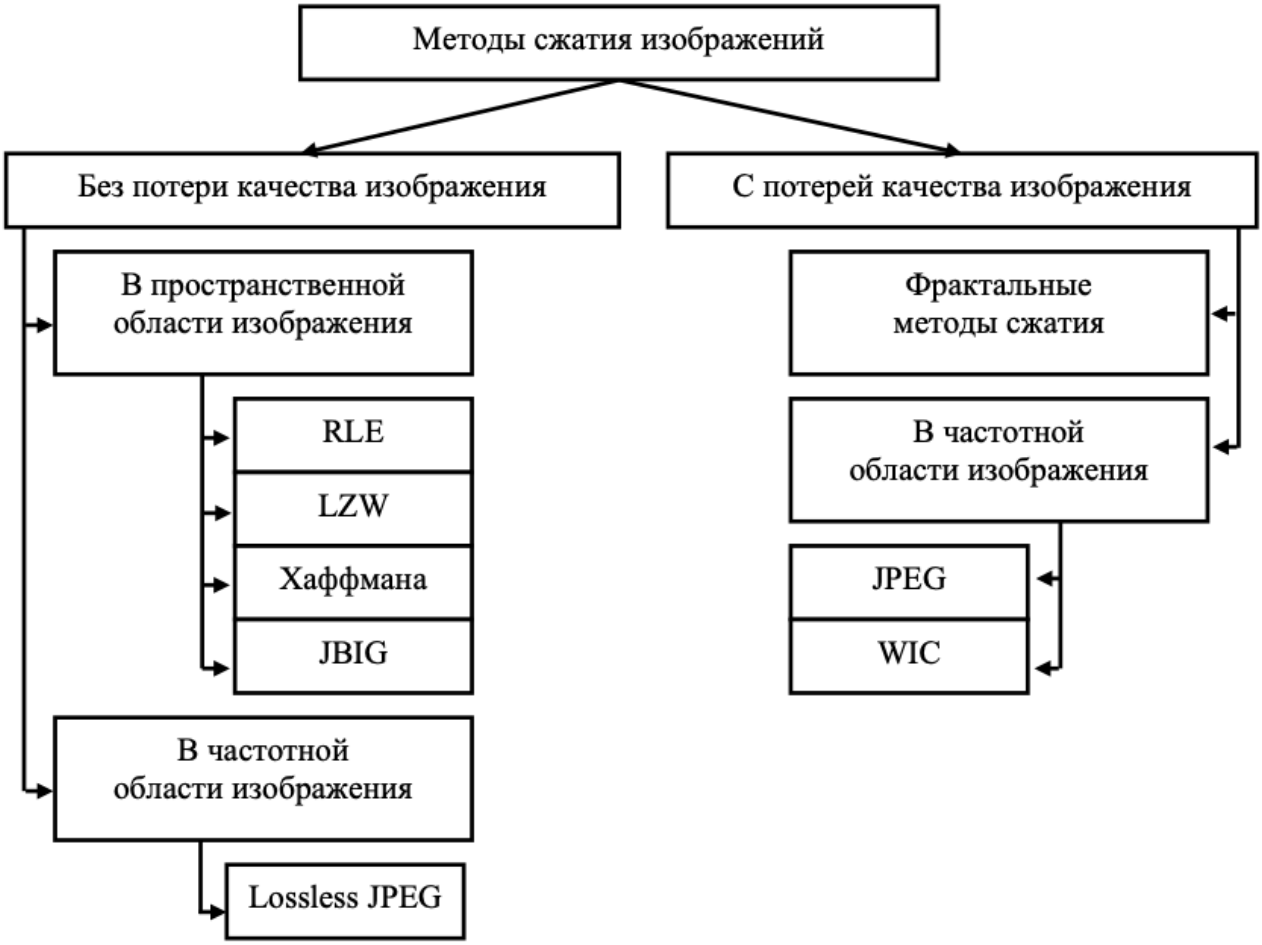



Рисунок 19 – Классификация методов сжатия изображений

*Пример 2*. В международной кодировке символов ASCII для кодирования любого символа отводится одинаковое количество битов (8), в то время как всем давно и хорошо известно, что наиболее часто встречающиеся символы имеет смысл кодировать меньшим количеством знаков. Так, например, в «азбуке Морзе» буквы «Е» и «Т», которые встречаются часто, кодируются одним знаком (соответственно это точка и тире). А такие редкие буквы, как «Ю» (• • - -) и «Ц» (- • - •), кодируются четырьмя знаками. Неэффективная кодировка – второе основание для избыточности. Программы, выполняющие сжатие информации, могут вводить свою кодировку (разную для разных файлов) и приписывать к сжатому файлу некую таблицу (словарь), из которой распаковывающая программа узнает, как в данном файле закодированы те или иные символы или их группы. Алгоритмы, основанные на перекодировании информации, называют алгоритмами Хафмана.

Наличие повторяющихся фрагментов – третье основание для избыточности. В текстах это встречается редко, но в таблицах и в графике повторение кодов – обычное явление. Так, например, если число 0 повторяется двадцать раз подряд, то нет смысла ставить двадцать нулевых байтов. Вместо них ставят один ноль и коэффициент 20. Такие алгоритмы, основанные на выявлении повторов, называют методами RLE (Run Length Encoding).

Большими повторяющимися последовательностями одинаковых байтов особенно отличаются графические иллюстрации, но не фотографические (там много шумов и соседние точки существенно различаются по параметрам), а такие, которые художники рисуют «гладким» цветом, как в мультипликационных фильмах.

***Необратимое сжатие (сжатие с потерями)***. Под необратимым сжатием подразумевают такое преобразование входного потока данных, при котором выходной поток, основанный на определенном формате информации, представляет достаточно похожий по внешним характеристикам на входной поток объект, однако отличается от него объемом. Степень сходства входного и выходного потоков определяется степенью соответствия некоторых свойств объектов (т.е. сжатой и несжатой информацией в соответствии с некоторым определенным форматом данных), представляемого данным потоком информации.

Такие алгоритмы неприменимы для текстовых документов, таблиц баз данных и особенно для программ.

Такие алгоритмы используются для сжатия, например данных растровых графических файлов с низкой степенью повторяемости байтов в потоке. При таком подходе используется свойство структуры формата графического файла и возможность представить графическую картинку приблизительно схожую по качеству отображения (для восприятия человеческим глазом) несколькими способами. Поэтому, кроме степени или величины сжатия, в таких алгоритмах возникает понятие качества, т.к. исходное изображение в процессе сжатия изменяется. Под качеством можно понимать степень соответствия исходного и результирующего изображения. Для графических файлов такое соответствие определяется визуально, хотя имеются и соответствующие формализованные методики и оценки. Необратимое сжатие невозможно применять в областях, в которых необходимо иметь точное соответствие информационной структуры входного и выходного потоков.

К алгоритмам сжатия с потерей информации относятся такие известные алгоритмы как JPEG и MPEG. Алгоритм JPEG используется при сжатии фотоизображений. Графические файлы, сжатые этим методом, имеют расширение JPG. Алгоритмы MPEG используют при сжатии видео и музыки. Эти файлы могут иметь различные расширения, в зависимости от конкретной программы, но наиболее известными являются .MPG для видео и .MP3 для музыки.

Алгоритмы сжатия с потерей информации применяют только для потребительских задач. Это значит, например, что если фотография передается для просмотра, а музыка для воспроизведения, то подобные алгоритмы применять можно. Если же они передаются для дальнейшей обработки, например для редактирования, то никакая потеря информации в исходном материале недопустима.

Величиной допустимой потери при сжатии обычно можно управлять. Это позволяет экспериментировать и добиваться оптимального соотношения размер/качество. На фотографических иллюстрациях, предназначенных для воспроизведения на экране, потеря 5% информации обычно некритична, а в некоторых случаях можно допустить и 20-25%.

Методы сжатия без потерь используются в основном в научных и медицинских приложениях, когда потеря информации недопустима или сами шумы изображения являются главной информацией, например в системах оценки качества

оптико-электронных систем. Коэффициент сжатия, достигаемый этими методами не более 1,5 для реальных сцен.

Методы сжатия с потерями позволяют получить существенно большие коэффициенты сжатия. Однако при этом происходит искажение исходного изображения, ухудшение его качества. В связи с этим при сравнении различных методов сжатия помимо коэффициента сжатия нужно учитывать качество восстановления изображения.

Для симметричных методов сжатия процедуры сжатия и восстановления однотипны. Время сжатия и восстановления для таких методов сравнимы. Для несимметричных методов процедура сжатия отличается от процедуры восстановления и обычно занимает большее машинное время.

## 6.1 Алгоритмы сжатия изображений без потерь

### 6.1.1 RLE-кодирование

Наиболее известный и простой алгоритм сжатия информации обратимым путем – это кодирование серий последовательностей (Run Length Encoding - RLE)[16]. Суть данного подхода состоит в замене цепочек или серий повторяющихся байтов или их последовательностей на один кодирующий байт и счетчик числа их повторений. Проблема всех аналогичных методов заключается лишь в определении способа, при помощи которого распаковывающий алгоритм мог бы отличить в результирующем потоке байтов кодированную серию от других – некодированных последовательностей байтов. Решение проблемы достигается обычно простановкой меток вначале кодированных цепочек. Такими метками могут быть, например, характерные значения битов в первом байте кодированной серии, значения первого байта кодированной серии и т.п.

Пример. Рассмотрим пример сжатия методом RLE. Пусть дана некоторая последовательность из 12 байтов:

11111111 11111111 11111111 11111111 11111111 11110000

00001111 11000011 10101010 10101010 10101010 10101010.

В начале исходной двоичной последовательности 5 раз повторяется байт 11111111. Чтобы упаковать эти 5 байтов, необходимо записать сначала управляющий байт 10000101, а затем повторяемый байт 11111111. В результате сжатия этого фрагмента данных выигрыш составит 3 байта. Далее идут 3 разных (неповторяющихся) байта: 11110000 00001111 и 11000011. Чтобы их «упаковать», нужно записать управляющий байт 00000011, а затем указать эти 3 неповторяющихся байта. В результате архивации этого фрагмента двоичной последовательности получается увеличение объема архива на 1 байт. Далее в последовательности 4 раза повторяется байт 10101010. Для архивации этого

[16] Тропченко А.Ю., Тропченко А.А. Методы сжатия изображений, аудиосигналов и видео: Учебное пособие – СПб: СПбГУ ИТМО, 2009. – 108 с.

фрагмента двоичных данных нужно сформировать управляющий байт 10000100 и записать повторяемый байт 10101010. Сжатие последнего фрагмента даст выигрыш 2 байта.

В результате такой архивации получена новая последовательность данных (архив), состоящая из 8 байтов:

10000101 11111111 00000011 11110000

00001111 11000011 10000100 10101010.

Таким образом, 12 байт исходной двоичной последовательности удалось сжать до 8 байт.

$$K_c = \frac{12}{8} = 1{,}5.$$

Пример. Выполним сжатие сообщения методом RLE.

Текст сообщения: ИНН 222221333.

| Фраза | Десятичный код | № | Двоичный код | Архив | № |
|---|---|---|---|---|---|
| И | 200 | 1 | 11001000 | **00000001** | 1 |
| Н | 205 | 2 | 11001101 | 11001000 | 2 |
| Н | 205 | 3 | 11001101 | **10000010** | 3 |
| | 32 | 4 | 00100000 | 11001101 | 4 |
| 2 | 50 | 5 | 00110010 | **00000001** | 5 |
| 2 | 50 | 6 | 00110010 | 00100000 | 6 |
| 2 | 50 | 7 | 00110010 | **10000101** | 7 |
| 2 | 50 | 8 | 00110010 | 00110010 | 8 |
| 2 | 50 | 9 | 00110010 | **00000001** | 9 |
| 1 | 49 | 10 | 00110001 | 00110001 | 10 |
| 3 | 51 | 11 | 00110011 | **10000011** | 11 |
| 3 | 51 | 12 | 00110011 | 00110011 | **12** |
| 3 | 51 | **13** | 00110011 | | |

Коэффициент сжатия в данном случае составил: $K_c = \frac{13}{12} = 1{,}08.$

Данные методы, как правило, достаточно эффективны для сжатия растровых графических изображений (BMP, PCX, TIFF), т.к. последние содержат достаточно длинных серий повторяющихся последовательностей байтов.

Недостатком метода RLE является достаточно низкая степень сжатия или стоимость кодирования файлов с малым числом серий и, что еще хуже – с малым числом повторяющихся байтов в сериях. К положительным сторонам алгоритма, пожалуй, можно отнести только то, что он не требует дополнительной памяти при работе, и быстро выполняется. [Видео]

### 6.1.1 Алгоритм Лемпеля-Зива (LZ-compression) LZ77

Метод Лемпеля-Зива – это простой, чёткий и легко реализуемый алгоритм сжатия данных, использующий повторяемость подстрок. Его главная мысль чрезвычайно проста: всякий раз, когда ранее прочитанная подстрока встречается повторно, алгоритм пишет вместо неё ссылку на её предыдущее вхождение.

Алгоритм LZ77 (Зив и Лемпель [1977]): на каждом шаге выводится тройка ($d$, $l$, $a$), что означает: сперва повторяется подстрока длины $l$, ранее встречавшаяся $d$ символов назад, а потом идёт символ $a$. Допускается $l > d$, в этом случае кодируется подстрока с периодическим окончанием. В конкретных реализациях всё это представляется по-разному, иногда разделяются ($d$, $l$) и $a$.

**Пример 1.** *Строка $w = abaababa$, её код* — $(0,0,a)(0,0,b)(2,1,a)(3,2,b)(0,0,a)$. *Или так:* $ab(2,1)(3,3)(2,2)$.

У алгоритма LZ77 есть определённая свобода выбора ранее виденной подстроки на каждом шаге сжатия. Оказывается, что жадная стратегия, выбирающая на каждом шаге самую длинную такую подстроку, даёт наилучшее сжатие. Следующая теорема приводится без доказательства.

**Теорема 1.** *Жадный LZ77 оптимален.*

Самая простая реализация LZ77 – через суффиксное дерево, в котором на каждом шаге читаются следующие символы входной строки, пока не находится самая длинная раньше встречавшаяся подстрока. Недостатки: во-первых, если строка длинная, то сколь бы эффективные структуры данных мы ни использовали, алгоритм неизбежно начнёт замедляться. Во-вторых, будут находиться очень старые подстроки, и потому количество битов, необходимое для записи $d$, будет расти; придётся придумывать, как сжимать $d$, и алгоритм усложнится.

***«Скользящее окно»***. Обычно используется вариант этого алгоритма, потенциально сжимающий немного хуже, однако более удобный в реализации. Подстроки ищутся среди последних $m$ прочитанных символов – в «скользящем окне» (sliding window). Более старые символы будут забываться. Чтобы искать их эффективно, тоже можно воспользоваться суффиксным деревом – но для этого нужно научиться удалять из него все суффиксы длиннее $m$. Для этого в недостроенном суффиксном дереве, создаваемом алгоритмом Укконена, надо на каждом шаге стирать самый длинный суффикс за совокупное линейное время. Алгоритм будет помнить лист, соответствующий самому длинному суффиксу, и на каждом шаге будет удаляться дуга, ведущая в этот лист; после этого алгоритм, следуя по суффиксной ссылке, будет запоминать следующий по порядку суффикс. Если же самый длинный недостроенный суффикс находится как раз на этой дуге, то дуга будет делиться.

### 6.1.2 Метод Лемпеля–Зива LZ78

Ранее рассмотренный метод LZ77 основан на том, что повторяющиеся подстроки представляются в виде ссылок на свои предыдущие вхождения. Таким образом, удаётся обойтись без составления таблицы часто встречающихся подстрок. Другой похожий алгоритм, LZ78 (Зив и Лемпель [1978]), строит-таки таблицу часто встречающихся подстрок – «словарь» (dictionary) – но делает он это по мере чтения строки. Словарь хранится в префиксном дереве, что позволяет легко находить самое длинное продолжение входной строки, уже присутствующее в словаре. При декодировании строится в точности этот же словарь. В сжатом представлении строки словарь не хранится.

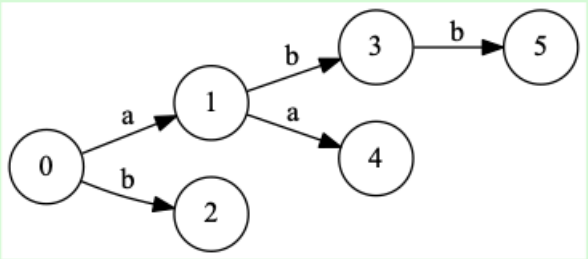
В начале работы в словаре содержится единственный элемент под номером ноль: $T_0 = \varepsilon$; иными словами, префиксное дерево состоит из корня, помеченного номером 0. На каждом шаге читается самая длинная строка $T_j = \vartheta$, уже имеющаяся в словаре, и выводится её код $j$; также читается и выводится следующий символ $a$. При этом в словарь добавляется новая строка $\vartheta a$ – конкатенация только что прочитанной со следующим входным символом.

На префиксном дереве это выглядит так: после вывода очередной пары $(j, a)$ алгоритм переходит в корень префиксного дерева, и дальше читает столько входных символов, сколько возможно. Когда очередной символ прочитать нельзя, создаётся новый лист, при этом выводится номер предыдущей вершины и прочитанный символ.

**Пример 2.** *Строка* $w = ababaaabb$, *в таблице* $T_0 = \varepsilon$.

***Кодирование:***

*Читается* $a$, *выводится* $0a$, *добавляется* $T_1 = a$.
*Читается* $b$, *выводится* $0b$, *добавляется* $T_2 = b$.
*Читается* $ab$, *выводится* $1b$, *добавляется* $T_3 = ab$.
*Читается* $aa$, *выводится* $1a$, *добавляется* $T_4 = aa$.
*Читается* $abb$, *выводится* $3b$, *добавляется* $T_5 = abb$.
*Код* — $0a0b1b1a3b$.




***Декодирование:***

*Читается* $0a$, *выводится* $a$, *добавляется* $T_1 = a$.
*Читается* $0b$, *выводится* $b$, *добавляется* $T_2 = b$.
*Читается* $1b$, *выводится* $ab$, *добавляется* $T_3 = ab$.
*Читается* $1a$, *выводится* $aa$, *добавляется* $T_4 = aa$.
*Читается* $3b$, *выводится* $abb$, *добавляется* $T_5 = abb$.

