

# Основы теории информации и кодирования. Измерение информации. Кодирование. Форматы файлов.

Александра Игоревна Кононова / [illinc@mail.ru](mailto:illinc@mail.ru)  
+7-985-148-32-64 (телефон), +7-977-977-97-29 (WhatsApp),  
[gitlab.com/illinc/raspisanie](https://gitlab.com/illinc/raspisanie)

МИЭТ

25 сентября 2021 г. — актуальную версию можно найти на  
<https://gitlab.com/illinc/otik>

# Регламент

См. <https://gitlab.com/illinc/otik/>

Дополнительные баллы:

- 1 бонусные задания л/р;
- 2 вычитка материала — 1 – 4 балла за принятое замечание, 2 – 8 за принятое исправление;
- 3 пополнение списка литературы — (–1) – (+8) баллов.

Экзамен (оценка):

5 86 – 100

4 70 – 85

3 50 – 69

2 0 – 49

Консультации — см. [gitlab.com/illinc/raspisanie](https://gitlab.com/illinc/raspisanie)

# Предмет теории информации. Источник информации

**Теория информации** — математическая теория, посвящённая измерению информации, её потока, «размеров» канала связи и т. п., особенно применительно к средствам связи:

$$x \in X \sim I(x)$$

$x$  — сообщение,  $X = \{x, p(x)\}$  — источник (сл. процесс/сл. величина).

Дискретное  $x$  может состоять из символов или быть отдельным символом.

**Информация** — нематериальная сущность, при помощи которой с любой точностью можно описывать реальные (материальные), виртуальные (возможные) и понятийные сущности.

$I(x)$ : ① **Новизна** (неизмеряемость в быту).

② **Объёмный** (длина — измерение в технике).

③ **Вероятностный** (снятая неопределённость — измерение в ТИ).

# Энтропия и информация

1865 г. — Рудольф Клаузиус ввёл в статистическую физику понятие **энтропии** — меры уравниваемости [Дж/К].

1877 г. — Людвиг Больцман установил связь энтропии с вероятностью.

1901 г. — Макс Планк определил энтропию как  $H = k \cdot \ln(\Omega)$ , где  $k$  — коэффициент Больцмана [Дж/К].

1921 г. — Роналд Фишер ввёл термин «информация» (информация, которую можно извлечь из имеющихся данных, **имеет предел**).

1928 г. — Ральф Хартли — логарифмическая мера информации для **равновероятных** событий.

1948 г. — Клод Шеннон — вычисление количества информации и энтропии.

Основное соотношение между энтропией и информацией:

$$I + \frac{\log_2 e}{k} H = \text{const} \quad [\text{бит}] \quad \left( \frac{dI}{dt} = - \frac{\log_2 e}{k} \frac{dH}{dt} \quad [\text{бит/с}] \right).$$

Предмет теории информации. Источник информации

Вероятностная мера информации

Задачи: измерение информации

Кодирование и структуры данных

Простые коды (1)

Простые коды (2)

Энтропия и информация

Виды источников информации

Единица измерения информации

Требования к мере информации  $I(x)$

# Виды источников информации

По сообщениям:

- дискретные (цифровые)/непрерывные (аналоговые);
- дискретные: качественные/количественные.

Элемент качественной информации — **символ**  $a \in A$  (множество  $A$  — алфавит);  
конечная последовательность символов — **слово**  $x \in A^+$  (строка, фраза).

Источник символов алфавита  $A$  (можно прочитать строку):

- 1 **стационарный** (вероятность символа не зависит от времени/позиции: только от контекста) / **нестационарный** (при сдвиге вероятности меняются);
- 2 **марковский** источник — вероятность символа определяется состоянием; состояние изменяется после порождения символа (новое состояние однозначно определяется предыдущим и порождённым символом); марковский источник порядка  $m$  — вероятность символа на  $i$ -м шаге зависит от  $m$  предыдущих символов:  $i-1, i-2, \dots, i-m$ ;
- 3 **стационарный источник без памяти** — вероятность символа  $a \in A$  постоянна (равна  $p(a)$ );
- 4 **равновероятный источник** — вероятность символа  $a \in A$  постоянна и одинакова для всех символов (равна  $\frac{1}{|A|}$ );

