

Кодирование информации

- § 4. Дискретное кодирование
- § 5. Равномерное и неравномерное кодирование
- § 6. Декодирование
- § 7. Алфавитный подход к измерению количества информации

Кодирование информации

§ 4. Дискретное кодирование

Вспомним известное

Кодирование — это представление информации в форме, удобной для её хранения, передачи и автоматической обработки.

Код — это правило, по которому сообщение преобразуется в цепочку знаков.

Язык — это система знаков и правил, используемая для записи и передачи информации.

Формальный язык — это язык, в котором однозначно определяется значение каждого слова, а также правила построения предложений и придания им смысла.

Знаковые системы

Знак — это «заменитель» объекта, вызывает в сознании объект.



– пиктограмма

Символ — это знак, о значении которого люди договорились.

§ – параграф ₹ – рубль

Знаковая система определяется алфавитом (набором используемых знаков) и правилами выполнения операций с этими знаками.

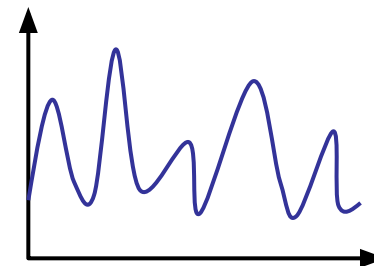


Знаковая система в компьютерах?

010101

Аналоговые сигналы и устройства

Аналоговый сигнал — это сигнал, который в любой момент времени может принимать любые значения в заданном диапазоне.

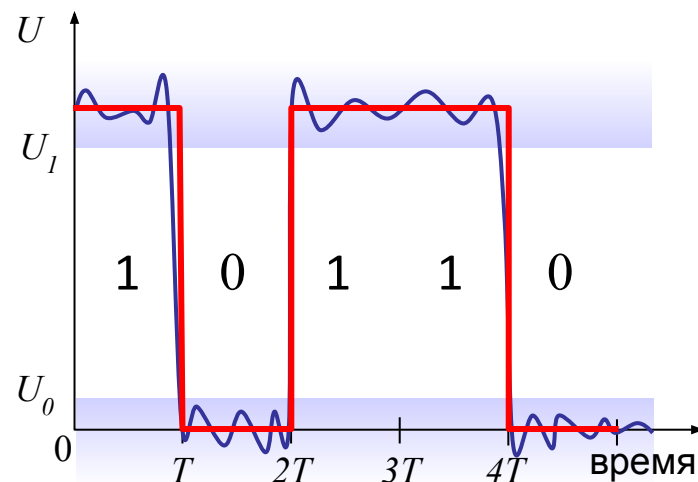


Аналоговые компьютеры



- невозможно «очистить» сигнал от помех
- при измерении сигнала вносится ошибка
- при копировании аналоговая информация искажается

Дискретные (цифровые) сигналы



Свойства:

- сигнал изменяется только в отдельные моменты времени (*дискретность по времени*);
- принимают только несколько возможных значений (*дискретность по уровню*).

Дискретный сигнал — это последовательность значений, каждое из которых принадлежит некоторому конечному множеству.

Дискретность

Цель – максимально точно передавать сообщения при сильных помехах.



Pacta sunt servanda.

• — — • — • • • • — • —
01000011001



Компьютеры могут хранить и обрабатывать

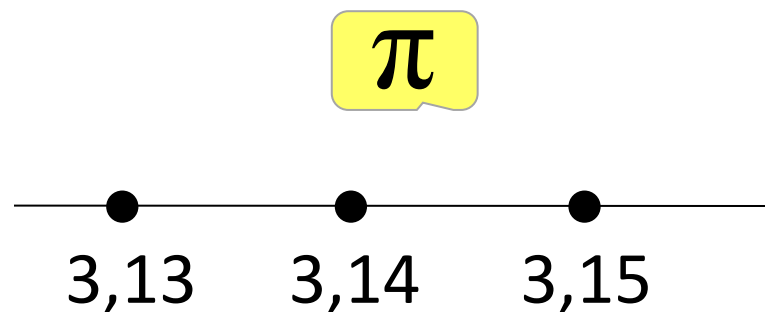
... только дискретную информацию нечеловеческого количества знаков некоторого алфавита.