При декодировании строится то же самое префиксное дерево. Алгоритму при этом потребуется находить строку по номеру в таблице: можно, например, завести для этого отдельный массив строк, или же читать в префиксном дереве путь из вершины в корень. Прочитав очередную пару $(j, a)$, алгоритм выводит строку $T_j$ и символ $a$, после чего сразу перескакивает в вершину $j$ и присоединяет к ней новый лист, с переходом по символу $a$.

В практических реализациях словарь имеет ограниченный размер (например, $2^{12}$ элементов), и когда он заполняется, есть несколько возможностей. Во-первых, можно продолжать кодирование с имеющимся словарём. Во-вторых, можно очистить словарь. Наконец, есть варианты алгоритма, которые при переполнении словаря начинают удалять из него наименее нужные строки (тем или иным образом выбираемые).

Также при реализации никогда не вредно добавить в конце арифметическое кодирование: можно на каждом шаге оценивать вероятности разных подстрок и выполнять один шаг арифметического кодирования с этими вероятностями.

Существует множество разнообразных вариантов метода Лемпеля–Зива – таких, например, как полученный Велчем [1984] алгоритм Лемпеля–Зива–Велча (LZW) – вариант алгоритма LZ78.

### 6.1.3 Алгоритм Лемпеля-Зива-Велча (Lempel-Ziv-Welch - LZW)

Данный алгоритм отличают высокая скорость работы как при упаковке, так и при распаковке, достаточно скромные требования к памяти и простая аппаратная реализация. Недостаток – низкая степень сжатия по сравнению со схемой двухступенчатого кодирования. Алгоритм преобразует поток символов на входе в поток индексов ячеек словаря на выходе. Существует довольно большое семейство LZW-подобных алгоритмов, различающихся, например, методом поиска повторяющихся цепочек.

Коэффициент сжатия $K_c = 1000$ достигается только на одноцветных изображениях размером больше 4 Мб. Ситуация, когда алгоритм увеличивает изображение, встречается крайне редко. Сжатие обеспечивается за счет одинаковых подцепочек в потоке. Алгоритм является почти симметричным, при условии оптимальной реализации операции поиска строки в таблице.

LZW универсален – именно его варианты используются в обычных архиваторах. Он реализован в форматах GIF, TIFF и TGA.

### 6.1.4 Алгоритм JBIG

Алгоритм разработан группой экспертов ISO (*Joint Bi-level Experts Group*) специально для сжатия однобитных черно-белых изображений (например, для факсов или отсканированных документов). Может применяться и к 2-х, и к 4-х битовым картинкам. При этом алгоритм разбивает их на отдельные битовые плоскости. JBIG позволяет управлять такими параметрами, как порядок разбиения изображения на битовые плоскости, ширина полос в изображении, уровни

масштабирования. Последняя возможность позволяет легко ориентироваться в базе больших по размерам изображений, просматривая сначала их уменьшенные копии. Настраивая эти параметры, можно использовать интересный эффект при получении изображения по сети или по любому другому каналу, пропускная способность которого мала по сравнению с возможностями процессора. Распаковываться изображение на экране будет постепенно, как бы медленно "проявляясь". При этом человек начинает анализировать картинку задолго до конца процесса разархивации. Алгоритм построен на базе Q-кодировщика, патентом на который владеет IBM. Q-кодер также, как и алгоритм Хаффмана, использует для чаще появляющихся символов короткие цепочки, а для реже появляющихся длинные. Однако, в отличие от него, в алгоритме используются и последовательности символов. Характерной особенностью JBIG является резкое снижение степени сжатия при повышении уровня шумов исходного изображения.

### 6.1.5 Алгоритм Lossless JPEG

Этот алгоритм разработан группой экспертов в области фотографии (*Joint Photographic Expert Group*). В отличие от JBIG, Lossless JPEG ориентирован на полноцветные 24-битные изображения.

Стандарт сжатия изображений JPEG включает два способа сжатия: первый предназначен для сжатия без потерь, второй – сжатия с потерей качества. Метод сжатия без потерь, используемый в стандарте lossless JPEG основан на методе разностного (дифференциального) кодирования. Основная идея дифференциального кодирования состоит в следующем. Обычно изображения характеризуются сильной корреляцией между точками изображения. Этот факт учитывается при разностном кодировании, а именно, вместо сжатия последовательности точек изображения $x_1, x_2, \ldots x_N$, сжатию подвергается последовательность разностей $y_i = x_i - x_{i-1}$, $i = 1,2,...x_N$, $x_0 = 0$. Числа $y_i$ называют ошибками предсказания $x_i$. В стандарте losslessJPEG предусмотрено формирование ошибок предсказания с использованием предыдущих закодированных точек в текущей строке и\или в предыдущей строке.

Lossless JPEG рекомендуется применять в тех приложениях, где необходимо побитовое соответствие исходного и разархивированного изображений.

Приведенные алгоритмы достаточно универсальны и покрывают все типы изображений, с другой – они, по сегодняшним меркам, обеспечивают слишком маленький коэффициент архивации. Используя один из алгоритмов без потерь, можно обеспечить коэффициент архивации изображения примерно в два раза. В то же время сжатия с потерями оперируют с коэффициентами 10-200 раз. Помимо возможности модификации изображения, одна из основных причин подобной разницы заключается в том, что традиционные алгоритмы ориентированы на работу с цепочкой. Они не учитывают так называемую "когерентность областей" в изображениях. Идея когерентности областей заключается в малом изменении цвета и структуры изображения на небольшом участке. Все алгоритмы, о которых речь пойдет ниже, были созданы позднее специально для сжатия графики и используют эту идею.

# 7. Шифрование текстовой информации
## (Элементы криптологии)

***Криптология*** (от греч. cryptos – тайный и logos – слово) – наука, занимающаяся шифрованием и дешифрованием. Криптология состоит из двух частей – криптографии и криптоанализа. ***Криптография*** – наука о построении криптографических систем, используемых с целью защиты информации. Криптоанализ – наука о методах анализа криптографических систем, цель анализа – разработка методов раскрытия информации, защищаемой криптосистемой. На протяжении всей истории человечества основным фактором развития криптологии было противоборство методов защиты информации и методов её раскрытия. ***Шифр*** – совокупность заранее оговоренных способов преобразования исходного секретного сообщения с целью его защиты.

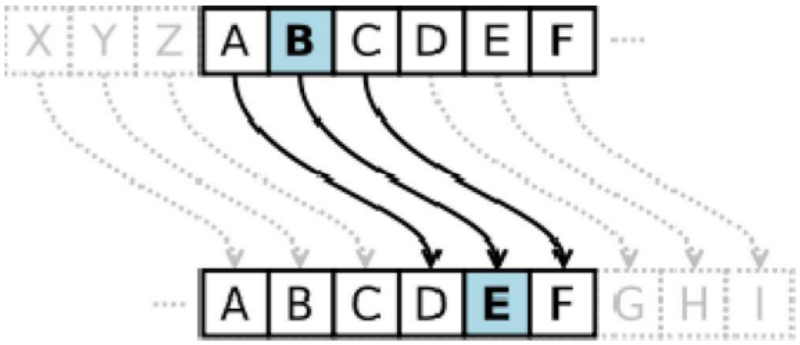
***Стеганогра́фия*** (от греч. στεγανός «скрытый» + γράφω «пишу»; букв. «тайнопись») – способ передачи или хранения информации с учётом сохранения в тайне самого факта такой передачи (хранения). Этот термин ввёл в 1499 году аббат бенедиктинского монастыря Св. Мартина в Шпонгейме Иоганн Тритемий в своём трактате «Стеганография» (лат. *Steganographia*), зашифрованном под магическую книгу.

В отличие от криптографии, которая скрывает содержимое тайного сообщения, стеганография скрывает сам факт его существования. Как правило, сообщение будет выглядеть как что-либо иное, например, как изображение, статья, список покупок, письмо или судоку. Стеганографию обычно используют совместно с методами криптографии, таким образом, дополняя её.

Преимущество стеганографии над чистой криптографией состоит в том, что сообщения не привлекают к себе внимания. Сообщения, факт шифрования которых не скрыт, вызывают подозрение и могут быть сами по себе уличающими в тех странах, в которых запрещена криптография. Таким образом, криптография защищает содержание сообщения, а стеганография – сам факт наличия каких-либо скрытых посланий от обличения.

### 7.1 Шифры простой замены

***Система шифрования Цезаря*** – частный случай шифра простой замены. Метод основан на замене каждой буквы сообщения на другую букву того же алфавита, путем смещения от исходной буквы на *K* букв.



Известная фраза Юлия Цезаря VENI VI D I VICI, где

| A | B | C | B | E | F | G | H | I | J | K | L | M |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| N | O | P | Q | R | S | T | U | V | W | X | Y | Z |

*пришел, увидел, победил*, зашифрованная с помощью данного метода, преобразуется в

SBKF SFAF SFZF

при смещении на 4 символа влево.

Последующие римские цезари модифицировали шифр, используя смещение в алфавите на четыре, пять и более букв. Мы можем описать их шифр в общем виде, если пронумеруем (закодируем) буквы русского алфавита числами от 0 до 31 (исключив букву Ё). Тогда правило шифрования запишется следующим образом:

$$c = (m + k) \mod 32, \tag{7.1}$$

где $m$ и $c$ – номера букв соответственно сообщения и шифротекста, а $k$ – некоторое целое число, называемое ключом шифра (в рассмотренном выше шифре Цезаря $k = 4$). (Здесь и в дальнейшем $a \mod b$ обозначает остаток от деления целого числа $a$ на целое число $b$, причем остаток берется из множества $\{0, 1, \ldots, b - 1\}$. Например, 13 mod 5 = 3.)

Чтобы дешифровать зашифрованный текст, нужно применить «обратный» алгоритм

$$m = (c - k) \mod 32 \tag{7.2}$$

Можно представить себе ситуацию, когда источник и получатель сообщений договорились использовать шифр (7.1), но для того, чтобы усложнить задачу противника, решили иногда менять ключ шифра. Отправитель сообщений и их получатель могут быть физическими лицами, организациями, какими-либо техническими системами. Иногда об А (отправитель сообщения) и В (получатель) говорят как об абонентах некоторой сети, о пользователях некоторой компьютерной системы или, еще более формально, как об абстрактных «сторонах» (англоязычный термин «party») или «сущностях» (entity), участвующих в информационном взаимодействии. Но чаще бывает удобно отождествлять участников обмена с некоторыми людьми и заменить формальные обозначения А и В на Алиса и Боб.Для этого Алиса каким-либо образом генерирует число $k$, передает его Бобу по закрытому каналу связи, и после этого они обмениваются сообщениями, зашифрованными с помощью этого ключа $k$. Замену ключа можно проводить, например, перед каждым сеансом связи или после передачи фиксированного числа букв (скажем, каждую десятку символов шифровать со своим $k$) и т.п. В таком случае говорят, что ключ порождается источником ключа. Схема рассмотренной криптосистемы с секретным ключом приведена на рис. 20.

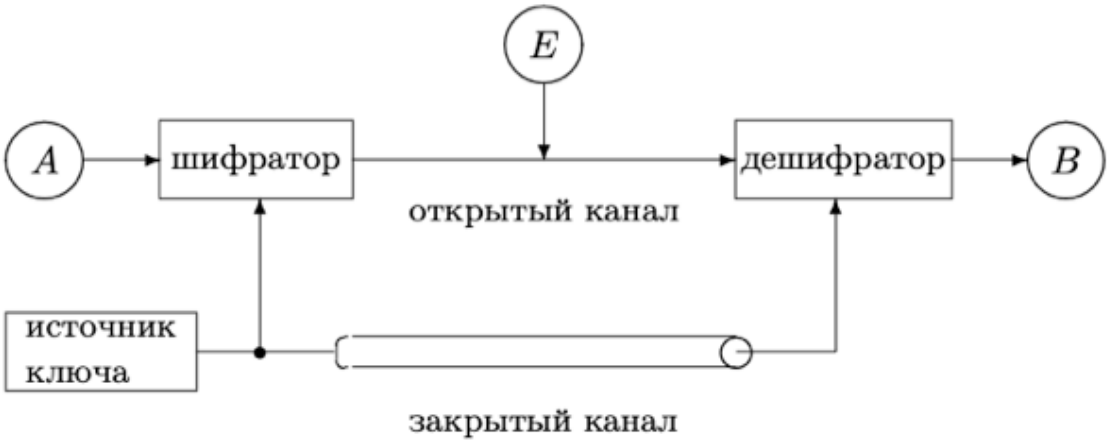



Рисунок 20 – Классическая система секретной связи

Обратимся теперь к анализу действий противника, пытающегося расшифровать сообщение и узнать секретный ключ, иными словами, вскрыть, или взломать шифр. Каждая попытка вскрытия шифра называется атакой на шифр (или на криптосистему). В криптографии принято считать, что противник может знать использованный алгоритм шифрования, характер передаваемых сообщений и перехваченный шифротекст, но не знает секретный ключ. Это называется «правилом Керкхоффса» в честь ученого, впервые сформулировавшего основные требования к шифрам. Иногда это правило кажется «перестраховкой», но такая «перестраховка» отнюдь не лишняя, если, скажем, передается распоряжение о переводе миллиона долларов с одного счета на другой.

*Пример*. Необходимо зашифровать слово

ПЕРЕМЕНА.

После применения к нему шифра Цезаря превращается в

ТИУИПИРГ

(если исключить букву Ё и считать, что в алфавите 32 буквы) ($k=3$). В нашем примере Ева (сторонний объект) знает, что шифр был построен в соответствии с (7.1), что исходное сообщение было на русском языке и что был передан шифротекст ТИУИПИРГ, но ключ Еве не известен.

Наиболее очевидная попытка расшифровки – последовательный перебор всех возможных ключей (это так называемый метод «грубой силы» (brute-force attack)). Итак, Ева перебирает последовательно все возможные ключи $k = 1, 2, \ldots,$ подставляя их в алгоритм дешифрования и оценивая получающиеся результаты. Попробуем и мы использовать этот метод. Результаты дешифрования по (1.2) при различных ключах и шифротексте ТИУИПИРГ сведены в таблицу 7.1. В большинстве случаев нам достаточно было расшифровать две-три буквы, чтобы отвергнуть соответствующий ключ (из-за отсутствия слова в русском языке, начинающегося с такого фрагмента).

Таблица 7.1 – Расшифровка слова ТИУИПИРГ

| $k$ | $m$ | $k$ | $m$ | $k$ | $m$ | $k$ | $m$ |
|---|---|---|---|---|---|---|---|
| 1 | СЗТ | 9 | ЙЯ | 17 | БЧ | 25 | ЩП |
| 2 | РЖС | 10 | ИЮЙ | 18 | АЦБ | 26 | ШОЩ |
| 3 | ПЕРЕМЕНА | 11 | ЗЭИ | 19 | ЯХА | 27 | ЧН |
| 4 | ОДП | 12 | ЖЬ | 20 | ЮФ | 28 | ЦМ |
| 5 | НГ | 13 | ЕЫ | 21 | ЭУ | 29 | ХЛЦ |
| 6 | МВ | 14 | ДЪ | 22 | Ь | 30 | ФК |
| 7 | ЛБМ | 15 | ГЩ | 23 | Ы | 31 | УЙ |
| 8 | КАЛАЗ | 16 | ВШГ | 24 | Ъ | 32 | ТИУИПИРГ |

Из таблицы 7.1 мы видим, что был использован ключ $k = 3$ и зашифровано сообщение ПЕРЕМЕНА. Причем для того, чтобы проверить остальные возможные значения ключа, нам не требовалось дешифровать все восемь букв, а в большинстве случаев после анализа двух-трех букв ключ отвергался (только при $k = 8$ надо было дешифровать пять букв, зато при $k = 22, 23, 24$ хватало и одной, так как в русском языке нет слов, начинающихся с Ь, Ъ, Ы).

Из этого примера мы видим, что рассмотренный шифр совершенно нестоек, для его вскрытия достаточно проанализировать несколько первых букв сообщения и после этого ключ $k$ однозначно определяется (и, следовательно, однозначно дешифруется все сообщение).

Греческим писателем Полибием за 100 лет до н.э. был изобретен так называемый *полибианский квадрат* размером 5*5, заполненный алфавитом в случайном порядке. Греческий алфавит имеет 24 буквы, а 25-м символом является пробел. Для шифрования на квадрате находили букву текста и записывали в зашифрованное сообщение букву, расположенную ниже ее в том же столбце. Если буква оказывалась в нижней строке таблицы, то брали верхнюю букву из того же столбца.

| | | | | |
|---|---|---|---|---|
| M | T | L | E | X |
| A | K | F | Q | Y |
| N | B | R | O | W |
| C | J | H | D | P |
| U | I | S | G | V |

***Схема*** шифрования Вижинера

Таблица Вижинера представляет собой квадратную матрицу с $n^2$ элементами, где n – число символов используемого алфавита. На рисунке показана верхняя часть таблицы Вижинера для кириллицы. Каждая строка получена циклическим сдвигом алфавита на символ. Для шифрования выбирается буквенный ключ, в соответствии с которым формируется рабочая матрица шифрования. Степень надежности закрытия информации повышается за счет того, что метод шифрования предусматривает

нарушение статистических закономерностей появления букв алфавита. Однако обладает достаточно высокой надежностью закрытия только при использовании весьма длинных ключей.

Шифр Вижинера с ключом, состоящим из одной буквы, известен как шифр Цезаря, а с неограниченным неповторяющимся ключом – как шифр Верната.

Таблица Вижинера

| 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 | ь | э | ю | я |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 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 | ь | э | ю | я | а |
| 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 | ь | э | ю | я | а | б |
| 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 | ь | э | ю | я | а | б | в |
| 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 | ь | э | ю | я | а | б | в | г |
| е | ё | ж | з | и | й | к | л | м | н | о | п | р | с | т | у | ф | х | ц | ч | ш | щ | ъ | ы | ь | э | ю | я | а | б | в | г | д |
| И т.д. до 33-ей строки.. | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Алгоритм шифрования Вижинера следующий. Из полной таблицы выбирается первая строка и те строки, первые буквы которых соответствуют буквам ключа. Первой размещается первая строка, а под нею – строки, соответствующие буквам ключа в порядке следования этих букв в ключе шифрования.

Процесс шифрования осуществляется следующим образом:

1. под каждой буквой шифруемого текста записываются буквы ключа. Ключ при этом повторяется необходимое число раз.