равновероятный  $\subseteq$  стационарный без памяти  $\subseteq$  марковский  $\subseteq$  стационарный



Предмет теории информации. Источник информации

Вероятностная мера информации

Задачи: измерение информации

Кодирование и структуры данных

Простые коды (1)

Простые коды (2)

Энтропия и информация

Виды источников информации

Единица измерения информации

Требования к мере информации  $I(x)$

# Единица измерения информации

**Бит** — количество информации в сообщении, уменьшающем неопределённость знания в два раза.

**Источник с двумя равновероятными состояниями** — симметричная монета

.	?	2 возможных варианта
P	Решка	1 вариант

---

Неопределённость уменьшилась в 2 раза:  $I(P) = 1$  бит

..	Две симметричные монеты	
0.	Первая — вверх орлом	2 раза (+1 бит)
0P	Вторая — вверх решкой	2 раза (+1 бит)
4 возможных варианта		$I(OP) = 2$ бита

# Требования к мере информации $I(x)$

- ❶  $I(x) \geq 0$ .
- ❷ Вероятностный подход:  $I(x) = f(p_x)$ .
- ❸ Объёмный подход:  $I(x)$  монотонно связана с затратами на передачу
  - два равновероятных сообщения — 0 и 1 (1 бит),  
четыре — 00, 01, 10, 11 (2 бита) и т. д.:  
 $f\left(\frac{1}{2}\right) = 1, \quad f\left(\frac{1}{4}\right) = 2, \quad f\left(\frac{1}{8}\right) = 3, \dots$
  - затраты на передачу независимых сообщений складываются:  
 $I(x_1, \dots, x_n) = I(x_1) + \dots + I(x_n)$   
при этом вероятности независимых событий умножаются  
 $f(p_1 \times \dots \times p_n) = f(p_1) + \dots + f(p_n)$ .

# Формула Хартли для равновероятных событий

Источник  $X$  порождает  $N$  **равновероятных** сообщений  $x$   
 $(\forall x \in X : p(x) = p = \frac{1}{N})$ .

$$I(x) = I(X) = I = \log_2 N = -\log_2(p) \quad \text{или} \quad 2^I = N$$

где  $I(x)$  — количество информации в сообщении  $x$ ;

$I(X)$  — **среднее** кол-во информации в одном сообщении источника  $X$ .

Если  $N = 2$ , то  $I = 1$  бит.

Подбрасывание монеты

.. 4 варианта 2 бита

Угадывание слов по словарю

..... 175 слов 7,5 бит

.а.и.а 122 слова 6,9 бит

р.б.т. 4 слова 2 бита



# Формула Шеннона для неравновероятных событий

Количество информации  $I$  в сообщении с вероятностью  $p(x)$ :

$$I(x) = -\log_2 p(x)$$

Свойства:

- ❶ Неотрицательность:  $I(x) \geq 0, x \in X$ .
- ❷ Монотонность:  $x_1, x_2 \in X, p(x_1) \geq p(x_2) \rightarrow I(x_1) \leq I(x_2)$ .
- ❸ Аддитивность: для независимых сообщений  $x_1, \dots, x_n$

$$I(x_1, \dots, x_n) = \sum_{i=1}^n I(x_i)$$

- ❹ Для равновероятных событий соответствует формуле Хартли.

Среднее количество информации дискретного источника  $X = \{x, p(x)\}$ :

$$I(X) = \sum_{x_i \in X} \left( p(x_i) \cdot I(x_i) \right) = - \sum_{x_i \in X} \left( p(x_i) \cdot \log_2 p(x_i) \right)$$

# Количество информации в тексте

Из источника символов  $X$  можно прочитать текст

$$\vec{x} = x_1 x_2 \dots x_k$$

- ❶ Источник без памяти: сообщения  $x_1, x_2, \dots, x_k$  независимы

$$p(\vec{x}) = p(x_1) \cdot p(x_2) \cdot \dots \cdot p(x_k)$$

$$I(\vec{x}) = I(x_1) + I(x_2) + \dots + I(x_k)$$

- ❷ Источник с памятью:

$$p(\vec{x}) = p(x_1) \cdot p(x_2|x_1) \cdot \dots \cdot p(x_k|x_1 x_2 \dots x_{k-1})$$

$$I(\vec{x}) = I(x_1) + I(x_2|x_1) + \dots + I(x_k|x_1 x_2 \dots x_{k-1})$$

Если источник марковский порядка  $m$ :

$$I(\vec{x}) = I(x_1) + \dots + I(x_i|x_{i-m} \dots x_{i-1}) + \dots + I(x_k|x_{k-1} \dots x_{k-1})$$