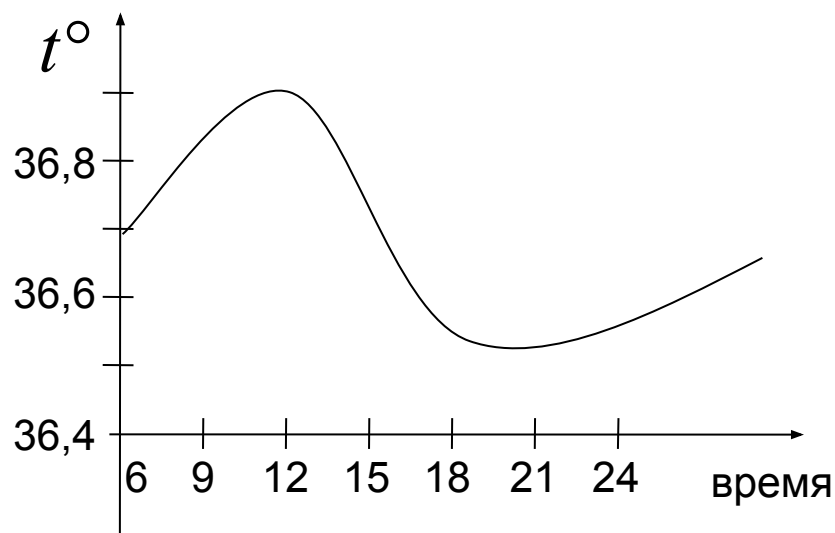
Все виды информации нужно перевести в дискретный вид!

Дискретизация

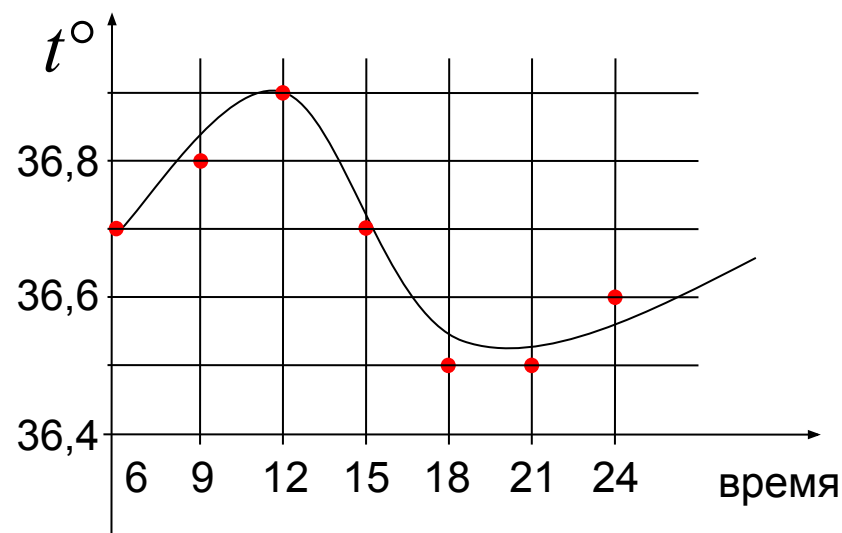
Дискретизация — это представление единого объекта в виде множества отдельных элементов.

 π 

Дискретизация



аналоговая информация



дискретизация

6 ч. 36,7°

9 ч. 36,8°

12 ч.

36,9°

15 ч.

36,7°

18 ч.

36,5°

21 ч.

36,5°

24 ч.