2. каждая буква шифруемого текста заменяется по подматрице буквами находящимися на пересечении линий, соединяющих буквы шифруемого текста в первой строке подматрицы и находящимися под ними букв ключа.

3. полученный текст может разбиваться на группы по несколько знаков.

Пусть, например, требуется зашифровать сообщение: *максимально допустимой ценой является пятьсот руб. за штуку*. В соответствии с первым правилом записываем под буквами шифруемого текста буквы ключа.

Получаем:

*максимально допустимой ценой является пятьсот руб. за штуку*
*книгакнигак нигакнигак нигак нигакниг акнигак ниг ак нигак*

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 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 | ь | э | ю | я |
| **к** | л | м | н | о | п | р | с | т | у | ф | х | ц | ч | ш | щ | ъ | ы | ь | э | ю | я | а | б | в | г | д | е | ё | ж | з | и | й |
| **н** | о | п | р | с | т | у | ф | х | ц | ч | ш | щ | ъ | ы | ь | э | ю | я | а | б | в | г | д | е | ё | ж | з | и | й | к | л | м |
| **и** | й | к | л | м | н | о | п | р | с | т | у | ф | х | ц | ч | ш | щ | ъ | ы | ь | э | ю | я | а | б | в | г | д | е | ё | ж | з |
| **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 | ь | э | ю | я | а | б | в |
| **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 | ь | э | ю | я |

Дальше осуществляется непосредственное шифрование в соответствии со вторым правилом, а именно: берем первую букву шифруемого текста (М) и соответствующую ей букву ключа (К); по букве шифруемого текста (М) входим в рабочую матрицу шифрования и выбираем под ней букву, расположенную в строке, соответствующей букве ключа (К), – в нашем примере такой буквой является (Ч); выбранную таким образом букву помещаем в зашифрованный текст. Эта процедура циклически повторяется до зашифрования всего текста.

Эксперименты показали, что при использовании такого метода статистические характеристики исходного текста практически не проявляются в зашифрованном сообщении. Нетрудно видеть, что замена по таблице Вижинера эквивалентна простой замене с циклическим изменением алфавита, т.е. здесь мы имеем полиалфавитную подстановку, причем число используемых алфавитов определяется числом букв в слове ключа. Поэтому стойкость такой замены определяется произведением стойкости прямой замены на число используемых алфавитов, т.е. число букв в ключе.

*Расшифровка* текста производится в следующей последовательности:

1. над буквами зашифрованного текста последовательно надписываются буквы ключа, причем ключ повторяется необходимое число раз.
2. в строке подматрицы Вижинера, соответствующей букве ключа отыскивается буква, соответствующая знаку зашифрованного текста. Находящаяся под ней буква первой строки подматрицы и будет буквой исходного текста.
3. полученный текст группируется в слова по смыслу.

Нетрудно видеть, что процедуры как прямого, так и обратного преобразования являются строго формальными, что позволяет реализовать их алгоритмически. Более того, обе процедуры легко реализуются по одному и тому же алгоритму.

Одним из недостатков шифрования по таблице Вижинера является то, что при небольшой длине ключа надежность шифрования остается невысокой, а формирование длинных ключей сопряжено с трудностями.

Нецелесообразно выбирать ключи с повторяющимися буквами, так как при этом стойкость шифра не возрастает. В то же время ключ должен легко запоминаться, чтобы его можно было не записывать. Последовательность же букв не имеющих смысла, запомнить трудно.

С целью повышения стойкости шифрования можно использовать усовершенствованные варианты таблицы Вижинера. Приведем только некоторые из них:

- во всех (кроме первой) строках таблицы буквы располагаются в произвольном порядке.
- в качестве ключа используется случайность последовательных чисел. Из таблицы Вижинера выбираются десять произвольных строк, которые кодируются натуральными числами от 0 до 10. Эти строки используются в соответствии с чередованием цифр в выбранном ключе.

Разновидностью шифра Вижинера является шифр Бофора, но при этом при определении цифрового эквивалента используют формулы

$$y_i = x_i - k_i(\mod 33) \text{ и } y_i = k_i - x_i(\mod 33).$$

При рассмотрении этих видов шифров становится очевидным, что чем больше длина ключа (например в шифре Вижинера), тем лучше шифр. Существенного улучшения свойств шифртекста можно достигнуть при использовании шифров с автоключом.

Шифр, в котором сам открытый текст или получающаяся криптограмма используются в качестве ключа, называется шифром с автоключом. Шифрование в этом случае начинается с ключа, называемого первичным, и продолжается с помощью открытого текста или криптограммы, смещенной на длину первичного ключа.

***Алгоритм перестановки***

Этот метод заключается в том, что символы шифруемого текста переставляются по определенным правилам внутри шифруемого блока символов. Рассмотрим некоторые разновидности этого метода, которые могут быть использованы в автоматизированных системах.

Самая простая перестановка – написать исходный текст задом наперед и одновременно разбить шифрограмму на пятерки букв. Например, из фразы

ПУСТЬ БУДЕТ ТАК, КАК МЫ ХОТЕЛИ.

получится такой шифротекст:

ИЛЕТО ХЫМКА ККАТТ ЕДУБЪ ТСУП

В последней группе (пятерке) не хватает одной буквы. Значит, прежде чем шифровать исходное выражение, следует его дополнить незначащей буквой (например, О) до числа, кратного пяти:

ПУСТЬ-БУДЕТ-ТАККА-КМЫХО-ТЕЛИО.

Тогда шифрограмма, несмотря на столь незначительные изменения, будет выглядеть по-другому:

ОИЛЕТ ОХЫМК АККАТ ТЕДУБ ЬТСУП

Кажется, ничего сложного, но при расшифровке проявляются серьезные неудобства.

Во время Гражданской войны в США в ходу был такой шифр: исходную фразу писали в несколько строк. Например, по пятнадцать букв в каждой (с заполнением последней строки незначащими буквами).

П У С Т Ь Б У Д Е Т Т А К К А
К М Ы Х О Т Е Л И К Л М Н О П

После этого вертикальные столбцы по порядку писали в строку с разбивкой на пятерки букв:

П К У М С  Ы Т Х Ь О  Б Т У Е Д  Л Е И Т К  Т Л А М К  Н К О А П

Если строки укоротить, а количество строк увеличить, то получится прямоугольник-решетка, в который можно записывать исходный текст. Но тут уже потребуется предварительная договоренность между адресатом и отправителем посланий, поскольку сама решетка может быть различной длины-высоты, записывать к нее можно по строкам, по столбцам, по спирали туда или по спирали обратно, можно писать и по диагоналями, а для шифрования можно брать тоже различные направления.

## 7.1 Шифры сложной замены

***Шифр Гронсфельда*** состоит в модификации шифра Цезаря числовым ключом. Для этого под буквами сообщения записывают цифры числового ключа. Если ключ короче сообщения, то его запись циклически повторяют. Зашифрованное сообщение получают примерно также, как в шифре Цезаря, но используют не одно жестко заданное смещение а фрагменты ключа.

Пусть в качестве ключа используется группа из трех цифр – 314, тогда сообщение

С О В Е Р Ш Е Н Н О С Е К Р Е Т Н О
3 1 4 3 1 4 3 1 4 3 1 4 3 1 4 3 1 4
Ф П Ё С Ь З О С С А Х З Л Ф З У С С

В ***шифрах многоалфавитной замены*** для шифрования каждого символа исходного сообщения применяется свой шифр простой замены (свой алфавит). Каждая строка в этой таблице соответствует одному шифру замены аналогично шифру Цезаря для алфавита, дополненного пробелом.

| | | А | Б | В | Г | Д | Е | Ж | З | И | Й | К | Л | М | Н | О | П | Р | С | Т | У | Ф | Х | Ц | Ч | Ш | Щ | Ъ | Ы | Э | Ю | Я |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | А | Б | В | Г | Д | Е | Ж | З | И | Й | К | Л | М | Н | О | П | Р | С | Т | У | Ф | Х | Ц | Ч | Ш | Щ | Ъ | Ы | Э | Ю | Я |
| **А** | А | Б | В | Г | Д | Е | Ж | З | И | Й | К | Л | М | Н | О | П | Р | С | Т | У | Ф | Х | Ц | Ч | Ш | Щ | Ъ | Ы | Э | Ю | Я | |
| **Б** | Б | В | Г | Д | Е | Ж | З | И | Й | К | Л | М | Н | О | П | Р | С | Т | У | Ф | Х | Ц | Ч | Ш | Щ | Ъ | Ы | Э | Ю | Я | | А |
| **В** | В | Г | Д | Е | Ж | З | И | Й | К | Л | М | Н | О | П | Р | С | Т | У | Ф | Х | Ц | Ч | Ш | Щ | Ъ | Ы | Э | Ю | Я | | А | Б |
| **Г** | Г | Д | Е | Ж | З | И | Й | К | Л | М | Н | О | П | Р | С | Т | У | Ф | Х | Ц | Ч | Ш | Щ | Ъ | Ы | Э | Ю | Я | | А | Б | В |
| **Д** | Д | Е | Ж | З | И | Й | К | Л | М | Н | О | П | Р | С | Т | У | Ф | Х | Ц | Ч | Ш | Щ | Ъ | Ы | Э | Ю | Я | | А | Б | В | Г |
| **Е** | Е | Ж | З | И | Й | К | Л | М | Н | О | П | Р | С | Т | У | Ф | Х | Ц | Ч | Ш | Щ | Ъ | Ы | Э | Ю | Я | | А | Б | В | Г | Д |
| **Ж** | Ж | З | И | Й | К | Л | М | Н | О | П | Р | С | Т | У | Ф | Х | Ц | Ч | Ш | Щ | Ъ | Ы | Э | Ю | Я | | А | Б | В | Г | Д | Е |
| **З** | З | И | Й | К | Л | М | Н | О | П | Р | С | Т | У | Ф | Х | Ц | Ч | Ш | Щ | Ъ | Ы | Э | Ю | Я | | А | Б | В | Г | Д | Е | Ж |
| **И** | И | Й | К | Л | М | Н | О | П | Р | С | Т | У | Ф | Х | Ц | Ч | Ш | Щ | Ъ | Ы | Э | Ю | Я | | А | Б | В | Г | Д | Е | Ж | З |
| **Й** | Й | К | Л | М | Н | О | П | Р | С | Т | У | Ф | Х | Ц | Ч | Ш | Щ | Ъ | Ы | Э | Ю | Я | | А | Б | В | Г | Д | Е | Ж | З | И |
| **К** | К | Л | М | Н | О | П | Р | С | Т | У | Ф | Х | Ц | Ч | Ш | Щ | Ъ | Ы | Э | Ю | Я | | А | Б | В | Г | Д | Е | Ж | З | И | Й |
| **Л** | Л | М | Н | О | П | Р | С | Т | У | Ф | Х | Ц | Ч | Ш | Щ | Ъ | Ы | Э | Ю | Я | | А | Б | В | Г | Д | Е | Ж | З | И | Й | К |
| **М** | М | Н | О | П | Р | С | Т | У | Ф | Х | Ц | Ч | Ш | Щ | Ъ | Ы | Э | Ю | Я | | А | Б | В | Г | Д | Е | Ж | З | И | Й | К | Л |
| **Н** | Н | О | П | Р | С | Т | У | Ф | Х | Ц | Ч | Ш | Щ | Ъ | Ы | Э | Ю | Я | | А | Б | В | Г | Д | Е | Ж | З | И | Й | К | Л | М |
| **О** | О | П | Р | С | Т | У | Ф | Х | Ц | Ч | Ш | Щ | Ъ | Ы | Э | Ю | Я | | А | Б | В | Г | Д | Е | Ж | З | И | Й | К | Л | М | Н |
| **П** | П | Р | С | Т | У | Ф | Х | Ц | Ч | Ш | Щ | Ъ | Ы | Э | Ю | Я | | А | Б | В | Г | Д | Е | Ж | З | И | Й | К | Л | М | Н | О |
| **Р** | Р | С | Т | У | Ф | Х | Ц | Ч | Ш | Щ | Ъ | Ы | Э | Ю | Я | | А | Б | В | Г | Д | Е | Ж | З | И | Й | К | Л | М | Н | О | П |
| **С** | С | Т | У | Ф | Х | Ц | Ч | Ш | Щ | Ъ | Ы | Э | Ю | Я | | А | Б | В | Г | Д | Е | Ж | З | И | Й | К | Л | М | Н | О | П | Р |
| **Т** | Т | У | Ф | Х | Ц | Ч | Ш | Щ | Ъ | Ы | Э | Ю | Я | | А | Б | В | Г | Д | Е | Ж | З | И | Й | К | Л | М | Н | О | П | Р | С |
| **У** | У | Ф | Х | Ц | Ч | Ш | Щ | Ъ | Ы | Э | Ю | Я | | А | Б | В | Г | Д | Е | Ж | З | И | Й | К | Л | М | Н | О | П | Р | С | Т |
| **Ф** | Ф | Х | Ц | Ч | Ш | Щ | Ъ | Ы | Э | Ю | Я | | А | Б | В | Г | Д | Е | Ж | З | И | Й | К | Л | М | Н | О | П | Р | С | Т | У |
| **Х** | Х | Ц | Ч | Ш | Щ | Ъ | Ы | Э | Ю | Я | | А | Б | В | Г | Д | Е | Ж | З | И | Й | К | Л | М | Н | О | П | Р | С | Т | У | Ф |
| **Ц** | Ц | Ч | Ш | Щ | Ъ | Ы | Э | Ю | Я | | А | Б | В | Г | Д | Е | Ж | З | И | Й | К | Л | М | Н | О | П | Р | С | Т | У | Ф | Х |
| **Ч** | Ч | Ш | Щ | Ъ | Ы | Э | Ю | Я | | А | Б | В | Г | Д | Е | Ж | З | И | Й | К | Л | М | Н | О | П | Р | С | Т | У | Ф | Х | Ц |
| **Ш** | Ш | Щ | Ъ | Ы | Э | Ю | Я | | А | Б | В | Г | Д | Е | Ж | З | И | Й | К | Л | М | Н | О | П | Р | С | Т | У | Ф | Х | Ц | Ч |
| **Щ** | Щ | Ъ | Ы | Э | Ю | Я | | А | Б | В | Г | Д | Е | Ж | З | И | Й | К | Л | М | Н | О | П | Р | С | Т | У | Ф | Х | Ц | Ч | Ш |
| **Ъ** | Ъ | Ы | Э | Ю | Я | | А | Б | В | Г | Д | Е | Ж | З | И | Й | К | Л | М | Н | О | П | Р | С | Т | У | Ф | Х | Ц | Ч | Ш | Щ |
| **Ы** | Ы | Э | Ю | Я | | А | Б | В | Г | Д | Е | Ж | З | И | Й | К | Л | М | Н | О | П | Р | С | Т | У | Ф | Х | Ц | Ч | Ш | Щ | Ъ |
| **Э** | Э | Ю | Я | | А | Б | В | Г | Д | Е | Ж | З | И | Й | К | Л | М | Н | О | П | Р | С | Т | У | Ф | Х | Ц | Ч | Ш | Щ | Ъ | Ы |
| **Ю** | Ю | Я | | А | Б | В | Г | Д | Е | Ж | З | И | Й | К | Л | М | Н | О | П | Р | С | Т | У | Ф | Х | Ц | Ч | Ш | Щ | Ъ | Ы | Э |
| **Я** | Я | | А | Б | В | Г | Д | Е | Ж | З | И | Й | К | Л | М | Н | О | П | Р | С | Т | У | Ф | Х | Ц | Ч | Ш | Щ | Ъ | Ы | Э | Ю |

При шифровании сообщения его выписывают в строку, а под ним ключ. Если ключ оказался короче сообщения, то его циклически повторяют. Зашифрованное сообщение получают, находя символ в колонке таблицы по букве текста и строке, соответствующей букве ключа. Например, используя ключ АГАВА, из сообщения ПРИЕЗЖАЮ ШЕСТОГО получаем следующую шифровку:

П Р И Е З Ж А Ю _ Ш Е С Т О Г О
А Г А В А А Г А В А А Г А В А А
П Ф И И З Ж Д Ю В Ш Е Х Т С Г О

Такая операция соответствует сложению кодов ASCII символов сообщения и ключа по модулю 256.

Прогрессивный ***ключ Тритемиуса***, является примером многоалфавитного шифра. Строка, обозначенная как сдвиг 0, совпадает с обычным порядком букв в алфавите. Буквы в следующей строке сдвинуты на один символ влево с циклическим сдвигом оставшихся позиций.

Каждая последующая строка получается с помощью такого же сдвига алфавита на один символ влево относительно предыдущей строки. Это продолжается до тех пор, пока в результате циклических сдвигов алфавита не будет смещен на все возможные позиции. Один из методов использования такого алфавита заключается в выборе первого символа шифрованного сообщения из строки,

полученной при сдвиге на 1 символ, второго символа – из строки, полученной при сдвиге на 2 символа, и т.д.

В начале 1850-х гг. Чарлз Уитстон придумал так называемый «прямоугольный шифр». Леон Плейфер, близкий друг Уитстона, рассказал об этом шифре во время официального обеда в 1854 г. министру внутренних дел лорду Пальмерстону и принцу Альберту. А поскольку Плейфер был хорошо известен в военных и дипломатических кругах, то за творением Уитстона навечно закрепилось название «шифр Плейфера».

Данный шифр стал первым буквенным ***биграммным шифром*** (в биграммной таблице Порты использовались символы, а не буквы). Он был предназначен для обеспечения секретности телеграфной связи и применялся британскими войсками в Англо-бурской и Первой мировой войнах. Им пользовалась также австралийская служба береговой охраны островов во время Второй мировой войны.

Шифр предусматривает шифрование пар символов (биграмм). Таким образом, этот шифр более устойчив к взлому по сравнению с шифром простой замены, так как затрудняется частотный анализ. Он может быть проведен, но не для 26 возможных символов (латинский алфавит), а для 26 x 26 = 676 возможных биграмм. Анализ частоты биграмм возможен, но является значительно более трудным и требует намного большего объема зашифрованного текста.