## Модель источника: $X$ неизвестен

Оценка алфавита  $A_1$  и вероятностей источника по сообщению:  $x = \text{«молоко»}$

- ①  $A_1$  — кои-8, равновероятные символы:  $p = \frac{1}{256}$ ,  $I(x) = 6 \cdot \log_2(256) = 48$  (бит)
- ②  $A_1$  — русский алфавит, равновероятные:  $p = \frac{1}{33}$ ,  $I(x) = 6 \cdot \log_2(33) \approx 30,3$
- ③  $A_1$  — Unicode 12.1, равновероятные:  $p = \frac{1}{137994}$ ,  $I(x) \approx 6 \cdot 17,1 \approx 102,4$
- ④  $A_1 = \{\text{к, л, м, о}\}$ , равновероятные:  $p = \frac{1}{4}$ ,  $I(x) = 6 \cdot \log_2(4) = 12$
- ⑤  $A_1 = \{\text{к, л, м, о}\}$  или кои-8, неравновероятные, стац-й источник без памяти:  
о (3) + к (1) + л (1) + м (1):  $p(o) = \frac{3}{6}$ ,  $p(k) = p(l) = p(m) = \frac{1}{6}$   
 $I(x) = -3 \cdot \log_2(\frac{3}{6}) - \log_2(\frac{1}{6}) - \log_2(\frac{1}{6}) - \log_2(\frac{1}{6}) = 3 \cdot \log_2(2) + 3 \cdot \log_2(6) \approx 10,8$
- ⑥  $A_1 = \{\text{к, л, м, о}\}$  или кои-8, марковский источник первого порядка:  

предыдущий	$p(k)$	$p(l)$	$p(m)$	$p(o)$
—	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
к, л, м	0	0	0	1
о	$\frac{1}{2}$	$\frac{1}{2}$	0	0

$$I(x) = -\log_2(\frac{1}{4}) - \log_2(1) - \log_2(\frac{1}{2}) - \log_2(1) - \log_2(\frac{1}{2}) - \log_2(1) = 2 + 1 + 1 = 4$$
- ⑦  $A_1 = \{\text{молоко, чай}\}$ , равновероятные символы:  $p = \frac{1}{2}$ ,  $I(x) = 1$
- ⑧  $A_1 = \{\text{молоко}\}$ :  $p = 1$ ,  $I(x) = 0$

# Задачи (равновероятный источник)

- 1 Найти количество информации в событии «три разные симметричные монеты выпали все вверх решкой».
- 2 Найти количество информации в источнике «три разные симметричные монеты».

## Задачи (стационарный источник без памяти)

- 1 Найти количество информации в событии «две из трёх неразличимых симметричных монет выпали вверх решкой, третья — орлом».
- 2 Найти количество информации в источнике «три неразличимые симметричные монеты».
- 3 Найти количество информации в событии «из урны с 3 белыми и 5 чёрными шарами извлекли чёрный шар».
- 4 Найти количество информации в событии «из урны с 3 белыми и 5 чёрными шарами извлекли белый шар».
- 5 Найти количество информации в источнике «урна с 3 белыми и 5 чёрными шарами».

# Задачи (стационарный источник с памятью)

- ❶ Источник  $X$  генерирует последовательность подстрок «хрю» и «мяу» (с равной вероятностью), не разделяя их (например, «хрюхрюхрюмяухрюмяумяухрюмяумя...»). Из случайного места последовательности (не обязательно с начала подстроки) читается три символа подряд (сообщение  $x$ ). Найти количество информации в событии « $x = \text{рюх}$ ».
- ❷ Источник  $X$  аналогично генерирует посл-ть «ку» и «кукареку» (например, «кукукукукарекукукукукарекукукукукарекукукукукареку...»). Из случайного места посл-ти читается два (три) символа подряд ( $x$ ). Найти количество информации в событиях:

- $x = \text{ка};$
- $x = \text{ку};$
- $x = \text{ек}.$
- $x = \text{кар};$
- $x = \text{ук};$

Подсказка: основная проблема в том, что часть символов — одинаковые. Пусть они разные...

