

# Измерение информации. Алфавитный подход. Вес символа. Кодировки

Логарифмическая мера информации, алфавитный подход, информационный вес символа, сравнение кодировок и практическая работа в текстовом процессоре.

## Цели урока

- Понимать алфавитный подход:  $m = \lceil \log_2 N \rceil, i = \log_2 N$ .
- Вычислять объём сообщений и информационный вес символа.
- Различать кодировки ASCII/Windows-1251/UTF-8 и понимать их отличия.
- На практике определять код символа в разных кодировках.

## Сценарий видео (7–11 минут)

- 0:00–1:30 Алфавитный подход и логарифмическая мера:  $i = \log_2 N$ ;  $m = \lceil \log_2 N \rceil$ ; связь с битами.
- 1:30–3:30 Объём сообщения:  $I = k \times i$  (бит) и практическое  $I \approx k \times m$ .
- 3:30–5:30 Кодировки: ASCII (7/8 бит), Windows-1251 (однобайтная кириллица), UTF-8 (переменная длина, совместима с ASCII).
- 5:30–7:30 Как один и тот же текст «весит» по-разному в разных кодировках; байты и отображение.

## Быстрые примеры

### Пример 1: Алфавит

$N=70 \rightarrow m=7$

Текст 1200 символов:  $I=1200 \times 7=8400$  бит = 1050 В.

### Пример 2: Кодировки

Слово «Привет»

ASCII: недоступно для кириллицы; Windows-1251: 6 байт; UTF-8: 12 байт (по 2 байта на букву кириллицы).

### Пример 3: Вес символа

Равновероятный алфавит N

$i = \log_2 N$  бит/символ; для  $N=32 \rightarrow 5$  бит.

## Закрепление: задачи с подробными решениями

### 1. Объём текста по алфавиту

*Условие:* алфавит 50 символов, текст 500 символов. Найти объём в байтах.

*Решение:*  $m = \lceil \log_2 50 \rceil = 6$ ;  $I = 500 \times 6 = 3000$  бит = 375 В.

Ответ: 375 В.

## 2. Вес символа (Хартли)

Условие: алфавит 32 символа, символы равновероятны. Найти  $i$ .

Решение:  $i = \log_2 32 = 5$  бит/символ.

## 3. Сравнение кодировок

Условие: строка «Test» и «Тест». Сравнить объём в Windows-1251 и UTF-8.

Решение: «Test»: 4 В в обеих; «Тест»: Windows-1251 — 4 В; UTF-8 — 8 В (каждая кириллическая буква — 2 байта).

## 4. Идентификатор символа

Условие: определить код точки «.» в ASCII, и код «Я» в Windows-1251 и UTF-8.

Решение: «.»: 46 (0x2E, ASCII). «Я»: 0xDF (223) в Windows-1251; UTF-8: D0 AF.

## Практическая работа: определение кода символа в разных кодировках

1. Откройте текстовый процессор (например, LibreOffice Writer) или редактор кода.
2. Вставьте фразу: «Привет, мир! Test.»
3. Сохраните файл как *Windows-1251*. Затем как *UTF-8*.
4. Откройте каждый файл в редакторе в шестнадцатеричном режиме (или командой `xxd` в терминале).
5. Сравните байты для символов «П», «я», «Т», «.» в двух вариантах.
6. Зафиксируйте:
  - Windows-1251: «П»=0xCF, «я»=0xFF; «Т»=0x54; «.»=0x2E.
  - UTF-8: «П»=D0 9F, «я»=D1 8F; «Т»=54; «.»=2E.

В UTF-8 ASCII-символы кодируются одним байтом и совпадают по значению с ASCII. Кириллица занимает 2 байта.

## Мини-викторина

- Что такое  $m$ ? → Минимальное целое число бит на символ.
- В чём разница Windows-1251 и UTF-8? → Однобайтная локальная vs универсальная переменной длины.
- Код «А» в ASCII? → 65 (0x41).
- Почему «Тест» больше в UTF-8? → Кириллица кодируется двумя байтами.

## Конспект (коротко)

- **Алфавитный подход:**  $m = \lceil \log_2 N \rceil$ ,  $i = \log_2 N$ .
- **Объём:**  $I = k \times i$  (бит), на практике  $I \approx k \times m$ .

- **Кодировки:** ASCII (базовый латиница), Windows-1251 (кириллица 1 байт), UTF-8 (универсальная, переменная длина).
- **Сопоставление байтов:** ASCII символы совпадают в UTF-8, кириллица — 2 байта.

## Домашнее задание (самопроверка)

1. **Задача А:** алфавит 70 символов, текст 1000 символов. Найти объём в байтах.

*Ответ:*  $m=7$ ;  $I=7000$  бит; 875 В.

2. **Задача В:** Сравнить объём строки «Hello, мир!» в Windows-1251 и UTF-8.

*Подсказка:* латиница — 1 байт в обоих; кириллица — 1 байт (1251) и 2 байта (UTF-8).

3. **Задача С:** Найти  $i$  для алфавита 128 символов.

*Ответ:*  $i = 7$  бит.

Подготовлено для урока «Алфавитный подход и кодировки» · Печать: Ctrl/Cmd + P