Для шифрования сообщения необходимо разбить его на биграммы (группы из двух символов), при этом, если в биграмме встретятся два одинаковых символа, то между ними добавляется заранее оговоренный вспомогательный символ (в оригинале – *X*, для русского алфавита — *Я*). Например, «зашифрованное сообщение» становится «за ши фр ов ан но ес о*Я* об ще ни е*Я*». Для формирования ключевой таблицы выбирается лозунг и далее она заполняется по правилам шифрующей системы Трисемуса. Например, лозунг «ДЯДИНА»

| | | | | | |
|---|---|---|---|---|---|
| Д | Я | И | Н | А | Б |
| В | Г | Е | Ё | Ж | З |
| Й | К | Л | М | О | П |
| Р | С | Т | У | Ф | Х |
| Ц | Ч | Ш | Щ | Ы | Ь |
| Ъ | Э | Ю | - | 1 | 2 |

Затем, руководствуясь следующими правилами, выполняется зашифровывание пар символов исходного текста:

1. Если символы биграммы исходного текста встречаются в одной строке, то эти символы замещаются на символы, расположенные в ближайших столбцах справа от соответствующих символов. Если символ является последним в строке, то он заменяется на первый символ этой же строки.

2. Если символы биграммы исходного текста встречаются в одном столбце, то они преобразуются в символы того же столбца, находящимися непосредственно под ними. Если символ является нижним в столбце, то он заменяется на первый символ этого же столбца.

3. Если символы биграммы исходного текста находятся в разных столбцах и разных строках, то они заменяются на символы, находящиеся в тех же строках, но соответствующие другим углам прямоугольника.

*Пример шифрования*:

— биграмма «за» формирует прямоугольник – заменяется на «жб»;
— биграмма «ши» находятся в одном столбце – заменяется на «юе»;
— биграмма «фр» находятся в одной строке – заменяется на «хс»;
— биграмма «ов» формирует прямоугольник – заменяется на «йж»;
— биграмма «ан» находятся в одной строке – заменяется на «ба»;
— биграмма «но» формирует прямоугольник – заменяется на «ам»;
— биграмма «ес» формирует прямоугольник – заменяется на «гт»;
— биграмма «оя» формирует прямоугольник – заменяется на «ка»;
— биграмма «об» формирует прямоугольник – заменяется на «па»;
— биграмма «ще» формирует прямоугольник – заменяется на «шё»;
— биграмма «ни» формирует прямоугольник – заменяется на «ан»;
— биграмма «ея» формирует прямоугольник – заменяется на «ги».

Шифрограмма – «жб юе хс йж ба ам гт ка па шё ан ги».

Для расшифровки необходимо использовать инверсию этих правил, откидывая символы *Я* (или *Х*), если они не несут смысла в исходном сообщении.

# РАЗДЕЛ 2. ПРАКТИЧЕСКИЙ

## ЛАБОРАТОРНАЯ РАБОТА № 1
## Системы счисления

*Цель работы:* Повторить и закрепить знания студентов по способам представления чисел в позиционных системах счисления, переводу чисел из десятичной системы счисления в любую другую и обратно.

Теоретические вопросы

1. Системы счисления.

*Задания к лабораторной работе*

В соответствии с выданным Вам вариантом выполнить:

1. Переведите данное число из десятичной системы счисления в двоичную, восьмеричную и шестнадцатеричную системы счисления.
2. Переведите данное число в десятичную систему счисления.
3. Сложите числа.
4. Выполните вычитание.
5. Выполните умножение.

Примечание: В заданиях 3 – 5 проверьте правильность вычислений переводом исходных данных и результатов в десятичную систему счисления. В задании 1 д) получите пять знаков после запятой в двоичном представлении.

### Вариант 1

1. а) $860_{(10)}$; б) $785_{(10)}$; в) $149{,}375_{(10)}$; г) $953{,}25_{(10)}$; д) $228{,}79_{(10)}$.
2. а) $1001010_{(2)}$; б) $1100111_{(2)}$; в) $110101101{,}00011_{(2)}$; г) $111111100{,}0001_{(2)}$; д) $775{,}11_{(8)}$; е) $294{,}3_{(16)}$.
3. а) $1101100000_{(2)} + 10110110_{(2)}$; б) $101110111_{(2)} + 1000100001_{(2)}$; в) $1001000111{,}01_{(2)} + 100001101{,}101_{(2)}$; г) $271{,}34_{(8)} + 1566{,}2_{(8)}$; д) $65{,}2_{(16)} + 3CA{,}8_{(16)}$.
4. а) $1011001001_{(2)} - 1000111011_{(2)}$; б) $1110000110_{(2)} - 101111101_{(2)}$; в) $101010000{,}10111_{(2)} - 11001100{,}01_{(2)}$; г) $731{,}6_{(8)} - 622{,}6_{(8)}$; д) $22D{,}1_{(16)} - 123{,}8_{(16)}$.
5. а) $1011001_{(2)} \cdot 1011011_{(2)}$; б) $723{,}1_{(8)} \cdot 50{,}2_{(8)}$; в) $69{,}4_{(16)} \cdot A{,}B_{(16)}$.

### Вариант 2

1. а) $250_{(10)}$; б) $757_{(10)}$; в) $711{,}25_{(10)}$; г) $914{,}625_{(10)}$; д) $261{,}78_{(10)}$.
2. а) $1111000_{(2)}$; б) $1111000000_{(2)}$; в) $111101100{,}01101_{(2)}$; г) $100111100{,}1101_{(2)}$; д) $1233{,}5_{(8)}$; е) $2B3{,}F4_{(16)}$.
3. а) $1010101_{(2)} + 10000101_{(2)}$; б) $1111011101_{(2)} + 101101000_{(2)}$; в) $100100111{,}001_{(2)} + 100111010{,}101_{(2)}$; г) $607{,}54_{(8)} + 1620{,}2_{(8)}$; д) $3BF{,}A_{(16)} + 313{,}A_{(16)}$.
4. а) $1001000011_{(2)} - 10110111_{(2)}$; б) $111011100_{(2)} - 10010100_{(2)}$; в) $1100110110{,}0011_{(2)} - 11111110{,}01_{(2)}$; г) $1360{,}14_{(8)} - 1216{,}4_{(8)}$; д) $33B{,}6_{(16)} - 11B{,}4_{(16)}$.
5. а) $11001_{(2)} \cdot 1011100_{(2)}$; б) $451{,}2_{(8)} \cdot 5{,}24_{(8)}$; в) $2B{,}A_{(16)} \cdot 36{,}6_{(16)}$.

*Контрольные вопросы:*

1. Что называется системой счисления?
2. На какие два типа можно разделить все системы счисления?
3. Какие системы счисления применяются в вычислительной технике: позиционные или непозиционные? Почему?
4. Что называется основанием системы счисления?
5. Охарактеризуйте двоичную, восьмеричную, шестнадцатеричную системы счисления: алфавит, основание системы счисления, запись числа.
6. По каким правилам выполняется сложение двух положительных целых чисел?
7. Каковы правила выполнения арифметических операций в двоичной системе счисления?
8. Для чего используется перевод чисел из одной системы счисления в другую?
9. Сформулируйте правила перевода чисел из системы счисления с основанием $p$ в десятичную систему счисления и обратного перевода: из десятичной системы счисления в систему счисления с основанием $p$. Приведите примеры.
10. Как выполнить перевод чисел из двоичной СС в восьмеричную и обратный перевод? Из двоичной СС в шестнадцатеричную и обратно? Приведите примеры. Почему эти правила так просты?
11. По каким правилам выполняется перевод из восьмеричной в шестнадцатеричную СС и наоборот? Приведите примеры.
12. Чему равны веса разрядов слева от точки, разделяющей целую и дробную части, в двоичной системе счисления (восьмеричной, шестнадцатеричной)?
13. Чему равны веса разрядов справа от точки, разделяющей целую и дробную части, в двоичной системе счисления (восьмеричной, шестнадцатеричной)?

## ЛАБОРАТОРНАЯ РАБОТА № 2
## **Методы оценки количества информации**

*Цель:* Ввести понятие «количество информации»; вероятности, равновероятных и неравновероятных событий; научиться определять количество информации.

Теоретические вопросы

1. Количество информации.
2. Понятие вероятности.

*Пример.* На экзамене приготовлено 30 билетов.

– Чему равно количество событий, которые могут произойти при вытягивании билета? (Ответ – 30).

– Равновероятны эти события или нет? (Ответ – равновероятны).

– Чему равна неопределенность знаний студента перед тем как он вытянет билет? (Ответ – 30).

– Во сколько раз уменьшится неопределенность знаний после того как студент билет вытянул? (Ответ – в 30 раз).

– Зависит ли этот показатель от номера вытянутого билета? (Ответ – нет, т. к. события равновероятны).

Можно сделать следующий вывод: чем больше начальное число возможных равновероятных событий, тем в большее количество раз уменьшается неопределенность наших знаний, и тем большее количество информации будет содержать сообщение о результатах опыта.

*Пример.* Книга лежит на одной из двух полок – верхней или нижней. Сообщение о том, что книга лежит на верхней полке, уменьшает неопределенность ровно вдвое и несет 1 бит информации.

*Сообщение о том, что произошло одно событие из двух равновероятных, несет 1 бит информации.*

*Пример.* Нестеров живет на Ленинградской улице. Мы получили сообщение, что номер его дома есть число четное, которое уменьшило неопределенность. После получения такой информации, мы стали знать больше, но информационная неопределенность осталась, хотя и уменьшилась в два раза.

*Пример.* Ваш друг живет в 16-ти этажном доме. Сколько информации содержит сообщение о том, что друг живет на 7 этаже.

*Решение*: Информационная неопределенность (количество возможных результатов события) равна 16. Будем задавать вопросы, на которые можно ответить только «да» или «нет». Вопрос будем ставить так, чтобы каждый ответ приносил 1 бит информации, т.е. уменьшал информационную неопределенность в два раза.

Задаем вопросы: - Друг живет выше 8-го этажа?
- Нет.

После этого ответа число вариантов уменьшилось в два раза, следовательно, информационная неопределенность уменьшилась в два раза. Получен 1 бит информации.

- Друг живет выше 4-го этажа?
- Да.

Число вариантов уменьшилось еще в два раза, получен еще 1 бит информации.

- Друг живет выше 6-го этажа?
- Да.

После данного ответа осталось два варианта: друг живет или на 7 этаже, или на 8 этаже. Получен еще 1 бит информации.

- Друг живет на 8-м этаже?
- Нет.
- Все ясно. Друг живет на 7-м этаже.

Каждый ответ уменьшал информационную неопределенность в два раза.

Всего было задано 4 вопроса. Получено 4 бита информации. Сообщение о том, что друг живет на 7-м этаже 16-ти этажного дома несет 4 бита информации.

*Задания*

1. Какое количество информации будет получено при отгадывании числа из интервала от 1 до 64; от 1 до 20?
2. Какое количество информации будет получено после первого хода в игре «крестики-нолики» на поле 3 х 3; 4 х 4?
3. Сколько могло произойти событий, если при реализации одного из них получилось 6 бит информации?
4. В коробке лежат кубики: 10 красных, 8 зеленых, 5 желтых, 12 синих. Вычислите вероятность доставания кубика каждого цвета и количество информации, которое при этом будет получено.
5. Какое количество информации будет содержать зрительное сообщение о цвете вынутого шарика, если в непрозрачном мешочке находится 50 белых, 25 красных, 25 синих шариков.
6. В группе учатся 12 девочек и 8 мальчиков. Какое количество информации несет сообщение, что к доске вызовут девочку; мальчика?
7. В корзине лежит 16 шаров разного цвета. Сколько информации несет сообщение, что достали белый шар?
8. В велокроссе участвуют 119 спортсменов. Специальное устройство регистрирует прохождение каждым из участников промежуточного финиша, записывая его номер с использованием минимально возможного количества бит, одинакового для каждого спортсмена. Каков информационный объем сообщения, записанного устройством, после того как промежуточный финиш прошли 70 велосипедистов?
9. Словарный запас некоторого языка составляет 256 слов, каждое из которых состоит точно из 4 букв. Сколько букв в алфавите языка?

*Контрольные вопросы:*

1. Какое сообщение называется информативным?
2. Что значит событие равновероятно; неравновероятно?
3. Что такое 1 бит информации?
4. Как определить количество информации для равновероятных событий?
5. Как определить количество информации для неравновероятных событий?
6. В чем заключается алфавитный подход к измерению количества информации?

ЛАБОРАТОРНАЯ РАБОТА № 3

**Кодирование информации**

*Цель:* изучение способов кодирования информации.

Теоретические вопросы

1. Кодирование информации.
2. Символьное кодирование информации.

*Задание 1.* Дана кодовая таблица азбуки Морзе:

| | | |
|---|---|---|
| А • — | Л • — • • | Ц — • — • |
| Б — • • • | М — — | Ч — — — • |
| В • — — | Н — • | Ш — — — — |
| Г — — • | О — — — | Щ — — • — |
| Д — • • | П • — — • | Ъ • — — • — • |
| Е • | Р • — • | Ы — • — — |
| Ж • • • — | С • • • | Ь — • • — |
| З — — • • | Т — | Э • • — • • |
| И • • | У • • — | Ю • • — — |
| Й • — — — | Ф • • — • | Я • — • — |
| К — • — | Х • • • • | |

Декодируйте сообщение:

— — — — — • — • • — — — — • • — • — • — • — • — —

Закодируйте с помощью азбуки Морзе слова ПАРОЛЬ, ЭКРАНИРОВАНИЕ, КОДИРОВАНИЕ.

*Задание 2.* Дана кодировочная таблица (первая цифра кода – номер строки, вторая – номер столбца):

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| 0 | А | Б | В | Г | Д | Е | Ё | Ж | З |
| 1 | И | К | Л | М | Н | О | П | Р | С |
| 2 | Т | У | Ф | Х | Ц | Ч | Ш | Щ | Ъ |
| 3 | Ы | Ь | Э | Ю | Я | _ | . | , | ? |
| 4 | : | ; | - | ! | » | | | | |

С помощью этой кодировочной таблицы закодируйте фразу: Я ЗНАЮ МЕТОДЫ ШИФРОВАНИЯ.

*Задание 3.* Используя таблицу кодирования.

| | Символ | Двоичный код |
|---|---|---|
| 1 | Р | 101 |
| 2 | О | 100 |
| 3 | И | 010 |
| 4 | Е | 011 |
| 5 | П | 1110 |
| 6 | М | 1100 |
| 7 | Пробел | 1101 |
| 8 | А | 0010 |
| 9 | С | 0011 |
| 10 | Г | 00000 |
| 11 | В | 00001 |
| 12 | К | 00010 |
| 13 | Б | 00011 |
| 14 | З | 111100 |
| 15 | Т | 111101 |
| 16 | Ь | 111110 |
| 17 | Н | 111111 |

Закодируйте слово СИМВОЛ. Рассчитайте полученную степень сжатия. Раскодируйте слово 11101011000000010100101100110000010.

*Задание 4.* Используйте алгоритм кодирования LZ77 для сжатия сообщения ТЕОРИЯ ИНФОРМАЦИИ – ТЕОРИЯ КОДИРОВАНИЯ. Оцените эффективность сжатия данного сообщения.

*Задание 5.* Используйте алгоритм кодирования LZW для сжатия сообщения ТЕОРИЯ ИНФОРМАЦИИ – ТЕОРИЯ КОДИРОВАНИЯ. Оцените эффективность сжатия данного сообщения.

*Контрольные вопросы:*

1. Какие методы кодирования информации вы знаете? В чем суть данных методов?
2. Для чего необходимо кодировать информацию?

ЛАБОРАТОРНАЯ РАБОТА №4

## Применение классических шифров замены

*Цель*: научиться применять классические шифры замены.

Теоретические вопросы

1. Понятие криптографии.
2. Понятие шифра.
3. Шифр замены.
4. Шифр многоалфавитной замены.
5. Сходства и различия шифра Гронсфельда и шифра Цезаря.
6. Биграммный шифр замены.

*Задание 1*. Выбрать один из методов замены:

1. шифр Цезаря;
2. шифр Полибианский квадрат;
3. шифр Трисимуса;
4. шифр многоалфавитной замены Вижинера;
5. шифр биграммами;
6. шифр Гронсфельда.

Составить алгоритм программы шифрования по выбранному методу.

*Задание 2*. Составить программу шифрования по выбранному методу.

*Задание 3*. Составить алгоритм программы расшифрования по выбранному методу. Составить программу расшифрования по выбранному методу.

*Задание 4*. Расшифровать текст,

а) зашифрованный шифром Цезаря со сдвигом на 4 позиции:

У окдгнбэылмбаноюзыбожмдлокнднебиь

б) зашифрованный шифром Цезаря со сдвигом на 6 позиции:

Иыфщлзвмелнмцйкяиыкъбиьъзвгйякялмъзиьдъвбъжязъ

в) зашифрованный заменой по кодовому слову «пароль»:

випигьпжоймгсзпчгумйрпигяиьлйжбийржгясыипипльбийнсынгнсьзъ

*Контрольные вопросы:*

1. Какие методы шифрования сообщений вы знаете? В чем суть данных методов?
2. Для чего необходимо шифровать сообщения?

## ЛАБОРАТОРНАЯ РАБОТА №5
## **Криптоанализ шифра Вижинера**

*Цель*: научиться выполнять криптоанализ шифра Виженера.

Теоретические вопросы

1. Понятие криптоанализа.
2. Шифр Виженера.

З*адание*. Задан некоторый текст зашифрованный шифром Виженера, требуется определить ключевое слово и прочитать открытый текст.

*Шифрованный текст*

Влцдутжбюцхъяррмшбрхцэооэцгбрьцмйфктъъюьмшэсяцпунуящэйтаьэдкцибрьцгбр
пачкъуцпъбьсэгкцъгуущарцёэвърюуоюэкааэбрняфукабъарпяъафкъиьжяффнйо
яфывбнэнфуюгбрьсшьжэтбэёчюъюръегофкбьчябашвёэуъъюаднчжчужцёэвлрнчулб
юпцуруньъшсэюъзкцхъяррнрювяспэмасчкпэужьжыатуфуярюравртубурьпэщлафоуф
бюацмнубсюкйтаьэдйюнооэгюожбгкбрънцэпотчмёодзцвбцшщвщепчдчдръюьскасэг
ъппэгюкдойрсрэвоопчщшоказръббнэугнялёкьсрбёуыэбдэулбюасшоуэтъшкрсдугэфл
бубуъчнчтртпэгюкиугюэмэгюккъъпэгяапуфуэзьрадзьжчюрмфцхраююанчёчюъыхьъ
цомэфъцпоирькнщпэтэузуябащущбаыэйчдфрпэцъьрьцъцпоилуфэдцойэдятррачкубу
фнйтаьэдкцкрннцюабугюуубурьпйюэъжтгюркующоъуфъэгясуоичщщчдцсфырэдщэ
ъуяфшёчцюйрщвяхвмкршрпгюопэуцчйтаьэдкцибрьцыяжтюрбуэтэбдуящэубъибрюв
ъежагибрбагбрымпуноцшяжцечкфодщоъчжшйуъцхчщвуэбдлдъэгясуахзцэбдэулькнъ
щбжяцэьрёдъьвюврнуяфуоухфекьгцччгэъжтанопчынажпачкъуъмэнкйрэфщэъьбуд
эндадъярьеюэлэтчоубъцэфэвлнёэгфдсэвэёкбсчоукгаутэыпуббцчкпэгючсаъбэнэфърк
ацхёваетуфяепьрювържадфёжбьфутощоявьъгупчршуитеачйчирамчюфчоуяюонкяжы
кгсцбрясшчйотъъжрсщчл

*Контрольные вопросы:*

1. Поясните суть шифрования методом Виженера.
2. В чем плюсы и недостатки данного метода шифрования?