Или: пусть всего  $2N \gg 1$  слов, то есть  $N$  «ку» и  $N$  «кукареку»...



# Задачи (построение модели источника)

Оценить алфавит и построить модели источника: а) равновероятную, б) стационарную без памяти, в) марковскую первого порядка для сообщения  $x$ , по модели оценить  $I(x)$  и  $I(y)$ .

- ❶  $x = \text{хрюхрюхрюмяухрюмяумяухрюмяумя}$  (30 символов, 5 «хрю» (0) и 5 «мяу» (1) 0001011011);  $y = \text{рюх}$ .

В тексте 5 двухбуквенных сочетаний, начинающихся с «ю»:  
2 «юх» и 3 «юм»

- ❷  $x = \text{кукукукукарекукукукарекукукукарекукукукарекукукукареку}$  (50 символов, 5 «ку» (0) и 5 «кукареку» (1) аналогично);  
 $y = \text{кар}$ .

# Кодирование и структуры данных

**Кодирование** — преобразование дискретной информации

$$x \in X = A_1^+ \rightarrow \text{code}(x) \in A_2^+$$

смена алфавита, **сжатие, защита от шума**, шифрование.

**Декодирование** — обратное преобразование  $\text{code}(x) \rightarrow x$

$x$  — сообщение, исходный текст, исходная строка, блок;

$X$  — источник сообщений;

$A_1$  — первичный алфавит (до преобразования);

$A_2$  — вторичный (алфавит конечного представления).

Обычно  $A_1$  — байты, исходные тексты  $x$  — бинарные файлы.



# Характеристики кодов

- ❶ Первичный алфавит  $A_1$
  - ❷ Оптимальность (неизбыточность)
  - ❸ Избыточность (в том числе помехоустойчивость)
- } модель источника!
- ❹ Вторичный алфавит  $A_2$  ( $A_2 = \{0, 1\}$  — двоичный код)
  - ❺ Однозначная декодируемость [должна быть!]
  - ❻ Разделяемость — код  $code(x)$  любой последовательности  $x = \overline{a_1 \dots a_n}$  единственным образом разделим на кодовые слова  $c_i = code(a_i)$ ,  $a_i \in A_1$ :
    - ❶ коды фиксированной ширины —  $a, b, c \rightarrow 00, 01, 10$ ;
    - ❷ коды с разделителем —  $1, 11, 111$  (0 как разделитель символов);
    - ❸ префиксные коды (дерево) —  $0, 10, 11$ ;
    - ❹ прочие — например,  $11, 1110111, 11100111$ .



# Оптимальное кодирование источника $X$

Пусть  $X$  порождает последовательность из  $2^N$  возможных символов.

- 1 Равновероятный источник ( $I(X) = N$ ) — кодирование отдельных символов кодами фиксированной ширины  $N$  бит.
- 2 Стационарный источник без памяти, порождающий символы с разными постоянными вероятностями ( $I(X) < N$ ) — кодирование отдельных символов кодами переменной ширины: коды Хаффмана, методы семейства арифметического кодирования.
- 3 Стационарный источник с памятью, порождающий символы с вероятностями, зависящими от контекста ( $I(X) < N$ ) — кодирование сочетаний символов: словарные методы семейства LZ77 (словарь=текст) и семейства LZ78 (отдельный словарь в виде дерева/таблицы).

Если изначально каждый символ записан кодом фиксированной ширины из  $N$  бит  $\Rightarrow$  сжатие для 2 и 3.



# Метод кодирования и его реализация

**Идея кодирования:**  $x = A_1^+, x \in X \leftrightarrow \text{code}(x) \in A_2^+$

На практике: первичный алфавит — байты, исходный текст — произвольной длины  $n$  байт; причём там может встречаться любой символ или их комбинация.

**Алгоритм кодирования:**

- 1 собственно алгоритм;
- 2 представление данных.

