

Средняя общеобразовательная школа № 654 имени А.Д. Фридмана

Кодирование текстовой информации

Информатика

10 класс

Базовая подготовка

Учитель: *Ермаков Максим Геннадьевич*

Зачем нужно

Целью процесса кодирования информации является *определение количества информации* или *вычисление информационного объёма*, необходимого для сохранения информации для последующей обработки

Задача определения количества информации

Используются два подхода:

- *вероятностный* (содержательный);
количество информации зависит от содержания сообщения
(по объему информации отметка за зачет отличается от отметки за экзамен)
- *алфавитный*
количество информации не зависит от содержания сообщения
(сообщение из газеты может быть без информации)

Вероятностный подход

Сообщение несет в себе информацию о событии.

Бит – это количество информации, соответствующее выбору одного из двух равновероятных вариантов.

Количество информации зависит от *количества возможных вариантов события*.

Формула Шеннона: $i = \log_2 N$ (или $N = 2^i$)

i – информационный объем сообщения

N – количество равновероятных вариантов события

Алфавитный подход

Сообщение представляет собой последовательность символов алфавита

Алфавит – набор неповторяющихся символов для кодирования информации

Количество информации зависит от *количества символов* и *информационного веса символа*

$$V = K \cdot i \quad W = 2^i$$

V – информационный объем сообщения

K – количество символов сообщения

i – информационный вес одного символа алфавита

W – количество символов алфавита

Примеры задач

Определить объём каждого сообщения:

- **ЁЖИК В ТУМАНЕ**
- **Ёжик в тумане**

Формулы:

$$V = K \cdot i \quad W = 2^i$$

V – информационный
объем сообщения

K – количество символов
сообщения

i – информационный вес
одного символа
алфавита

W – количество символов
алфавита

Примеры задач

Для кодирования некоторой последовательности, состоящей из букв **К**, **Л**, **М**, **Н**, решили использовать неравномерный двоичный код, удовлетворяющий условию Фано. Для буквы **Н** использовали кодовое слово **0**, для буквы **К** — кодовое слово **10**. Какова наименьшая возможная суммарная длина всех четырёх кодовых слов?

Примечание.

Условие Фано означает, что никакое кодовое слово не является началом другого кодового слова. Это обеспечивает возможность однозначной расшифровки закодированных сообщений.

1) 7 2) 8 3) 9 4) 10

Задачи для решения

- Для кодирования некоторой последовательности, состоящей из букв А, Б, В, Г и Д, используется неравномерный двоичный код, позволяющий однозначно декодировать полученную двоичную последовательность. Вот этот код: А — 1; Б — 0100; В — 000; Г — 011; Д — 0101. Требуется сократить для одной из букв длину кодового слова так, чтобы код по-прежнему можно было декодировать однозначно. Коды остальных букв меняться не должны.
Каким из указанных способов это можно сделать?
1) для буквы Г — 11 2) для буквы В — 00 3) для буквы Г — 01 4) это невозможно
- **Определить количество символов алфавита:**
 - сообщение объёмом 200 бит содержит 40 символов
 - сообщения, содержащего 4096 символов, объёмом 1/512 части Мбайта
- **Определить количество символов сообщения объёмом 300 бит:**
 - написанного при помощи только строчных латинских букв
 - написанного при помощи только строчных русских букв
- **Перекодировка сообщения**
 - Автоматическое устройство осуществило перекодировку информационного сообщения на русском языке, первоначально записанного в 16-битном коде Unicode, в 8-битную кодировку КОИ-8. При этом информационное сообщение уменьшилось на 16 байтов. Сколько бит было в первоначальном сообщении?
 - Автоматическое устройство осуществило перекодировку информационного сообщения на русском языке, первоначально записанного в 16-битном коде Unicode, в 8-битную кодировку КОИ-8. При этом информационное сообщение уменьшилось на 480 бит. Какова длина сообщения в символах?

Домашнее задание

- Записать в тетрадь заголовок «Кодирование текстовой информации»
- Переписать в тетрадь виды (подходы) кодирования информации и характеристику каждого из них.
- Выписать определения понятий «бит» и «алфавит», формулы вычисления информационного объема сообщения при каждом виде кодирования информации
- Решить в тетради выделенные цветом задачи слайдов «Задачи для решения»