дискретная информация

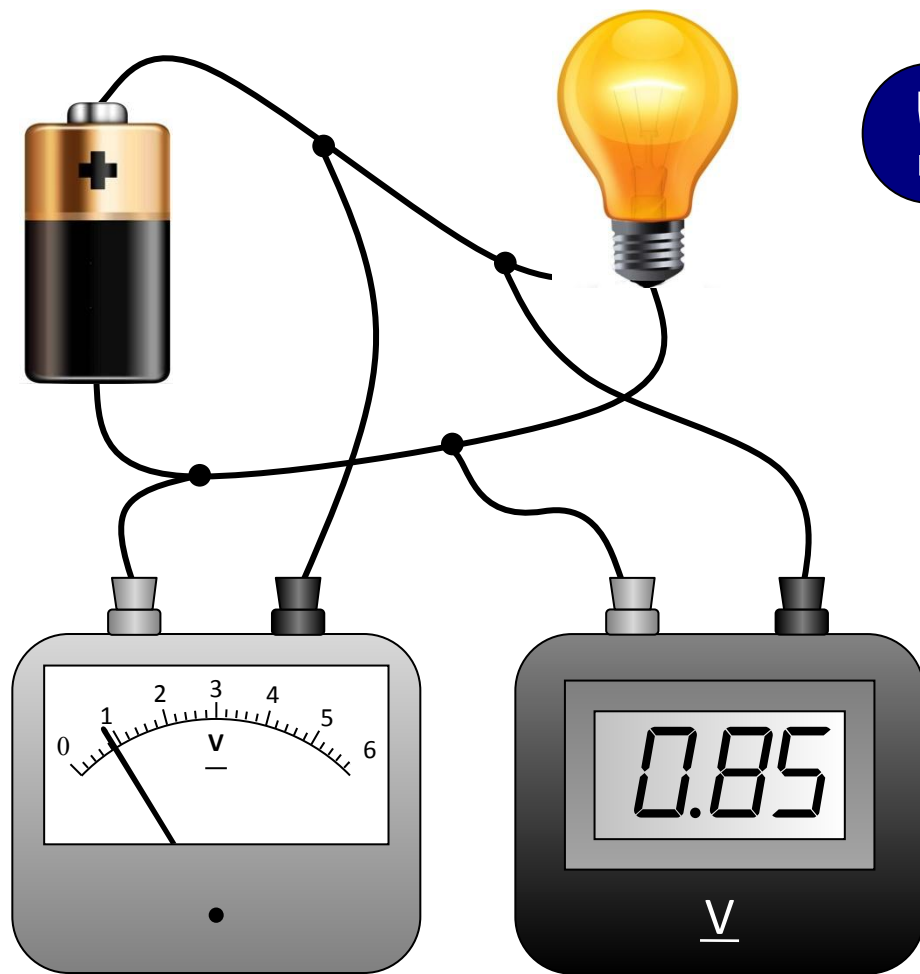


При дискретизации
есть потеря информации!



Как уменьшить потери?