**Программная реализация:**

- 1 дополнение исходного текста при необходимости (обычно нулями) и обрезка декодированного текста до длины  $n$ ;
- 2 при сжатии: анализ сжатия/увеличения (запись кода или копии);
- 3 формирование и чтение заголовка.

$$(n \text{ байт}) \leftrightarrow \left\{ \begin{array}{l} \text{модель } X, \\ x \in A_1^+ \end{array} \right\} \leftrightarrow \left\{ \begin{array}{l} \text{модель } X, \\ \text{алгоритм кодирования,} \\ \text{параметры кодирования,} \\ \text{code}(x) \in A_2^+ \end{array} \right\} \leftrightarrow \left\{ \begin{array}{l} \text{заголовок,} \\ \text{данные } y \text{ (} m \text{ байт)} \end{array} \right\}$$

# Формат файла

- ❶ Сигнатура (обычно первые 2-4 байта для общепринятых форматов)
  - быстрое распознавание типа файла (свой/чужой).
- ❷ Метаданные (заголовок)
  - версия формата;
  - исходная длина файла;
  - смещение начала данных, их размер и формат;
  - тип сжатия, параметры для распаковки (обычно чем нестандартнее модель источника, тем объёмнее);
  - тип защиты от помех, параметры для восстановления;
  - зарезервированные поля для выравнивания;
  - контрольная сумма заголовка;
  - контрольная сумма файла и т. д.
- ❸ Данные
  - могут включать вложенные заголовки (контейнеров) с сигнатурами.

# Некоторые типовые реализации

Алгоритм работает с блоком длины  $N$  байт (после кодирования  $M$  байт) — файл дополняется до  $kN$  и нарезается на блоки:

$$\text{Блоки по } N \rightarrow M \text{ байт: } (n \text{ байт}) \leftrightarrow \begin{cases} n, \\ k \text{ блоков по } N, \\ (k-1)N < n \leq kN \end{cases} \leftrightarrow \begin{cases} n, \\ \text{алгоритм,} \\ kM \text{ байт} \end{cases}$$

Алгоритм заменяет подстроку  $c_i \dots c_j$  на некоторый кортеж  $\alpha_i$  — предварительно кортеж  $\alpha_i$  **префиксом**  $p = c_k$  (выбираем самый редкий символ):  $c_i \dots c_j \rightarrow p \alpha_i$ , вхождения  $p = c_k$  как символа экранируем (заменяем на  $p \alpha_0$  такое, что никакое  $\alpha_i \neq \alpha_0$  и не начинается с  $\alpha_0$ ):

$$\begin{cases} \dots(c_i \dots c_j) \dots c_k \dots \\ c_i - \text{байты} \end{cases} \leftrightarrow \dots(\alpha_i) \dots c_k \dots \leftrightarrow \begin{cases} \text{алгоритм,} \\ \text{значение префикса } p = c_k, \\ \dots p \alpha_i \dots p \alpha_0 \dots \end{cases}$$

где  $c_i, p$  — символы (байты),  $\alpha_i, \alpha_0$  — цепочки символов (байтов).

# Простые коды

$x \in A_1 \leftrightarrow y \in A_2$  без сжатия, защиты от помех и шифрования

Простейший базовый код (подразумевается):

- 0 байт памяти  $\leftrightarrow$  беззнаковое целое число  $0 \dots B - 1$  (обычно: октет  $\leftrightarrow 0 \dots 255$ )  
натуральный двоичный код  $\implies$  биты байта имеют номер.

Порядок байтов (если файл читается и записывается на одной платформе — не важен и также подразумевается):

- 1  $N$  байтов  $(\chi_0, \dots, \chi_{N-1})$ ,  $N = 2^s \leftrightarrow$  беззнаковое целое число  $0 \dots B^N - 1$ ;
- 2  $N$  битов,  $N$  произвольное  $\leftrightarrow$  беззнаковое целое число  $0 \dots 2^N - 1$ .

Простые коды (фиксированной ширины): беззнаковое целое (код)  $\leftrightarrow$  ?

- 3  $0 \dots 127 \leftrightarrow$  символ из таблицы ASCII;
- 4 знаковые числа;
- 5 числа с плавающей или фиксированной запятой;
- 6 нестандартные цифровые коды (ДДК, Грея, Джонсона) и т. д.

# Натуральный двоичный код

Целые неотрицательные числа: от 0 до  $2^N - 1$ .