## ЛАБОРАТОРНАЯ РАБОТА №6
## **Программная реализация классических шифров**

*Цель:* изучение способов кодирования информации.

Теоретические вопросы

1. Основные понятия криптографии.
2. Основные понятия криптоанализа.
3. Методы шифрования и кодирования информации.

*Задание 1*

*Вариант 1*

В Средние века для шифрования перестановкой применялись и магические квадраты. Магическими квадратами называют квадратные таблицы с вписанными в их клетки последовательными натуральными числами, начиная от 1, которые дают в сумме по каждому столбцу, каждой строке и каждой диагонали одно и то же число. Шифруемый текст вписывали в магические квадраты в соответствии с нумерацией их клеток. Если затем выписать содержимое такой таблицы по строкам, то получится шифротекст, сформированный благодаря перестановке букв исходного сообщения. В те времена считалось, что созданные с помощью магических квадратов шифротексты охраняет не только ключ, но и магическая сила.

Пример магического квадрата и его заполнения сообщением «Прилетаю восьмого» показан ниже (рисунок ).

| | | | |
|---|---|---|---|
| 16 | 3 | 2 | 13 |
| 5 | 10 | 11 | 8 |
| 9 | 6 | 7 | 12 |
| 4 | 15 | 14 | 1 |

| | | | |
|---|---|---|---|
| О | И | Р | М |
| Е | О | С | Ю |
| В | Т | А | Ь |
| Л | Г | О | П |

Шифротекст, получаемый при считывании содержимого правой таблицы по строкам, имеет вполне загадочный вид

ОИРМ ЕОСЮ ВТАЬ ЛГОП.

Число магических квадратов быстро возрастает с увеличением размера квадрата. Существует только один магический квадрат размером 3 × 3 (если не учитывать его повороты). Количество магических квадратов 4 × 4 составляет уже 880, а количество магических квадратов 5 × 5 – около 250000.

Пользуясь изложенным способом создать программу, которая:

а) зашифрует введенный текст и сохранит его в файл;

б) считает зашифрованный текст из файла и расшифрует данный текст.

*Вариант 2*

Для обеспечения дополнительной скрытности можно повторно зашифровать сообщение, которое уже прошло шифрование. Такой метод шифрования называется двойной перестановкой. В случае двойной перестановки столбцов и строк таблицы перестановки определяются отдельно для столбцов и отдельно для строк. Сначала в таблицу записывается текст сообщения, а потом поочередно переставляются столбцы, а затем строки. При расшифровании порядок перестановок должен быть обратным.

Пример выполнения шифрования методом двойной перестановки показан на рис. Если считывать шифротекст из правой таблицы построчно блоками по четыре буквы, то получится следующее:

ТЮАЕ ООГМ РЛИП ОЬСВ.

| | 4 | 1 | 3 | 2 |
|---|---|---|---|---|
| 3 | П | Р | И | Л |
| 1 | Е | Т | А | Ю |
| 4 | В | О | С | Ь |
| 2 | М | О | Г | О |

Исходная таблица

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 3 | Р | Л | И | П |
| 1 | Т | Ю | А | Е |
| 4 | О | Ь | С | В |
| 2 | О | О | Г | М |

Перестановка столбцов

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | Т | Ю | А | Е |
| 2 | О | О | Г | М |
| 3 | Р | Л | И | П |
| 4 | О | Ь | С | В |

Перестановка строк

Ключом к шифру двойной перестановки служит последовательность номеров столбцов и номеров строк исходной таблицы (в нашем примере последовательности 4132 и 3142 соответственно).

Число вариантов двойной перестановки быстро возрастает при увеличении размера таблицы:

- для таблицы 3 × 3 – 36 вариантов;
- для таблицы 4 × 4 – 576 вариантов;
- для таблицы 5 ×5 – 14400 вариантов.

Пользуясь изложенным способом создать программу, которая:

а) зашифрует введенный текст и сохранит его в файл;

б) считает зашифрованный текст из файла и расшифрует данный текст.

# РАЗДЕЛ 3. КОНТРОЛЬ ЗНАНИЙ

## Методические указания по выполнению контрольной работы

По дисциплине «Теория информации» студент должен выполнить контрольную работу, которая предоставляется на кафедру и защищается студентом заочной (дистанционной) формы получения образования до начала лабораторно-зачетной сессии. При выполнении контрольных работ необходимо соблюдать следующие правила:

1. Контрольная работа может выполняться как в рукописном так и печатном виде, на титульном листе необходимо указать наименование дисциплины, фамилию и инициалы студента, выполненный вариант (соответствует последним двум цифрам зачетной книжки), шифр специальности и номер группы.
2. Контрольную работу следует выполнять аккуратно, оставляя поля для замечаний рецензента.
3. Для пояснения решения задачи, где это нужно, сделать чертеж.
4. Решение задач и выбор используемых при этом формул следует сопровождать пояснениями.
5. В пояснениях к задаче необходимо указывать основные законы и формулы, на использовании которых базируется решение данной задачи.
6. В контрольной работе следует указывать учебники и учебные пособия, которые использовались при решении задач.

## Общая формулировка заданий к контрольной работе

1. Перевести данное число из десятичной системы счисления в двоичную, восьмеричную и шестнадцатеричную системы счисления. (Согласно вариантам приведенным ниже).

2. Перевести данное число в десятичную систему счисления.

3. Выполните арифметические действия в заданных системах счисления.

4. Среди приведённых ниже трёх чисел, записанных в различных системах счисления, найдите максимальное и запишите его в ответе в десятичной системе счисления. В ответе запишите только число, основание системы счисления указывать не нужно.

5. Условие задачи см. ниже в приведенных вариантах.

6. Закодировать методом Хаффмана Вашу *Фамилию Имя Отчество*.

Указание:

*1. Определяете количество уникальных символов в ФИО.*

*2. Исходя из этого, узнаёте количество бит, необходимых для кодирования.*

*3. Рассчитываете частоту вхождения и вес каждого символа в строке.*

*4. Создаете дерево Хаффмана, получая код для каждой буквы.*

*5. Записываете с помощью этого кода необходимую информацию.*

7. Условие задачи см. ниже в приведенных вариантах.

8. Используя теоретический материал зашифровать текст, используя квадрат Полибия, согласно вариантам (см. ниже). Регистр не учитывается.

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | А | Б | В | Г | Д | Е |
| 2 | Ё | Ж | З | И | Й | Л |
| 3 | Л | М | Н | О | П | Р |
| 4 | С | Т | У | Ф | Х | Ц |
| 5 | Ч | Ш | Щ | Ъ | Ы | Ь |
| 6 | Э | Ю | Я | , | . | - |

9. Используя теоретический материал расшифровать текст, используя квадрат Полибия, согласно вариантам. Регистр не учитывается.

10. Выполнить сжатие заданного текста методом RLE, воспользовавшись таблицами ASCII. Вычислить коэффициент сжатия и контрольную сумму.

## Вариант 1

1. а) 777; б) 305; в) 153,25; г) 162,25; д) 248,46.

2. а) $1100111011_2$; б) $10000000111_2$; в) $10110101,1_2$; г) $100000110,10101_2$; д) $671,24_8$; е) $41A,6_{16}$.

3. $126_7 + 662_7$

4. $23_{16}$, $32_8$, $11110_2$.

5. Измерьте информационный объем сообщения «Ура! Скоро Новый год!» в битах, байтах, килобайтах (Кб), мегабайтах (Мб).

*Указание*: считается, что текст набран с помощью компьютера, один символ алфавита несет 1 байт информации. Пробел – это тоже символ в алфавите мощностью 256 символов.

6. Условие задачи см. выше в общая формулировка заданий.

7. На любой из позиций двоичного кода может быть с равной вероятностью переданы «0» и «1». Помехи преобразуют «1» в «0» с вероятностью 0,02 и «0» в «1» с вероятностью 0,04. Найти вероятность того, что был передан «0», если принят «0».

8. *Труд человека кормит, а лень портит.*

9. 26 11 12 11 33 51 24 26.

10. BBBBBBACCCABBBBBB

## Вариант 2

1. а) 164; б) 255; в) 712,25; г) 670,25; д) 11,89.

2. а) $1001110011_2$; б) $1001000_2$; в) $1111100111,01_2$; г) $1010001100,101101_2$; д) $413,41_8$; е) $118,8C_{16}$.

3. $126_8 + 662_8$

4. $38_{16}$, $75_8$, $110100_2$.

5. Измерьте примерную информационную емкость одной страницы любого своего учебника, всего учебника.

*Указание:* Для выполнения задания возьмите учебник по любимому предмету, посчитайте число строк на странице, число символов в строке, включая пробелы. Помните, что один символ алфавита несет 1 байт информации. Перемножив полученные значения, Вы найдете информационную емкость одной страницы учебника (в байтах).

Сколько таких учебников может поместиться на дискете 1,44 Мб, на винчестере в 1 Гб.

6. Условие задачи см. выше в общая формулировка заданий.

7. По линии связи посылаются сигналы «1» и «0» с вероятностями $P_1 = 0,6$ и $P_0 = 0,4$. Если посылается сигнал «1», то с вероятностью $p_{11} = 0,9$ принимается сигнал «1», с вероятностью $p_{10} = 0,1$ принимается сигнал «0». Если посылается сигнал «0», то с вероятностями $p_{01} = 0,3$ принимается сигнал «1», $p_{00} = 0,7$ принимается сигнал «0». Какова условная вероятность того, что посылается сигнал «1» при условии, что принимается сигнал «1»?

8. *Без труда не вытащить и рыбки из пруда*.

9. 11 15 36 16 41 11 33 42.

10. FFFFFKKKACCCCF

## Вариант 3

1. а) 273; б) 661; в) 156,25; г) 797,5; д) 53,74.

2. а) $1100000000_2$; б) $1101011111_2$; в) $1011001101{,}00011_2$; г) $1011110100{,}011_2$; д) $1017{,}2_8$; е) $111{,}B_{16}$.

3. $101101_2 + 11011_2$

4. $14_{16}$, $26_8$, $11000_2$.

5. 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.

6. Условие задачи см. выше в общая формулировка заданий.

7. По линии связи посылаются сигналы «1» и «0» с вероятностями $P_1 = 0{,}3$ и $P_0 = 0{,}7$. Если посылается сигнал «1», то с вероятностью $p_{11} = 0{,}8$ принимается сигнал «1», с вероятностью $p_{10} = 0{,}2$ принимается сигнал «0». Если посылается сигнал «0», то с вероятностями $p_{01} = 0{,}3$ принимается сигнал «1», $p_{00} = 0{,}7$ принимается сигнал «0». Какова условная вероятность того, что посылается сигнал «1» при условии, что принимается сигнал «1»?

8. *Без беды друга не узнаешь.*

9. 26 11 31 16 51 24 42 56

10. LLLLLPHJKKKKKEFFFF

## Вариант 4

1. а) 105; б) 358; в) 377,5; г) 247,25; д) 87,27.

2. а) $1100001001_2$; б) $1100100101_2$; в) $1111110110{,}01_2$; г) $11001100{,}011_2$; д) $112{,}04_8$; е) $334{,}A_{16}$.

3. $6346_8 \cdot 447_8$

4. $24_{16}$, $50_8$, $101100_2$.

5. Яд находится в одном из 16 бокалов. Сколько единиц информации будет содержать сообщение о бокале с ядом?

6. Условие задачи см. выше в общая формулировка заданий.

7. По линии связи посылаются сигналы «1» и «0» с вероятностями $P_1 = 0{,}6$ и $P_0 = 0{,}4$. Если посылается сигнал «1», то с вероятностью $p_{11} = 0{,}9$ принимается сигнал «1», с вероятностью $p_{10} = 0{,}1$ принимается сигнал «0». Если посылается сигнал «0», то с вероятностями $p_{01} = 0{,}3$ принимается сигнал «1», $p_{00} = 0{,}7$ принимается сигнал «0». Какова вероятность того, что принимается сигнал 1?

8. *Терпенье и труд все перетрут.*

9. 12 24 36 32 11 33 26 11

10. qqqqLLLkkkkPIOHHH

## Вариант 5

1. а) 500; б) 675; в) 810,25; г) 1017,25; д) 123,72.
2. а) $1101010001_2$; б) $100011100_2$; в) $1101110001{,}011011_2$; г) $110011000{,}111001_2$; д) $1347{,}178_8$; е) $155{,}6C16_{16}$.
3. $11044_6 \cdot 305_6$
4. $50_{16}$, $106_8$, $1001010_2$.
5. Какое количество информации приходится на одну букву алфавита, состоящего из 16, 25, 32 букв?
6. Условие задачи см. выше в общая формулировка заданий.
7. По линии связи посылаются сигналы «1» и «0» с вероятностями $P_1 = 0{,}5$ и $P_0 = 0{,}5$. Если посылается сигнал «1», то с вероятностью $p_{11} = 0{,}6$ принимается сигнал «1», с вероятностью $p_{10} = 0{,}4$ принимается сигнал «0». Если посылается сигнал «0», то с вероятностями $p_{01} = 0{,}2$ принимается сигнал «1», $p_{00} = 0{,}8$ принимается сигнал «0». Какова вероятность того, что принимается сигнал 1?
8. *Жизнь измеряется не годами, а трудами.*
9. 13 16 36 16 33 24 46 11
10. PPPPdJKKEEEC

## Вариант 6

1. а) 218; б) 808; в) 176,25; г) 284,25; д) 253,04.
2. а) $111000100_2$; б) $1011001101_2$; в) $10110011{,}01_2$; г) $1010111111{,}011_2$; д) $1665{,}3_8$; е) $FA{,}7_{16}$.
3. $1011_2 \cdot 1101_2$
4. $50_{16}$, $106_8$, $1001010_2$.
5. Сколько байт составляет фраза «Я изучаю информационные технологии», если для кодирования символов используется таблица ASCII?
6. Условие задачи см. выше в общая формулировка заданий.
7. По линии связи посылаются сигналы «1» и «0» с вероятностями $P_1 = 0{,}6$ и $P_0 = 0{,}4$. Если посылается сигнал «1», то с вероятностью $p_{11} = 0{,}9$ принимается сигнал «1», с вероятностью $p_{10} = 0{,}1$ принимается сигнал «0». Если посылается сигнал «0», то с вероятностями $p_{01} = 0{,}3$ принимается сигнал «1», $p_{00} = 0{,}7$ принимается сигнал «0». Какова вероятность того, что принимается сигнал 0?
8. *Язык болтает, а голова не знает.*
9. 13 55 46 16 15 24 42 56
10. RRGRRRLKDDDPQSSS

## Вариант 7

1. а) 306; б) 467; в) 218,5; г) 667,25; д) 318,87.

2. а) $1111000111_2$; б) $11010101_2$; в) $1001111010{,}010001_2$; г) $1000001111{,}01_2$; д) $465{,}3_8$; е) $252{,}38_{16}$.

3. $6512_7 + 566_7$

4. $41_{16}$, $77_8$, $1000010_2$.

5. Считая, что каждый символ кодируется 16-ю битами, оцените информационный объём следующего предложения в кодировке Unicode (в байтах):

*Каждый символ кодируется восемью битами.*

6. Условие задачи см. выше в общая формулировка заданий.

7. По линии связи посылаются сигналы «1» и «0» с вероятностями $P_1 = 0{,}8$ и $P_0 = 0{,}2$. Если посылается сигнал «1», то с вероятностью $p_{11} = 0{,}8$ принимается сигнал «1», с вероятностью $p_{10} = 0{,}2$ принимается сигнал «0». Если посылается сигнал «0», то с вероятностями $p_{01} = 0{,}4$ принимается сигнал «1», $p_{00} = 0{,}6$ принимается сигнал «0». Какова вероятность того, что принимается сигнал 0?

8. *Книги не говорят, а правду сказывают.*

9. 14 43 12 34 52 31 21 35

10. OOORRRDGESSSSSS

## Вариант 8

1. а) 167; б) 113; в) 607,5; г) 828,25; д) 314,71.

2. а) $110010001_2$; б) $100100000_2$; в) $1110011100{,}1112$2; г) $1010111010{,}1110111_2$; д) $704{,}6_8$; е) $367{,}38_{16}$.

3. $3765_8 + 122_8$

4. $32_{16}$, $60_8$, $110110_2$.

5. Юстасу необходимо передать следующее сообщение:

*Дорогой Алекс! От всей души поздравляю с успешной сдачей экзамена по информатике. Желаю дальнейших успехов. Ваш Юстас.*

Пеленгатор определяет место передачи, если она длится не менее 3 минут. С какой скоростью (бит/с) Юстас должен передавать радиограмму?

6. Условие задачи см. выше в общая формулировка заданий.

7. На любой из позиций двоичного кода может быть с равной вероятностью переданы «0» и «1». Помехи преобразуют «1» в «0» с вероятностью 0,02 и «0» в «1» с вероятностью 0,04. Найти вероятность того, что был передан «1», если принят «1».

8. *Красота до вечера, а доброта навек.*

9. 14 36 63 23 33 43 42 56

10. POIFFFFRdsaRR

## Вариант 9

1. а) 342; б) 374; в) 164,25; г) 520,375; д) 97,14.

2. а) $1000110110_2$; б) $111100001_2$; в) $1110010100{,}1011001_2$; г) $1000000110{,}00101_2$; д) $666{,}16_8$; е) $1C7{,}68_{16}$.

3. $101_2 + 11_2$

4. $20_{16}$, $36_8$, $11100_2$.

5. 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?

6. Условие задачи см. выше в общая формулировка заданий.