Непрерывность и дискретность



аналоговые
данные

дискретные
данные

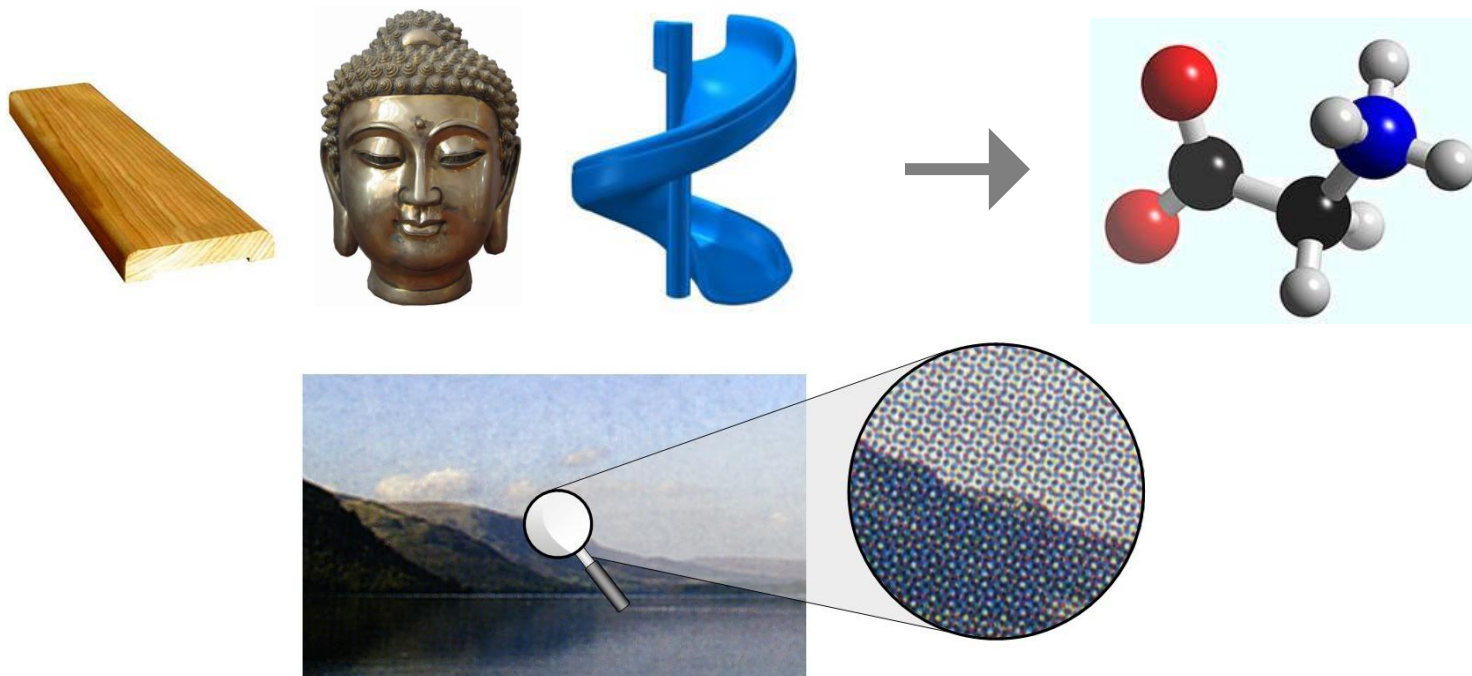


Дискретность —
это свойство не
информации, а её
представления.

Непрерывность и дискретность

! При увеличении точности дискретизации свойства аналоговой и дискретной информации практически совпадают!

$$\pi \approx 3,1415926$$



Кодирование информации

§ 5. Равномерное и неравномерное кодирование

Вспомним известное

Алфавит — это набор знаков, который используется в языке.

Мощность алфавита — это количество знаков в алфавите.

Равномерный код — это код, в котором все кодовые слова имеют одинаковую длину.

Неравномерный код — это код, в котором кодовые слова имеют различную длину.

Двоичное кодирование — это кодирование с помощью двух знаков.

1 бит — это одна двоичная цифра (один знак сообщения, записанного в двоичном коде).

Количество возможных сообщений

Если алфавит языка состоит из M символов (имеет мощность M), количество различных сообщений длиной L знаков равно

$$N = M^L$$

Для двоичного кода: $N = 2^L$

Сколько

- возможных 7-битовых двоичных кодов?
- возможных 5-буквенных слов в русском языке?
- возможных 3-буквенных слов в английском языке?

$$2^7$$

$$33^5$$

$$26^3$$

Количество возможных сообщений

Сколько

• различных чисел можно закодировать в 8-битовой ячейке?

2⁸

• различных чисел можно закодировать в 8-разрядной ячейке троичного компьютера (-1, 0, 1)?

3⁸

• сколько битов нужно выделить для хранения номера спортсмена от 1 до 1000?

10

$$512 = 2^9 < 1000 \leq 2^{10} = 1024$$

• сколько битов нужно выделить для хранения температуры от -50° до 80°?

8

$$128 = 2^7 < 131 \leq 2^8 = 256$$

Правило умножения

Если в сообщении длиной L на позиции i может стоять один из M_i символов, количество различных сообщений *равно*

$$N = M_1 \cdot M_2 \cdot \dots \cdot$$

$$M_L$$

Задача 1. Сколько существует различных сообщений длины 5 в алфавите {А, В, С, Х}, если буква «Х» может появляться только на первом или на последнем месте?

4	3	3	3	4
M_1	M_2	M_3	M_4	M_5

$$4 \cdot 3 \cdot 3 \cdot 3 \cdot 4 = 432$$

Правило умножения

Задача 2. Сколько существует 5-значных десятичных чисел, все цифры в которых различны?

9	9	8	7	6
M_1	M_2	M_3	M_4	M_5

$$9 \cdot 9 \cdot 8 \cdot 7 \cdot 6 = 27216$$

Не может быть 0!

Неравномерные коды



- можно уменьшить длину закодированного сообщения

Равномерный код:

А	Г	Р
00	01	10

ГАГАР

А

12 бит
→ 010001001000

Неравномерный код:

А	Г	Р
0	01	10

ГАГАР

А

9 бит
→ 010010100



- не всегда однозначно декодируется

010010100 → 010010100 ГАГАР

→ 010010100 АРАРА

А

Правило сложения

Задача 3. Сколько существует двоичных кодов длиной от 2 до 5 битов?

$$\begin{array}{ll} L = 2: & N_2 = 2^2 = 4 \\ L = 3: & N_3 = 2^3 = 8 \\ L = 4: & N_4 = 2^4 = 16 \\ L = 5: & N_5 = 2^5 = 32 \end{array}$$

$$N = 4 + 8 + 16 + 32 = 60$$

$$N = N_2 + N_3 + N_4 + N_5$$



Правило сложения!