Для  $N = 4$  — целые 0 до  $2^4 - 1 = 16 - 1 = 15$ :

0	1	2	3	4	5	6	7
0000	0001	0010	0011	0100	0101	0110	0111
8	9	A (10)	B (11)	C (12)	D (13)	E (14)	F (15)
1000	1001	1010	1011	1100	1101	1110	1111

Циклическая арифметика по модулю  $2^N$ :

$$2^N \equiv 0$$

то есть  $(2^N - 1) + 1 = 0$ .

Взвешенный:

$$x = 1 \cdot \text{bit}[0] + \dots + 2^{N-1} \cdot \text{bit}[N-1] = \alpha_0 \cdot \text{bit}[0] + \dots + \alpha_{N-1} \cdot \text{bit}[N-1].$$



# Код со смещением

Целые числа (возможно — знаковые) в произвольном диапазоне  $[a, b]$

— для  $x \in [a, b]$  записываем беззнаковое число  $y = x - a$  натуральным двоичным кодом.

# Дополнительный код

Целые знаковые числа, 0 и ближайшие к 0 положительные представляются как беззнаковые, **циклическая арифметика по модулю  $2^N$** :  $(-1) = 0 - 1 \equiv 2^N - 1$ ,  $(-2) \equiv 2^N - 2, \dots$   
 $(-2^{N-1}) \equiv 2^N - 2^{N-1} = 2^{N-1}$  (считается отрицательным).

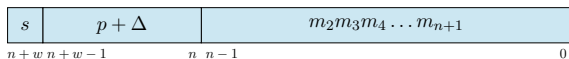
Целые числа от  $-2^{N-1}$  до  $+2^{N-1} - 1$ :

0	+1	+2	+3	+4	+5	+6	+7	
0000	0001	0010	0011	0100	0101	0110	0111	
	-1	-2	-3	-4	-5	-6	-7	-8
	1111	1110	1101	1100	1011	1010	1001	1000

$$(-x) = 0 - x = (-1 - x) + 1 = (\sim x) + 1;$$

$$max + 1 = min.$$

## Числа с плавающей запятой (IEEE 754)



$$(-1)^s \cdot 2^p \cdot \overline{0, 1m_2m_3m_4 \dots m_{n+1}}$$

$$p_{min} \leq p \leq p_{max}$$



$$(-1)^s \cdot 2^{p_{min}} \cdot \overline{0,0m_2m_3m_4 \dots m_{n+1}}$$



$$(-1)^s \cdot 0$$



$$(-1)^s \cdot \infty$$



нечисло (*nan*)

Нециклическая неассоциативная арифметика:  $x + (y + z) \neq (x + y) + z$

32 = 1 + 8 + 23 бита — одинарная точность, *float*

$$2^{-126} \leq |x| \leq 2^{127} \cdot (2 - 2^{-23})$$

64 = 1 + 11 + 52 бита — двойная, *double*

$$2^{-1022} \leq |x| \leq 2^{1023} \cdot (2 - 2^{-52})$$

Округление: к ближайшему|чётному, к ближайшему| $\infty$ , к 0, к  $+\infty$  (вверх), к  $-\infty$  (вниз)

C/C++:  $\text{float} \subseteq \text{double} \subseteq \text{long double} \subseteq \mathbb{R}$ , но даже  $\text{long double} \neq \mathbb{R}$

# Едини́чный код

Избыточный невзвешенный  
 рефлексный (при переходе между кодовыми комбинациями  
 изменяется только один бит)  
 нециклический ( $max + 1 \neq min$ ) двоичный код

Для  $N$  битов — целые 0 до  $N$ :

0	1	2	3	4
0000	0001	0011	0111	1111
	0010	0101	1011	
	0100	1001	1101	
	1000	0110	1110	
		1010		
		1100		

# Коды Грея и Джонсона

Код Грея — избыточный невзвешенный рефлексный циклический двоичный код

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

Код Джонсона — избыточный невзвешенный рефлексный циклический двоичный код

0	1	2	3	4	5	6	7
0000	0001	0011	0111	1111	1110	1100	1000

# ASCII и Unicode

ASCII — 128 символов и семибитная (~однобайтовая) кодировка

**ASCII Code Chart**

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT	LF	VT	FF	CR	SO	SI
1	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US
2		!	"	#	\$	%	&	'	(	)	*	+	,	-	.	/
3	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
6	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7	p	q	r	s	t	u	v	w	x	y	z	{		}	~	DEL