7. По линии связи посылаются сигналы «1» и «0» с вероятностями $P_1 = 0{,}6$ и $P_0 = 0{,}4$. Если посылается сигнал «1», то с вероятностью $p_{11} = 0{,}9$ принимается сигнал «1», с вероятностью $p_{10} = 0{,}1$ принимается сигнал «0». Если посылается сигнал «0», то с вероятностями $p_{01} = 0{,}3$ принимается сигнал «1», $p_{00} = 0{,}7$ принимается сигнал «0». Какова условная вероятность того, что посылается сигнал «0» при условии, что принимается сигнал «0»?

8. *Прежде соберись, а потом дерись.*

9. 26 36 16 15 24 42 14 13

10. POKJJJJJGHYTFFFF

## Вариант 10

1. а) 524; б) 222; в) 579,5; г) 847,625; д) 53,35.

2. а) $101111111_2$; б) $1111100110_2$; в) $10011000{,}1101011_2$; г) $1110001101{,}1001_2$; д) $140{,}22_8$; е) $1DE{,}54_{16}$.

3. $11_8 \cdot 11_8$

4. $14_{16}$, $17_8$, $10011_2$.

5. В коробке лежат 7 цветных карандашей. Какое количество информации содержит сообщение, что из коробки достали красный карандаш?

6. Условие задачи см. выше в общая формулировка заданий.

7. По линии связи посылаются сигналы «1» и «0» с вероятностями $P_1 = 0{,}6$ и $P_0 = 0{,}4$. Если посылается сигнал «1», то с вероятностью $p_{11} = 0{,}9$ принимается сигнал «1», с вероятностью $p_{10} = 0{,}1$ принимается сигнал «0». Если посылается сигнал «0», то с вероятностями $p_{01} = 0{,}3$ принимается сигнал «1», $p_{00} = 0{,}7$ принимается сигнал «0». Какова вероятность того, что принимается сигнал 0?

8. *Курица по зернышку клюет да сыта бывает.*

9. 26 36 11 41 11 13 26 11

10. PPPPTTTTiCCCC

## Вариант 11

1. а) 113; б) 875; в) 535,1875; г) 649,25; д) 6,52.
2. а) $11101000_2$; б) $1010001111_2$; в) $1101101000{,}01_2$; г) $1000000101{,}01011_2$; д) $1600{,}14_8$; е) $1E9{,}4_{16}$.
3. $135_6 \cdot 23_6$
4. $47_{16}$, $120_8$, $1001011_2$.
5. 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?
6. Условие задачи см. выше в общая формулировка заданий.
7. По линии связи посылаются сигналы «1» и «0» с вероятностями $P_1 = 0{,}3$ и $P_0 = 0{,}7$. Если посылается сигнал «1», то с вероятностью $p_{11} = 0{,}6$ принимается сигнал «1», с вероятностью $p_{10} = 0{,}4$ принимается сигнал «0». Если посылается сигнал «0», то с вероятностями $p_{01} = 0{,}2$ принимается сигнал «1», $p_{00} = 0{,}8$ принимается сигнал «0». Какова вероятность того, что принимается сигнал 1?
8. *Все течет, все изменяется.*
9. 26 34 52 51 34 33 26 11
10. ЯЯЯОООООЗЗЗОДДДД

## Вариант 12

1. а) 294; б) 723; в) 950,25; г) 976,625; д) 282,73.
2. а) $10000011001_2$; б) $10101100_2$; в) $1101100{,}01_2$; г) $1110001100{,}1_2$; д) $1053{,}2_8$; е) $200{,}6_{16}$.
3. $1111_2 \cdot 101_2$
4. $60_{16}$, $134_8$, $1100001_2$.
5. 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?
6. Условие задачи см. выше в общая формулировка заданий.
7. По линии связи посылаются сигналы «1» и «0» с вероятностями $P_1 = 0{,}9$ и $P_0 = 0{,}1$. Если посылается сигнал «1», то с вероятностью $p_{11} = 0{,}9$ принимается сигнал «1», с вероятностью $p_{10} = 0{,}1$ принимается сигнал «0». Если посылается сигнал «0», то с вероятностями $p_{01} = 0{,}2$ принимается сигнал «1», $p_{00} = 0{,}8$ принимается сигнал «0». Какова вероятность того, что принимается сигнал 0?
8. *В тихом омуте черти водятся.*
9. 32 24 56 16 13 11 42 56
10. DDDDDDDOOOOOOOOOGGGJDFFFFF

## Вариант 13

1. а) 617; б) 597; в) 412,25; г) 545,25; д) 84,82.
2. а) $110111101_2$; б) $1110011101_2$; в) $111001000{,}01_2$; г) $1100111001{,}1001_2$; д) $1471{,}17_8$; е) $3EC{,}5_{16}$.
3. $a81c_{16} + 9fb6_{16}$
4. $35_{16}$, $71_8$, $110111_2$.
5. Книга, набранная с помощью компьютера, содержит 150 страниц; на каждой странице – 40 строк, в каждой строке – 60 символов. Каков объем информации в книге?
6. Условие задачи см. выше в общей формулировке заданий.
7. По линии связи посылаются сигналы «1» и «0» с вероятностями $P_1 = 0{,}8$ и $P_0 = 0{,}2$. Если посылается сигнал «1», то с вероятностью $p_{11} = 0{,}8$ принимается сигнал «1», с вероятностью $p_{10} = 0{,}2$ принимается сигнал «0». Если посылается сигнал «0», то с вероятностями $p_{01} = 0{,}4$ принимается сигнал «1», $p_{00} = 0{,}6$ принимается сигнал «0». Какова вероятность того, что принимается сигнал 1?
8. *В зимний холод всякий молод.*
9. 32 24 14 23 11 42 24 42
10. ЖЖЖЖЖУУУУУУК

## Вариант 14

1. а) 1047; б) 335; в) 814,5; г) 518,625; д) 198,91.
2. а) $1101100000_2$; б) $100001010_2$; в) $1011010101{,}1_2$; г) $1010011111{,}1101_2$; д) $452{,}63_8$; е) $1E7{,}08_{16}$.
3. $a1c_{16} \cdot 6_{16}$
4. $59_{16}$, $126_8$, $1011100_2$.
5. При игре в кости используется кубик с шестью гранями. Сколько бит информации получает игрок при каждом бросании кубика?
6. Условие задачи см. выше в общая формулировка заданий.
7. По линии связи посылаются сигналы «1» и «0» с вероятностями $P_1 = 0{,}6$ и $P_0 = 0{,}4$. Если посылается сигнал «1», то с вероятностью $p_{11} = 0{,}9$ принимается сигнал «1», с вероятностью $p_{10} = 0{,}1$ принимается сигнал «0». Если посылается сигнал «0», то с вероятностями $p_{01} = 0{,}3$ принимается сигнал «1», $p_{00} = 0{,}7$ принимается сигнал «0». Какова условная вероятность того, что посылается сигнал «1» при условии, что принимается сигнал «0»?
8. *Раз солгал, а на век лгуном стал.*
9. 32 16 42 11 41 42 11 23
10. НННННННОРННННУR

## Вариант 15

1. а) 887; б) 233; в) 801,5; г) 936,3125; д) 218,73.

2. а) $1010100001_2$; б) $10000010101_2$; в) $1011110000{,}100101_2$; г) $1000110001{,}1011_2$; д) $1034{,}34_8$; е) $72{,}6_{16}$.

3. $1c_{16} \cdot ab_{16}$

4. $41_{16}$, $107_8$, $1000011_2$.

5. В барабане для розыгрыша лотереи находится 32 шара. Сколько информации содержит сообщение о первом выпавшем номере (например, выпал номер 15)?

6. Условие задачи см. выше в общая формулировка заданий.

7. По линии связи посылаются сигналы «1» и «0» с вероятностями $P_1 = 0{,}8$ и $P_0 = 0{,}2$. Если посылается сигнал «1», то с вероятностью $p_{11} = 0{,}9$ принимается сигнал «1», с вероятностью $p_{10} = 0{,}1$ принимается сигнал «0». Если посылается сигнал «0», то с вероятностями $p_{01} = 0{,}3$ принимается сигнал «1», $p_{00} = 0{,}7$ принимается сигнал «0». Какова условная вероятность того, что посылается сигнал «0» при условии, что принимается сигнал «1»?

8. *Красота до вечера, а доброта навек.*

9. 32 16 31 56 51 11 42 56

10. KKKKKLLLLLJSDPPPJ

## Вариант 16

1. а) 969; б) 549; в) 973,375; г) 508,5; д) 281,09.

2. а) $10100010_2$; б) $1110010111_2$; в) $110010010{,}101_2$; г) $1111011100{,}10011_2$; д) $605{,}02_8$; е) $3C8{,}8_{16}$.

3. $1111_2 - 101_2$

4. $26_{16}$, $26_8$, $11101_2$.

5. Скорость информационного потока – 20 бит/сек. Сколько времени потребуется для передачи информации объемом в 10 килобайт.

6. Условие задачи см. выше в общая формулировка заданий.

7. По каналу связи с помехами передается одно из двух сообщений:

1) 11111 с вероятностью равной 0,7;

2) 00000 с вероятностью равной 0,3.

Вероятность правильного приема каждого из символов 0 и 1 равна 0,6. Символы искажаются помехами независимо друг от друга. На выходе канала получают кодовое сообщение 10110. Определить вероятности передачи первого и второго сообщений.

8. *На грубое слово не сердись, на ласковое не сдавайся.*

9. 33 11 15 43 12 24 42 56

10. FFFFFOIJHGGGGGGGGHHHHDDDD

## Вариант 17

1. а) 163; б) 566; в) 694,375; г) 352,375; д) 288,61.

2. а) $1001101001_2$; б) $110011101_2$; в) $1000001101{,}01_2$; г) $1010001001{,}11011_2$; д) $247{,}1_8$; е) $81{,}4_{16}$.

3. $11_8 + 11_8$

4. $28_{16}$, $47_8$, $101010_2$.

5. Подсчитать в килобайтах количество информации в тексте, если текст состоит из 600 символов, а мощность используемого алфавита – 128 символов.

6. Условие задачи см. выше в общая формулировка заданий.

7. По каналу связи с помехами передается одно из двух сообщений:

1) 11111 с вероятностью равной 0,8;

2) 00000 с вероятностью равной 0,2.

Вероятность правильного приема каждого из символов 0 и 1 равна 0,6. Символы искажаются помехами независимо друг от друга. На выходе канала получают кодовое сообщение 11100. Определить вероятности передачи первого и второго сообщений.

8. *Вернемся к нашим баранам.*

9. 33 11 15 52 24 13 26 11

10. PPPPPKRTYYDDDDDDDDDDDDD

## Вариант 18

1. а) 917; б) 477; в) 74,5; г) 792,25; д) 84,33.

2. а) $1110011100_2$; б) $1111101111_2$; в) $111110100{,}101_2$; г) $110011110{,}1000011_2$; д) $1446{,}62_2$; е) $9C{,}D_{16}$.

3. $3765_8 - 122_8$

4. $28_{16}$, $47_8$, $101010_2$.

5. Информационное сообщение объемом 1.5 Кбайта содержит 3072 символа. Сколько символов содержит алфавит, при помощи которого было записано это сообщение?

6. Условие задачи см. выше в общая формулировка заданий.

7. По каналу связи с помехами передается одно из двух сообщений:

1) 11111 с вероятностью равной 0,6;

2) 00000 с вероятностью равной 0,4.

Вероятность правильного приема каждого из символов 0 и 1 равна 0,8. Символы искажаются помехами независимо друг от друга. На выходе канала получают кодовое сообщение 01010. Определить вероятности передачи первого и второго сообщений.

8. *Муж жену любит здоровую, а брат сестру богатую.*

9. 33 11 15 55 32 24 42 56

10. aaaaaooPPPPPJJJJJJa

## Вариант 19

1. а) 477; б) 182; в) 863,25; г) 882,25; д) 75,2.

2. а) $101011100_2$; б) $1000010011_2$; в) $11100011,1_2$; г) $100101010,00011_2$; д) $1762,7_8$; е) $1B5,6_{16}$.

3. $101101_2 - 11011_2$

4. $81_{16}$, $172_8$, $1110011_2$.

5. Проводят две лотереи: «4 из 32» и «5 из 64» Сообщение о результатах какой из лотерей несет больше информации?

6. Условие задачи см. выше в общая формулировка заданий.

7. По каналу связи с помехами передается одно из двух сообщений:

1) 11111 с вероятностью равной 0,1;

2) 00000 с вероятностью равной 0,9.

Вероятность правильного приема каждого из символов 0 и 1 равна 0,4. Символы искажаются помехами независимо друг от друга. На выходе канала получают кодовое сообщение 01010. Определить вероятности передачи первого и второго сообщений.

8. *Живу, как живётся, а не как люди хотят.*

9. 34 12 32 34 51 24 42 56

10. NNN0000KKKKFHHHHHHHHH

## Вариант 20

1. а) 804; б) 157; в) 207,625; г) 435,375; д) 30,43.

2. а) $10010000_2$; б) $11001010_2$; в) $1110101100,1011_2$; г) $110110101,10111_2$; д) $1164,36_8$; е) $1D5,C8_{16}$.

3. $100101_2 - 11011_2$

4. $49_{16}$, $102_8$, $1000111_2$.

5. Сколько бит информации несет сообщение о том, что из колоды в 32 карты достали «даму пик»?

6. Условие задачи см. выше в общая формулировка заданий.

7. По каналу связи с помехами передается одно из двух сообщений:

1) 11111 с вероятностью равной 0,5;

2) 00000 с вероятностью равной 0,5.

Вероятность правильного приема каждого из символов 0 и 1 равна 0,5. Символы искажаются помехами независимо друг от друга. На выходе канала получают кодовое сообщение 11011. Определить вероятности передачи первого и второго сообщений.

8. *В дороге и палка пригодится.*

9. 34 12 32 11 33 53 24 26

10. TTTTTTGGGGFJJJJJJH

***Образец оформления титульного лист***

Белорусский национальный технический университет

Международный институт дистанционного образования

Кафедра «Информационные системы и технологии»

Контрольная работа по дисциплине «Теория информации»

за _____ семестр

Вариант _____

Выполнил:

студент _ курса, группы _____

ФИО_____

Проверил:

ФИО преподавателя_____

Минск 20___

## РАЗДЕЛ 4. ВСПОМОГАТЕЛЬНЫЙ

### УЧЕБНАЯ ПРОГРАММА

Учебная программа по учебной дисциплине «Теория информации» разработана для специальности 1-40 01 01 «Программное обеспечение информационных технологий».

Целями освоения дисциплины «Теория информации» являются:

- знакомство студентов с основными положениями теории информации и кодированием, являющимися необходимым компонентом технического образования и освоение которых обеспечит осознанное понимание многих разделов специальных дисциплин.
- научить будущих специалистов применять полученные знания по теории информации и кодирование в практической деятельности;
- привить умение самостоятельно, посредством математического аппарата, осваивать реальные, характерные для специальности задачи;
- развить логическое мышление, аналитические способности, интеллект, необходимые для решения научных и практических задач.

Задачи учебной дисциплины:

- обеспечить овладение студентами теоретических основ дисциплины «Теория информации», добиться знания определений и основных теорем изучаемых разделов курса.
- выработать овладение основными методами решения задач.

В результате изучения дисциплины «Теория информации» формируются следующие компетенции:

СК-10 Применять основные положения теории информации, лежащие в основе современных криптографических преобразований информации, стеганографии и обфускации, для анализа и защиты данных.

В результате изучения учебной дисциплины обучаемый должен:

*знать:*

- основные факты, лежащие в основе построения теории информации и кодирования;
- основные положения и теоремы теории информации и кодирования.

*уметь:*

- применять методологические основы информации в практической деятельности;
- применять алгоритмы кодирования в практической деятельности;
- применять алгоритмы сжатия данных в практической деятельности;
- использовать криптографические методы в решении важных прикладных задач;
- ориентироваться в имеющейся литературе по теории информации и кодировании;

- самостоятельно расширять круг математических знаний по теории информации и кодирования, используя необходимую научную, учебную и справочную литературу.

Согласно учебному плану для заочной (дистанционной) формы получения высшего образования на изучение дисциплины отведено всего 110 часа, из них — **38** часов аудиторных.
Распределение аудиторных часов по курсам, семестрам и видам занятий приведено в таблице 1.

Таблица 1

| Заочная (дистанционная) форма получения высшего образования | | | | |
|---|---|---|---|---|
| Курс | Семестр | Лекции, ч. | Лабораторные занятия, ч. | Форма текущей аттестации |
| 2 | 4 | 8 | 8 | Зачет |

## СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

### Раздел 1. Информация по Хартли. Энтропия Шеннона.

Тема 1. Информация по Хартли.
Информация. Виды информации. Хранение, измерение, обработка и передача информации. Способы измерения информации. Классификация сигналов и их математические модели. Задачи и постулаты прикладной теории информации.

Тема 2. Энтропия как мера степени неопределенности.
Определение энтропии. Свойства энтропии. Энтропия сложной системы. Условная энтропия. Информационная и физическая энтропия. Семантическая информация.

### Раздел 2. Теория информации и оптимальная кодировка.

Тема 3. Сжатие информации.
Системы исчисления. Десятичная, двоичная, восьмеричная и шестнадцатиричная системы исчисления. Код, кодировка. Одноэлементная и многоэлементная кодировки. Классификация кодов. Простейшие алгоритмы сжатия информации. Сжатие информации с потерями. Особенности программ-архиваторов.

Тема 4. Кодирование информации
Общие понятия теории кодирования информации. Избыточность и оптимальное кодирование информации. Арифметическое кодирование. Код Шеннона-Фано, код Хаффмана. Недостатки системы эффективного кодирования. Простейшие алгоритмы сжатия информации. Алгоритмы сжатия изображений без потерь. RLE-кодирование. Алгоритм Лемпеля-Зива (LZ-compression) LZ77. Метод Лемпеля–Зива LZ78. Алгоритм JBIG. Алгоритм Lossless JPEG

Раздел 3. Криптология.

Тема 5. Элементы криптологии
Криптографические средства с древнего времени. Секретность и имитостойкость. Основные идеи криптологии. Криптография и криптоанализ.

**УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА**

| № пп | Наименование раздела, темы | Всего (часы) | Лекции (часы) | Практи-ческие занятия (часы) | Лабора-торные занятия (часы) |
|---|---|---|---|---|---|
| **1** | **2** | **3** | **4** | **5** | **6** |
| 1 | Раздел: Информация по Хартли.<br>Тема: Информация. Виды информации. Хранение, измерение, обработка и передача информации. Способы измерения информации. | 4 | 2 | | 2 |
| 2 | Раздел: Энтропия как мера степени неопределенности.<br>Тема: Определение энтропии. Свойства энтропии. Энтропия сложной системы. Условная энтропия.. | 4 | 2 | | 2 |
| 3 | Раздел: Сжатие информации<br>Тема: Системы исчисления. Десятичная, двоичная, восьмеричная и шестнадцатиричная системы исчисления. Код, кодировка. Одноэлементная и многоэлементная кодировки. Классификация кодов. Простейшие алгоритмы сжатия информации. | 4 | 2 | | 2 |
| 4 | Раздел: Кодирование информации<br>Тема: Общие понятия теории кодирования информации. Простейшие алгоритмы сжатия информации. | 2 | 2 | | |
| 5 | Раздел: Элементы криптологии<br>Тема: Шифрование текстовой информации | | | | 2 |
| | ИТОГО: | 16 | 8 | | 8 |

## 3. ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

ЛИТЕРАТУРА

ОСНОВНАЯ

1. Блинова И.В., Попов И.Ю. Теория информации. Учебное пособие. – СПб: Университет ИТМО, 2018. – 84 с.

2. Шавенько Н.К., Основы теории информации и кодирования. Учебное пособие. – М,: Изд-во МИИГАиК, 2012. – 125 с.

3. Бурькова Е.В., Теория информации: методические указания / Е.В. Бурькова; Оренбургский гос. ун-т. – Оренбург: ОГУ, 2018. – 50 с.

4. Гошин, Е. В. Практикум по теории информации и кодирования**:** учеб. пособие / Е.В. Гошин. – Самара: Изд-во Самарского ун-та, 2018. – 80 с.: ил.

5. Лидовский В. В. Теория информации: Учебное пособие. — М.: Компания Спутник+, 2004. — 111 с. — ISBN 5-93406-661-7.

ДОПОЛНИТЕЛЬНАЯ

6. Крупенкова, Т. Г. Криптографические средства защиты информации. В 2 ч. Ч. 1 [Электронный ресурс] : учебно-методическое пособие для студентов спец. 1-38 02 03 "Техническое обеспечение безопасности" специализации 1-38 02 03 02 ""Аппаратно-программные средства защиты компьютерной информации"/ Т. Г. Крупенкова ; Белорусский национальный технический университет, Кафедра "Инженерная математика". – Минск : БНТУ, 2012.

7. Мотовилова О.В. Основы теории информации: Учебно-методическое пособие для студентов и преподавателей специальности 230701 Прикладная информатика (по отраслям) учреждений среднего профессионального образования . – Ростов н/Д, 2012. – 95 с.

8. Митюхин, А. И. Прикладная теория информации : учеб.- метод. пособие / А. И. Митюхин. – Минск : БГУИР, 2018. – 168 с. : ил.

## Вопросы к зачету

1. Понятие системы счисления
2. Непозиционный системы счисления
3. Позиционные системы счисления
4. Двоичная система счисления
5. Шестнадцатеричная система счисления
6. Восьмеричная система счисления
7. Основные понятия теории информации
8. Общие сведения о передаче информации
9. Классификация сигналов и их математические модели
10. Детерминированные и случайные сигналы
11. Периодические и непериодические сигналы
12. Импульсные сигналы
13. Задачи и постулаты прикладной теории информации
14. Энтропия. Количество информации. Единицы измерения информации
15. Свойства энтропии
16. Энтропия сложной системы
17. Условная энтропия
18. Частная информация о системе
19. Информационные характеристики каналов связи
20. Общие понятия теории кодирования информации
21. Избыточность и оптимальное кодирование информации
22. Метод Шеннона-Фано
23. Метод Хаффмана
24. Избыточность и оптимальное кодирование
25. Префиксные коды
26. Недостатки системы эффективного кодирования
27. Простейшие алгоритмы сжатия информации
28. Алгоритмы сжатия изображений без потерь
29. RLE-кодирование
30. Алгоритм Лемпеля-Зива (LZ-compression) LZ77
31. Метод Лемпеля–Зива LZ78
32. Алгоритм JBIG
33. Алгоритм Lossless JPEG
34. Шифрование текстовой информации
35. Шифры простой замены
36. Шифры сложной замены

# ГЛОССАРИЙ

**Алфавит** – набор знаков, в котором установлен порядок их следования (лексикографический порядок).

**Анализ** – метод исследования, основанный на выделении отдельных компонентов системы и рассмотрении их свойств и связей.

**Байт** – часть машинного слова, состоящая обычно из восьми битов

**Бит –** единица измерения энтропии при двух возможных равновероятных исходах опыта.

**Бод** (англ. baud) в связи и электронике – единица измерения символьной скорости, количество изменений информационного параметра несущего периодического сигнала в секунду. Названа по имени Эмиля Бодо, изобретателя кода Бодо – кодировки символов для телетайпов. В системах с двоичным кодированием, когда каждый бит может принимать только два значения (0 и 1), **бод** соответствует количеству бит, переданных в секунду.

**Вероятность** – число между 0 и 1, ассоциируемое с событием, которое является одним из множества возможных; событие, которое обязательно происходит, имеет вероятность 1. Вероятность события имеет ограниченную величину, определяемую относительной частотой

**Внешние запоминающие устройства (ВЗУ)** – устройства, выполняющие операции, связанные с сохранения и считывания данных на материальном носителе.

**Данные** – это сведения, характеризующие какую-то систему, явление, процесс или объект, представленные в определенной форме и предназначенные для дальнейшего использования.

**Датчик –** любое устройство, которое преобразует энергию в форме звука, света, давления и т.д. в эквивалентный электрический сигнал, или наоборот. Например, фотоэлемент преобразует световое или ультрафиолетовое излучение в электрическую энергию, пьезоэлемент преобразует механическое усилие в электрическую энергию (и наоборот).

**Декодирование** – операция, обратная кодированию, т.е. восстановление информации в первичном алфавите по полученной последовательности кодов.

**Демодулятор –** устройство, преобразующее входные аналоговые сигналы в выходные цифровые. Принцип действия демодулятора обратен

**Дискретные устройства** – устройства, у которых дискретны множества внутренних состояний, входных и выходных сигналов, а также множество моментов времени, в которые поступают входные сигналы, меняются внутренние состояния и выдаются выходные сигналы.

**Длина кода** – число знаков, применяемых для представления кодируемой информации.

**Документ** – продукт, сформированный в результате исполнения некоторой программы.

**Запись логическая** – поименованная совокупность элементарных данных, имеющая смысловую завершенность.

**Запись физическая** – элемент поверхности носителя, на котором в соответствии с физическими принципами функционирования носителя размещаются данные, составляющие логическую запись.

**Запоминающие устройства с произвольным доступом** – те, в которых доступ к данным осуществляется по адресу ячейки, где они хранятся.
**Знак** – элемент некоторого конечного множества отличных друг от друга сущностей, используемого для представления дискретных сигналов.
**Избыточность кода относительная** – характеристика, показывающая, во сколько раз требуется удлинить сообщение, чтобы обеспечить его надежную (безошибочную) передачу (хранение).
**Информатика** – фундаментальная естественная наука, изучающая общие свойства информации, процессы, методы и средства ее обработки (сбор, хранение, преобразование, перемещение, выдача) (определение А.П. Ершова и Б.Н. Наумова).
**Информация (статистическое определение)** – это содержание сообщения, понижающего неопределенность некоторого опыта с неоднозначным исходом; убыль связанной с ним энтропии является количественной мерой информации.
**Информационный процесс** – это изменение с течением времени содержания информации или представляющего его сообщения.
**Искажение –** нежелательные изменения формы сигнала, возникающие между двумя точками в системах передачи.
**Источник информации** – это субъект или объект, порождающий информацию и представляющий ее в виде сообщения.
**Канал передачи данных** – путь передачи информации со всеми необходимыми схемами, который используется для пересылки данных между системами или частями системы. В случае интерфейса, состоящего из нескольких параллельных каналов, каждый канал выделяется для передачи информации одного типа, например, данных или сигналов управления.
**Квантование** – процесс формирования дискретного представления количественной характеристики, которая обычно имеет непрерывный вид. Различают квантование по уровню и времени. Уровень квантования – одно из значений непрерывнонго сигнала, полученное в результате его квантования, шаг квантования – разность между соседними уровнями квантования. Квантованием во времени называется измерение в дискретные промежутки времени амплитуды непрерывного сигнала (термин квантование синонимичен термину дискретизация). В результате замены мгновенного значения сигнала $U$ соответствующим уровнем квантования $V$ возникает погрешность $p = U - V$, которую называют ошибкой квантования. Эта погрешность является случайной величиной. При квантовании сигнала $U(t)$ по уровню случайный процесс заменяется ступенчатой зависимостью – $U_1(t)$. Изменяющуюся во времени ошибку квантования $d(t)$, также представляющую собой случайный процесс, называют шумом квантования $d(t) = U(t) - U_1(t)$.
**Класс** – это множество объектов, обладающих одним или несколькими одинаковыми атрибутами; эти атрибуты называются полем свойств класса.
**Классификаци**я – это распределение однотипных объектов в соответствии с выделенными свойствами (признаками, категориями, классами).
**Код** – (1) правило, описывающее соответствие знаков или их сочетаний одного алфавита знакам или их сочетаниям другого алфавита.
(2) знаки вторичного алфавита, используемые для представления знаков или их сочетаний первичного алфавита.

**Кодирование** – перевод информации, представленной посредством первичного алфавита, в последовательность кодов.
**Коэффициент сжатия** – отношение длин сообщения до и после его сжатого кодирования (в общем случае такое кодирование выполняется для укорачивания сообщений).
**Массив** – упорядоченная линейная совокупность однородных данных.
**Материальный носитель информации** – материальный объект или среда, которые служат для представления или передачи информации.
**Машинное слово** – (1) совокупность двоичных элементов, обрабатываемая как единое целое в устройствах и памяти компьютера; (2) данные, содержащиеся в одной ячейке памяти компьютера.
**Моделирование** – построение упрощенного варианта прототипа, обеспечивающего приемлемую для данной задачи точность описания его строения или поведения.
**Моделирование имитационное** – метод исследования, основанный на том, что изучаемый прототип заменяется ее имитатором – натурной или информационной моделью – с которым и проводятся эксперименты с целью получения информации об особенностях прототипа.
**Модель** – это объединение составных частей (элементов) и связей между ними, отражающая существенные для данной задачи свойства прототипа.
**Модель математическая** – это множество элементов произвольной природы, на которых определено конечное множество отношений.
**Модель проверяемая** – та, у которой результат ее использования может быть соотнесен (сравнен) с прототипом.
**Набор знаков** – набор знаков, в котором установлен порядок их следования.
**Полоса пропускания** – диапазон частот с нижним и верхним пределами. Все частоты между этими пределами (но, может быть, и какие-то другие) пропускаются фильтром или каналом с небольшим затуханием.
**Помеха** – любой сигнал, который возникает в электронной или коммуникационной системе и не является передаваемым полезным сигналом (случайное колебание, значение некоторых параметров которого предсказать невозможно). Помехи могут появляться, например, от внешних неблагоприятных воздействий и расстраивать систему, поскольку они могут формировать ложные сигналы, то есть ошибки.
**Помехоустойчивость** – величина интенсивности внешнего воздействия, при которой цифровая схема может работать безошибочно. Логические величины на выходе схемы представляются двумя различными уровнями электрического потенциала. Любая помеха, наведенная в логической схеме внешним воздействием, прибавляется (или вычитается) к передаваемому цифровому логическому сигналу. Запасом по помехоустойчивости является максимальное шумовое напряжение, которое может быть добавлено или вычтено из логического сигнала, и которое не повлияет на пороговое напряжение, необходимое для достижения устойчивого логического состояния.
**Сжатие данных** – любой из многих методов в теории информации, с помощью которого производится кодирование данных с целью сокращения их избыточности
**Сигнал** – изменение характеристики материального носителя, которое используется для представления информации.

**Сигнал непрерывный (аналоговый)** – его параметр может принимать любое значение в пределах некоторого интервала.
**Сигнал дискретный** – его параметр может принимать конечное число значений в пределах некоторого интервала.
**Синтез** – (1) метод исследования (изучения) системы в целом (т.е. компонентов в их взаимосвязи), сведение в единое целое данных, полученных в результате анализа; (2) создание системы путем соединения отдельных компонентов на основании законов, определяющих их взаимосвязь.
**Система** – совокупность взаимодействующих компонентов, каждый из которых в отдельности не обладает свойствами системы в целом, но является ее неотъемлемой частью.
**Система счисления** – это правило записи чисел с помощью заданного набора специальных знаков - цифр.
**Система счисления позиционная** – те, в которых значение каждой цифры в изображении числа определяется ее положением (позицией) в ряду других цифр.
**Сообщение** – последовательность сигналов.
**Сообщения шенноновские** – те, в которых вероятность появления каждого отдельного знака не меняется со временем.
**Структура данных** – перечень объединяемых одиночных данных, их характеристики, а также особенности связей между ними образуют.
**Схема** – это комбинация базисных элементов, в которой выходы одних элементов присоединяются к входам других.
**Теорема Котельникова (теорема отсчетов):** Непрерывный сигнал можно полностью отобразить и точно воссоздать по последовательности измерений или отсчетов величины этого сигнала через одинаковые интервалы времени, меньшие или равные половине периода максимальной частоты, имеющейся в сигнале.
**Терема Шеннона (первая):** при отсутствии помех передачи всегда возможен такой вариант кодирования сообщения, при котором среднее число знаков кода, приходящихся на один знак кодируемого алфавита, будет сколь угодно близко к отношению средних информации на знак первичного и вторичного алфавитов.
**Терема Шеннона (вторая):** при передаче информации по каналу с шумом всегда имеется способ кодирования, при котором сообщение будет передаваться со сколь угодно высокой достоверностью, если скорость передачи не превышает пропускной способности канала.
**Условие Фано:** неравномерный код может быть однозначно декодирован, если никакой из кодов не совпадает с началом какого-либо иного более длинного кода.
**Файл** – определенным образом оформленная совокупность физических записей, рассматриваемая как единое целое и имеющая описание в системе хранения информации.
**Формальный язык** – конечное или бесконечное подмножество множества всех слов, образованных из некоторого конечного набора символов. Множество называется алфавитом языка. Таким образом, в теории формальных языков под языком понимается просто совокупность строк без всякой связи с их возможной семантикой.
**Ширина полосы пропускания** – интервал частот, используемый данным каналом связи для передачи сигналов.

**Шифрование** – обратимое преобразование информации в целях сокрытия от неавторизованных лиц, с предоставлением, в это же время, авторизованным пользователям доступа к ней. Главным образом, шифрование служит задачей соблюдения конфиденциальности передаваемой информации. Важной особенностью любого алгоритма шифрования является использование ключа, который утверждает выбор конкретного преобразования из совокупности возможных для данного алгоритма.

**Экономичность системы счисления** – то количество чисел, которое можно записать в данной системе с помощью определенного количества цифр.

**Энтропия** есть мера неопределенности опыта, в котором проявляются случайные события, равная средней неопределенности всех возможных его исходов.