Правила умножения и сложения

Задача 4. Сколько существует различных 3-буквенных слов в алфавите {К, Р, О, Т}, в которых буква К встречается ровно 1 раз?

К	*	*		
1	3	3	$1 \cdot 3 \cdot 3 = 9$	
*	К	*	$3 \cdot 1 \cdot 3 = 9$	} $9 + 9 + 9 = 27$
*	*	К	$3 \cdot 3 \cdot 1 = 9$	

Задачи

1. Сколько существует в коде Морзе различных последовательностей из точек и тире, длина которых от 4 до 6 символов?
2. Вася и Петя передают друг другу сообщения, используя синий, красный и зелёный фонарики. Это они делают, включая по одному фонарику на одинаковое короткое время в некоторой последовательности. Количество вспышек в одном сообщении — 3 или 4, между сообщениями — паузы. Сколько различных сообщений могут передавать мальчики?

Задачи

3. Шахматная доска состоит из 8 столбцов и 8 строк. Какое минимальное количество битов потребуется для кодирования координат одной шахматной фигуры?
4. Для кодирования значений температуры воздуха (целое число в интервале от -50 до 40) используется двоичный код. Какова минимальная длина двоичного кода?
5. Дорожный светофор подаёт шесть видов сигналов (непрерывные красный, жёлтый и зелёный, мигающие жёлтый и зелёный, мигающие красный и жёлтый одновременно). Подряд записано 100 сигналов светофора. Определите информационный объём этого сообщения в битах.

Задачи

6. Автомобильный номер длиной 6 символов составляется из заглавных букв (всего используется 12 букв) и десятичных цифр в любом порядке. Каждый символ кодируется одинаковым и минимально возможным количеством битов, а каждый номер — одинаковым и минимально возможным количеством байтов. Определите объём памяти, необходимый для хранения 32 автомобильных номеров.

Кодирование информации

§ 6. Декодирование

Декодирование

Декодирование — это восстановление сообщения из последовательности кодов.