Unicode — 137 994 символ (в версии 12.1) и набор кодировок: UTF-8, UTF-16 (UTF-16BE, UTF-16LE) и UTF-32 (UTF-32BE, UTF-32LE)

UTF-8 (до 6 байт)	1 байт 0aaa aaaa	2 байта 110x xxxx 10xx xxxx
3 байта	1110 xxxx 10xx xxxx 10xx xxxx	
4 байта	1111 0xxx 10xx xxxx 10xx xxxx 10xx xxxx	
5 байт	1111 10xx 10xx xxxx 10xx xxxx 10xx xxxx 10xx xxxx	
6 байт	1111 110x 10xx xxxx 10xx xxxx 10xx xxxx 10xx xxxx 10xx xxxx	



# Код Морзе (Фридрих Герке)



Азбука Морзе				
А • —	К — • —	Ф • • — •	1 • — — — —	. • • • • •
Б — • • •	Л • — • •	Х • • • •	2 • • — — —	, • — — — —
В • — —	М — — —	Ц — • — •	3 • • • — —	; — • — — •
Г — — •	Н — •	Ч — — — •	4 • • • • —	: — — — • •
Д — • •	О — — — —	Ш — — — —	5 • • • • •	? • • — — • •
Е •	П • — — •	Щ — — — •	6 — • • • •	! — — • • — —
Ж • • • —	Р • — •	Ъ, Ъ — • • —	7 — — — • •	- — — • • • —
З — — • •	С • • •	Ы — • — —	8 — — — — • •	« • — — — •
И • •	Т —	Э • • — • •	9 — — — — •	( — • — — • —
Й • — — —	У • • —	Ю • • — — —	0 — — — — —	/ — • • — •
		Я • — • —		

Каждой букве или знаку соответствует определённая комбинация кратковременных (точка) и вдвое более длинных (тире) импульсов тока, разделённых бестоковым интервалом, равным длительности точки.

Для разделения букв в словах и цифр в многозначных числах применяется тройной бестоковый интервал, заканчивающий каждую комбинацию.

Для разделения слов в тексте служит пятикратный бестоковый интервал.

$$A_2 = \{ \cdot, -, \text{межсимвольный интервал}, \text{межсловный интервал} \}$$



# Код Бодо (Дональд Мюррей)

Международный телеграфный код №2 (ITA2) +  
+ 00000 = МТК-2

Русский шрифт	Е	Перевод строки ≡	Пробел ∨	Т	А	И	Н	О	С	Р	Х	Д	Л	З	У	Ц	М	Ф	Й	Г	П	Ы	Б	В		К	Ж	Ь	Я			
Цифры	3			5	-	8	,	9	'	Ч	Щ	кто там?	)	+	7	:	.	Э	Ю (ив)	Ш	0	5	?	2		Цифры	(	=	/	1		
Латинский шрифт	Е			Т	А	І	Н	О	Ѕ	Р	Н	Д	Л	Ζ	U	С	М	F	J	G	P	Y	B	W		К	V	X	Q			
Ведущие отверстия	1	●			●				●			●		●	●			●	●			●	●	●	●	●		●	●	●	●	
	2	●			●	●				●			●		●	●			●	●	●			●	●	●	●	●	●	●	●	
		○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
	3		●		●	●			●		●			●	●	●	●			●	●				●	●	●	●	●	●	●	
	4			●			●	●		●		●			●	●	●	●	●				●		●	●	●	●	●		●	
5				●				●			●		●	●			●			●	●	●	●	●	●	●	●	●	●	●	●	

фиксированной ширины 5, режимы



# Спасибо за внимание!

МИЭТ

[www.miet.ru](http://www.miet.ru)

Александра Игоревна Кононова / [illinc@mail.ru](mailto:illinc@mail.ru)

+7-985-148-32-64 (телефон), +7-977-977-97-29 (WhatsApp),

[gitlab.com/illinc/raspisanie](https://gitlab.com/illinc/raspisanie)

<https://gitlab.com/illinc/otik/>