# ПРИЛОЖЕНИЕ 1

Таблица величин $\eta(p) = -plogp$

| p | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| **0,00** | – | 0,0100 | 0,0179 | 0,0251 | 0,0319 | 0,0382 | 0,0443 | 0,0501 | 0,0557 | 0,0612 |
| **0,01** | 0,0664 | 0,0716 | 0,0766 | 0,0815 | 0,0862 | 0,0909 | 0,0955 | 0,0999 | 0,1043 | 0,1086 |
| **0,02** | 0,1129 | 0,1170 | 0,1211 | 0,1252 | 0,1291 | 0,0330 | 0,1369 | 0,1407 | 0,1444 | 0,1481 |
| **0,03** | 0,1518 | 0,1554 | 0,1589 | 0,1624 | 0,1659 | 0,1693 | 0,1727 | 0,1760 | 0,1793 | 0,1825 |
| **0,04** | 0,1858 | 0,1889 | 0,1921 | 0,1952 | 0,1983 | 0,2013 | 0,2043 | 0,2073 | 0,2103 | 0,2132 |
| **0,05** | 0,2161 | 0,2190 | 0,2218 | 0,2246 | 0,2274 | 0,2301 | 0,2329 | 0,2356 | 0,2383 | 0,2409 |
| **0,06** | 0,2435 | 0,2461 | 0,2487 | 0,2513 | 0,2538 | 0,2563 | 0,2588 | 0,2613 | 0,2637 | 0,2661 |
| **0,07** | 0,2686 | 0,2709 | 0,2733 | 0,2756 | 0,2780 | 0,2803 | 0,2826 | 0,2848 | 0,2871 | 0,2893 |
| **0,08** | 0,2915 | 0,2937 | 0,2959 | 0,2980 | 0,3002 | 0,3023 | 0,3044 | 0,3065 | 0,3086 | 0,3106 |
| **0,09** | 0,3127 | 0,3147 | 0,3167 | 0,3187 | 0,3207 | 0,3226 | 0,3246 | 0,3265 | 0,3284 | 0,3303 |
| **0,10** | 0,3322 | 0,3341 | 0,3359 | 0,3378 | 0,3398 | 0,3414 | 0,3432 | 0,3450 | 0,3468 | 0,3485 |
| **0,11** | 0,3503 | 0,3520 | 0,3537 | 0,3555 | 0,3571 | 0,3588 | 0,3605 | 0,3622 | 0,3638 | 0,3654 |
| **0,12** | 0,3671 | 0,3687 | 0,3703 | 0,3719 | 0,3734 | 0,3750 | 0,3766 | 0,3781 | 0,3796 | 0,3811 |
| **0,13** | 0,3826 | 0,3841 | 0,3856 | 0,3871 | 0,3886 | 0,3900 | 0,3915 | 0,3929 | 0,3943 | 0,3957 |
| **0,14** | 0,3971 | 0,3985 | 0,3999 | 0,4012 | 0,4026 | 0,4040 | 0,4053 | 0,4066 | 0,4079 | 0,4092 |
| **0,15** | 0,4105 | 0,4118 | 0,4131 | 0,4144 | 0,4156 | 0,4169 | 0,4181 | 0,4194 | 0,4206 | 0,4218 |
| **0,16** | 0,4230 | 0,4242 | 0,4254 | 0,4266 | 0,4277 | 0,4289 | 0,4301 | 0,4312 | 0,4323 | 0,4335 |
| **0,17** | 0,4346 | 0,4357 | 0,4368 | 0,4379 | 0,4390 | 0,4400 | 0,4411 | 0,4422 | 0,4432 | 0,4443 |
| **0,18** | 0,4453 | 0,4463 | 0,4474 | 0,4484 | 0,4494 | 0,4504 | 0,4514 | 0,4523 | 0,4533 | 0,4543 |
| **0,19** | 0,4552 | 0,4562 | 0,4571 | 0,4581 | 0,4590 | 0,4599 | 0,4608 | 0,4617 | 0,4626 | 0,4635 |
| **0,20** | 0,4644 | 0,4653 | 0,4661 | 0,4670 | 0,4678 | 0,4687 | 0,4695 | 0,4704 | 0,4712 | 0,4720 |
| **0,21** | 0,4728 | 0,4736 | 0,4744 | 0,4752 | 0,4760 | 0,4768 | 0,4776 | 0,4783 | 0,4791 | 0,4798 |
| **0,22** | 0,4806 | 0,4813 | 0,4820 | 0,4828 | 0,4835 | 0,4842 | 0,4849 | 0,4856 | 0,4863 | 0,4870 |
| **0,23** | 0,4877 | 0,4883 | 0,4890 | 0,4897 | 0,4903 | 0,4010 | 0,4916 | 0,4923 | 0,4929 | 0,4935 |
| **0,24** | 0,4941 | 0,4947 | 0,4954 | 0,4960 | 0,4966 | 0,4971 | 0,4977 | 0,4983 | 0,4989 | 0,4994 |
| **0,25** | 0,5000 | 0,5006 | 0,5011 | 0,5016 | 0,5022 | 0,5027 | 0,5032 | 0,5038 | 0,5043 | 0,5048 |
| **0,26** | 0,5053 | 0,5058 | 0,5063 | 0,5068 | 0,5072 | 0,5077 | 0,5082 | 0,5087 | 0,5091 | 0,5096 |
| **0,27** | 0,5100 | 0,5105 | 0,5109 | 0,5113 | 0,5118 | 0,5122 | 0,5126 | 0,5130 | 0,5134 | 0,5138 |
| **0,28** | 0,5142 | 0,5146 | 0,5150 | 0,5154 | 0,5158 | 0,5161 | 0,5165 | 0,5169 | 0,5172 | 0,5176 |
| **0,29** | 0,5179 | 0,5182 | 0,5186 | 0,5189 | 0,5192 | 0,5196 | 0,5199 | 0,5202 | 0,5205 | 0,5208 |
| **0,30** | 0,5211 | 0,5214 | 0,5217 | 0,5220 | 0,5222 | 0,5225 | 0,5228 | 0,5230 | 0,5233 | 0,5235 |
| **0,31** | 0,5238 | 0,5240 | 0,5243 | 0,5245 | 0,5247 | 0,5250 | 0,5252 | 0,5254 | 0,5256 | 0,5258 |
| **0,32** | 0,5260 | 0,5262 | 0,5264 | 0,5266 | 0,5268 | 0,5270 | 0,5272 | 0,5273 | 0,5275 | 0,5277 |
| **0,33** | 0,5278 | 0,5280 | 0,5281 | 0,5283 | 0,5284 | 0,5286 | 0,5287 | 0,5288 | 0,5289 | 0,5290 |
| **0,34** | 0,5292 | 0,5293 | 0,5294 | 0,5295 | 0,5296 | 0,5297 | 0,5298 | 0,5299 | 0,5299 | 0,5300 |
| **0,35** | 0,5301 | 0,5302 | 0,5302 | 0,5303 | 0,5304 | 0,5304 | 0,5305 | 0,5305 | 0,5305 | 0,5306 |

| p | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| **0,35** | 0,5301 | 0,5302 | 0,5302 | 0,5303 | 0,5304 | 0,5304 | 0,5305 | 0,5305 | 0,5305 | 0,5306 |
| **0,36** | 0,5306 | 0,5306 | 0,5307 | 0,5307 | 0,5307 | 0,5307 | 0,5307 | 0,5307 | 0,5307 | 0,5307 |
| **0,37** | 0,5307 | 0,5307 | 0,5307 | 0,5307 | 0,5307 | 0,5306 | 0,5306 | 0,5306 | 0,5305 | 0,5305 |
| **0,38** | 0,5304 | 0,5304 | 0,5303 | 0,5303 | 0,5302 | 0,5302 | 0,5301 | 0,5300 | 0,5300 | 0,5299 |
| **0,39** | 0,5298 | 0,5297 | 0,5296 | 0,5295 | 0,5294 | 0,5293 | 0,5292 | 0,5291 | 0,5290 | 0,5289 |
| **0,40** | 0,5288 | 0,5286 | 0,5285 | 0,5284 | 0,5283 | 0,5281 | 0,5280 | 0,5278 | 0,5277 | 0,5275 |
| **0,41** | 0,5274 | 0,5272 | 0,5271 | 0,5269 | 0,5267 | 0,5266 | 0,5264 | 0,5262 | 0,5260 | 0,5258 |
| **0,42** | 0,5256 | 0,5255 | 0,5253 | 0,5251 | 0,5249 | 0,5246 | 0,5244 | 0,5242 | 0,5240 | 0,5238 |
| **0,43** | 0,5236 | 0,5233 | 0,5231 | 0,5229 | 0,5226 | 0,5224 | 0,5222 | 0,5219 | 0,5217 | 0,5214 |
| **0,44** | 0,5211 | 0,5209 | 0,5206 | 0,5204 | 0,5201 | 0,5198 | 0,5195 | 0,5193 | 0,5190 | 0,5187 |
| **0,45** | 0,5184 | 0,5181 | 0,5187 | 0,5175 | 0,5172 | 0,5169 | 0,5166 | 0,5163 | 0,5160 | 0,5157 |
| **0,46** | 0,5153 | 0,5150 | 0,5147 | 0,5144 | 0,5140 | 0,5137 | 0,5133 | 0,5130 | 0,5127 | 0,5123 |
| **0,47** | 0,5120 | 0,5116 | 0,5112 | 0,5109 | 0,5105 | 0,5102 | 0,5098 | 0,5094 | 0,5090 | 0,5087 |
| **0,48** | 0,5083 | 0,5079 | 0,5075 | 0,5071 | 0,5067 | 0,5063 | 0,5059 | 0,5055 | 0,5051 | 0,5047 |
| **0,49** | 0,5043 | 0,5039 | 0,5034 | 0,5030 | 0,5026 | 0,5022 | 0,5017 | 0,5013 | 0,5009 | 0,5004 |
| **0,50** | 0,5000 | 0,4996 | 0,4991 | 0,4987 | 0,4982 | 0,4978 | 0,4973 | 0,4968 | 0,4964 | 0,4959 |
| **0,51** | 0,4954 | 0,4950 | 0,4945 | 0,4940 | 0,4935 | 0,4930 | 0,4926 | 0,4921 | 0,4916 | 0,4911 |
| **0,52** | 0,4906 | 0,4901 | 0,4896 | 0,4891 | 0,4886 | 0,4880 | 0,4875 | 0,4870 | 0,4865 | 0,4860 |
| **0,53** | 0,4854 | 0,4849 | 0,4844 | 0,4839 | 0,4833 | 0,4828 | 0,4822 | 0,4817 | 0,4811 | 0,4806 |
| **0,54** | 0,4800 | 0,4795 | 0,4789 | 0,4784 | 0,4778 | 0,4772 | 0,4767 | 0,4761 | 0,4755 | 0,4750 |
| **0,55** | 0,4744 | 0,4738 | 0,4732 | 0,4726 | 0,4720 | 0,4714 | 0,4708 | 0,4702 | 0,4797 | 0,4691 |
| **0,56** | 0,4684 | 0,4678 | 0,4672 | 0,4666 | 0,4660 | 0,4654 | 0,4648 | 0,4641 | 0,4635 | 0,4629 |
| **0,57** | 0,4623 | 0,4616 | 0,4610 | 0,4603 | 0,4597 | 0,4591 | 0,4584 | 0,4578 | 0,4571 | 0,4565 |
| **0,58** | 0,4558 | 0,4551 | 0,4545 | 0,4538 | 0,4532 | 0,4525 | 0,4518 | 0,4512 | 0,4505 | 0,4498 |
| **0,59** | 0,4491 | 0,4484 | 0,4477 | 0,4471 | 0,4464 | 0,4457 | 0,4450 | 0,4443 | 0,4436 | 0,4429 |
| **0,60** | 0,4422 | 0,4415 | 0,4408 | 0,4401 | 0,4393 | 0,4386 | 0,4379 | 0,4372 | 0,4365 | 0,4357 |
| **0,61** | 0,4350 | 0,4343 | 0,4335 | 0,4328 | 0,4321 | 0,4313 | 0,4306 | 0,4298 | 0,4291 | 0,4383 |
| **0,62** | 0,4276 | 0,4268 | 0,4261 | 0,4253 | 0,4246 | 0,4238 | 0,4230 | 0,4223 | 0,4215 | 0,4207 |
| **0,63** | 0,4199 | 0,4192 | 0,4184 | 0,4176 | 0,4168 | 0,4160 | 0,4153 | 0,4145 | 0,4137 | 0,4129 |
| **0,64** | 0,4121 | 0,4113 | 0,4105 | 0,4097 | 0,4089 | 0,4080 | 0,4072 | 0,4064 | 0,4056 | 0,4048 |
| **0,65** | 0,4040 | 0,4032 | 0,4023 | 0,4015 | 0,4007 | 0,3998 | 0,3990 | 0,3982 | 0,3973 | 0,3965 |
| **0,66** | 0,3957 | 0,3948 | 0,3940 | 0,3931 | 0,3922 | 0,3914 | 0,3905 | 0,3897 | 0,3888 | 0,3880 |
| **0,67** | 0,3871 | 0,3862 | 0,3954 | 0,3845 | 0,3836 | 0,3828 | 0,3819 | 0,3810 | 0,3801 | 0,3792 |
| **0,68** | 0,3784 | 0,3775 | 0,3766 | 0,3757 | 0,3748 | 0,3739 | 0,3730 | 0,3721 | 0,3712 | 0,3703 |
| **0,69** | 0,3694 | 0,3685 | 0,3676 | 0,3666 | 0,3657 | 0,3648 | 0,3639 | 0,3630 | 0,3621 | 0,3611 |
| **0,70** | 0,3602 | 0,3593 | 0,3583 | 0,3574 | 0,3565 | 0,3555 | 0,3546 | 0,3536 | 0,3527 | 0,3518 |
| **0,71** | 0,3508 | 0,3499 | 0,3489 | 0,3480 | 0,3470 | 0,3461 | 0,3451 | 0,3441 | 0,3432 | 0,3422 |
| **0,72** | 0,3412 | 0,3403 | 0,3393 | 0,3383 | 0,3373 | 0,3364 | 0,3354 | 0,3344 | 0,3334 | 0,3324 |
| **0,73** | 0,3314 | 0,3304 | 0,3295 | 0,3285 | 0,3275 | 0,3265 | 0,3255 | 0,3245 | 0,3235 | 0,3225 |
| **0,74** | 0,3215 | 0,3204 | 0,3194 | 0,3184 | 0,3174 | 0,3164 | 0,3154 | 0,3144 | 0,3133 | 0,3123 |
| **0,75** | 0,3113 | 0,3103 | 0,3092 | 0,3082 | 0,3071 | 0,3061 | 0,3051 | 0,3040 | 0,3030 | 0,3019 |

| p | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| **0,76** | 0,3009 | 0,2999 | 0,2988 | 0,2978 | 0,2967 | 0,2956 | 0,2946 | 0,2935 | 0,2925 | 0,2914 |
| **0,77** | 0,2903 | 0,2893 | 0,2882 | 0,2871 | 0,2861 | 0,2850 | 0,2839 | 0,2828 | 0,2818 | 0,2807 |
| **0,78** | 0,2796 | 0,2785 | 0,2774 | 0,2763 | 0,2853 | 0,2741 | 0,2731 | 0,2720 | 0,2709 | 0,2698 |
| **0,79** | 0,2687 | 0,2676 | 0,2664 | 0,2653 | 0,2642 | 0,2631 | 0,2620 | 0,2609 | 0,2598 | 0,2587 |
| **0,80** | 0,2575 | 0,2564 | 0,2553 | 0,2542 | 0,2431 | 0,2519 | 0,2508 | 0,2497 | 0,2485 | 0,2474 |
| **0,81** | 0,2462 | 0,2451 | 0,2440 | 0,2428 | 0,2417 | 0,2405 | 0,2394 | 0,2382 | 0,2371 | 0,2359 |
| **0,82** | 0,2348 | 0,2336 | 0,2324 | 0,2313 | 0,2301 | 0,2290 | 0,2278 | 0,2268 | 0,2255 | 0,2243 |
| **0,83** | 0,2231 | 0,2220 | 0,2208 | 0,2196 | 0,2184 | 0,2172 | 0,2160 | 0,2149 | 0,2137 | 0,2125 |
| **0,84** | 0,2113 | 0,2101 | 0,2089 | 0,2077 | 0,2065 | 0,2053 | 0,2041 | 0,2029 | 0,2017 | 0,2005 |
| **0,85** | 0,1993 | 0,1981 | 0,1969 | 0,1957 | 0,1944 | 0,1932 | 0,1920 | 0,1908 | 0,1896 | 0,1884 |
| **0,86** | 0,1871 | 0,1859 | 0,1847 | 0,1834 | 0,1822 | 0,1810 | 0,1797 | 0,1785 | 0,1773 | 0,1760 |
| **0,87** | 0,1748 | 0,1735 | 0,1723 | 0,1711 | 0,1698 | 0,1686 | 0,1673 | 0,1661 | 0,1648 | 0,1635 |
| **0,88** | 0,1623 | 0,1610 | 0,1598 | 0,1585 | 0,1572 | 0,1560 | 0,1547 | 0,1534 | 0,1522 | 0,1509 |
| **0,89** | 0,1496 | 0,1484 | 0,1471 | 0,1458 | 0,1445 | 0,1432 | 0,1419 | 0,1407 | 0,1394 | 0,1381 |
| **0,90** | 0,1368 | 0,1355 | 0,1342 | 0,1329 | 0,1316 | 0,1303 | 0,1290 | 0,1277 | 0,1264 | 0,1251 |
| **0,91** | 0,1238 | 0,1225 | 0,1212 | 0,1199 | 0,1186 | 0,1173 | 0,1159 | 0,1146 | 0,1133 | 0,1120 |
| **0,92** | 0,1107 | 0,1094 | 0,1080 | 0,1067 | 0,1054 | 0,1040 | 0,1027 | 0,1014 | 0,1000 | 0,1987 |
| **0,93** | 0,0974 | 0,0960 | 0,0947 | 0,0933 | 0,0920 | 0,0907 | 0,0893 | 0,0880 | 0,0868 | 0,0853 |
| **0,94** | 0,0839 | 0,0826 | 0,0812 | 0,0798 | 0,0785 | 0,0771 | 0,0758 | 0,0744 | 0,0730 | 0,0717 |
| **0,95** | 0,0703 | 0,0689 | 0,0676 | 0,0662 | 0,0648 | 0,0634 | 0,0621 | 0,0607 | 0,0593 | 0,0579 |
| **0,96** | 0,0565 | 0,0552 | 0,0538 | 0,0524 | 0,0510 | 0,0496 | 0,0482 | 0,0468 | 0,0454 | 0,0440 |
| **0,97** | 0,0426 | 0,0412 | 0,0398 | 0,0384 | 0,0370 | 0,0356 | 0,0342 | 0,0328 | 0,0314 | 0,0300 |
| **0,98** | 0,0286 | 0,0271 | 0,0257 | 0,0243 | 0,0230 | 0,0214 | 0,0201 | 0,0186 | 0,0172 | 0,0158 |
| **0,99** | 0,0140 | 0,0129 | 0,0115 | 0,0101 | 0,0086 | 0,0072 | 0,0058 | 0,0043 | 0,0029 | 0,0014 |

## ПРИЛОЖЕНИЕ 2[17]

# Семантическая информация

Смысл сообщений не имеет никакого отношения к теории информации, целиком построенной на положениях теории вероятностей. Но в 50-х годах двадцатого века Бар-Хиллелом и Карнапом была предложена теория семантической информации. Семантическая информация трактовалась авторами, как синоним смыслового содержания.

Рассматривается предложение $s$. $p(s)$ – логическая вероятность предложения $s$. Предложены две основные меры семантической информации. Первая из них $cont(s) = 1 - p(s)$. Вторая $inf(s) = \log \frac{1}{1 - cont(s) = \log \frac{1}{p(s)} = -\log p(s)}$

Примером одной из таких мер является функция $inf(s) = -\log p(s)$, где $s$ – это предложение, смысловое содержание которого измеряется, $p(s)$ – вероятность истинности $s$. Некоторые свойства этих функций-мер:

**1.** если $s_1 \to s_2$ (из $s_1$ следует $s_2$) – истино, то $inf(s_1) \geqslant inf(s_2)$;

**2.** $inf(s) \geqslant 0$, $cont(s) \geqslant 0$;

**3.** если $s$ – истино, то $inf(s) = 0$;

**4.** Для двух логически независимых предложений $inf(s_1 \wedge s_2) = inf(s_1) + inf(s_2)$, но $cont(s_1 \wedge s_2) < cont(s_1) + cont(s_2)$, где $\wedge$ – знак логической связки "И".

Значение этой функции-меры $inf$ больше для предложений, исключающих большее количество возможностей.

**Пример 1.** Из $s_1$ – "$a < 8$"и $s_2$ – "$a = 3$"следует, что $s_2 \to s_1$, т.е. $inf(s_2) \geqslant inf(s_1)$. Действительно, $s_2$ исключает больше возможностей, чем $s_1$.

**Пример 2.** Известно, что высказываение $s_1$ истино на 50%, а высказывание $s_2$ истино на 25%. Найти $inf(s)$ и $cont(s)$ предложений $s_1$ и $s_2$.

$inf(s_1) = -\log(\frac{1}{2}) = 1$, $inf(s_2) = -\log(\frac{1}{4}) = 2$,

$cont(s_1) = 1 - \frac{1}{2} = \frac{1}{2}$, $cont(s_2) = 1 - \frac{1}{4} = \frac{3}{4}$.

[17] Блинова И.В., Попов И.Ю. Теория информации. Учебное пособие. – СПб: Университет ИТМО, 2018. – 84 с.