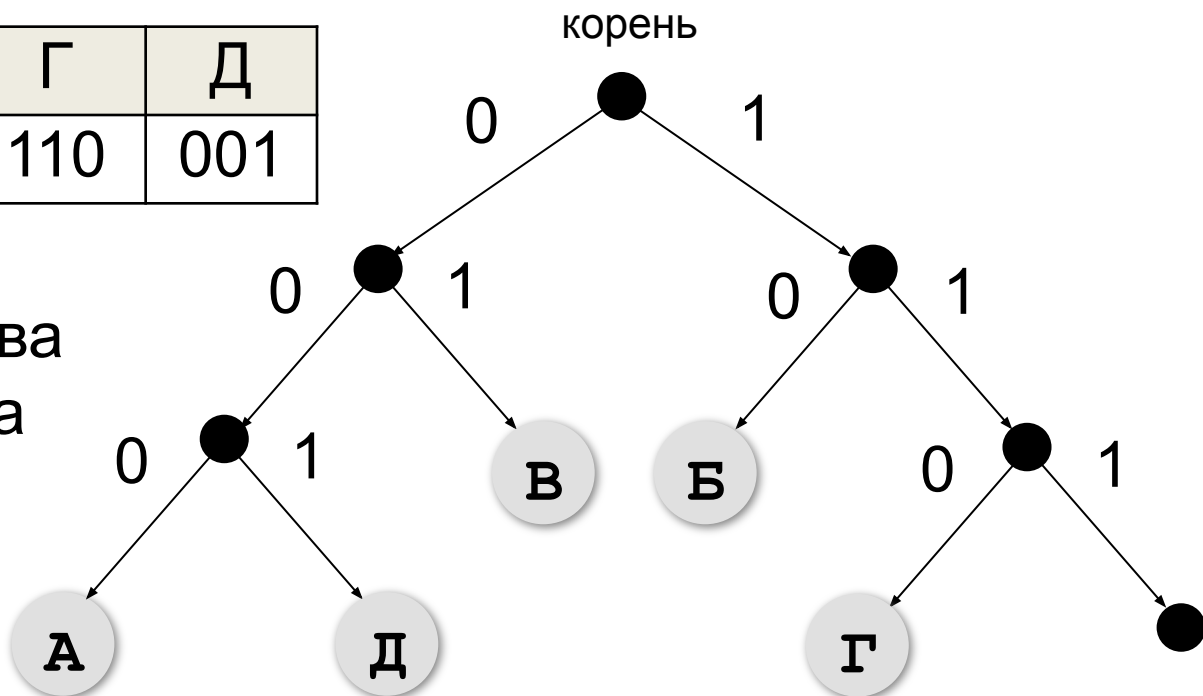
• - - • - • • • • - • - **ВАСЯ**



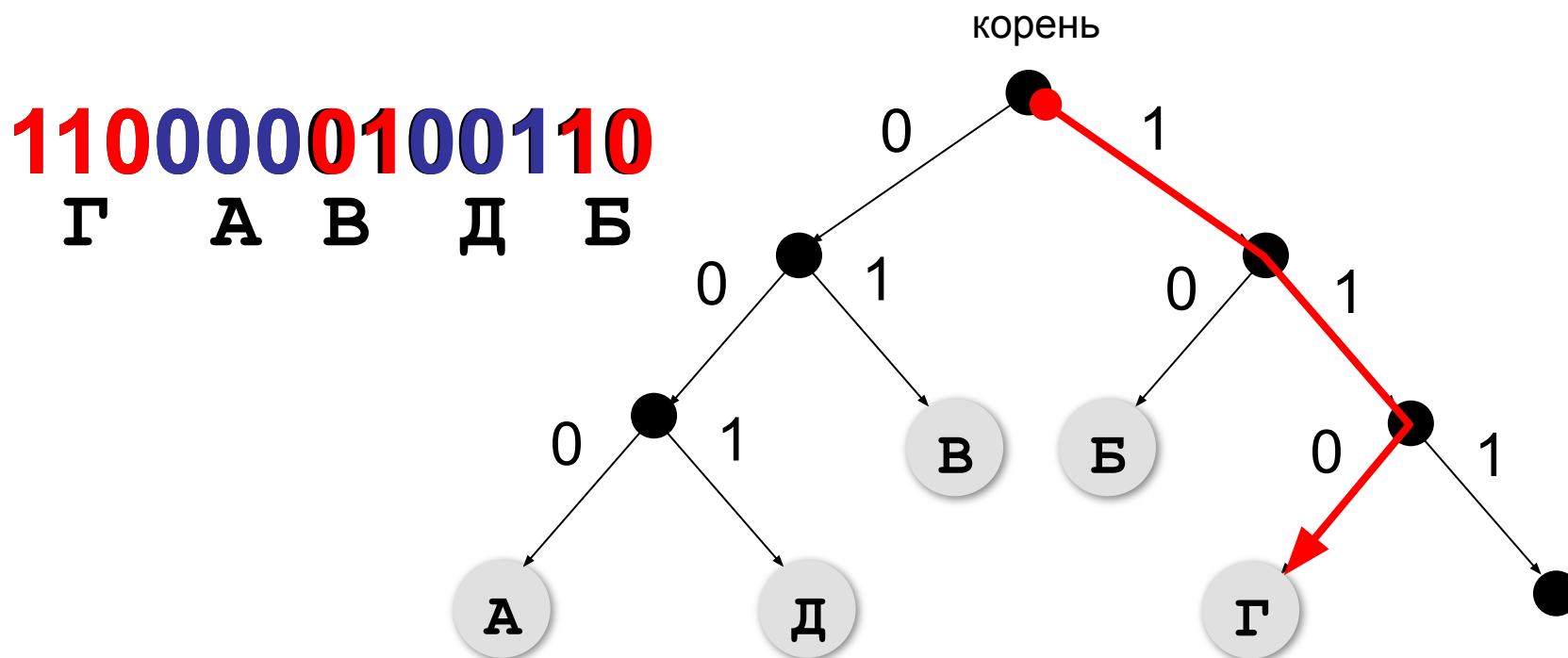
Когда разделитель не нужен?

А	Б	В	Г	Д
000	10	01	110	001

Все кодовые слова заканчиваются на листьях дерева!



Декодирование



Префиксный код — это код, в котором ни одно кодовое слово не совпадает с началом другого кодового слова (*условие Фано*). Сообщения декодируются однозначно.

Задачи

1. Для передачи сообщения, состоящего только из букв А, Б, В, Г, решили использовать неравномерный код:
- $$A = 0, B = 10, V = 110.$$

Как нужно закодировать букву Г, чтобы длина кода была минимальной и допускалось однозначное декодирование?

2. Для передачи сообщения, состоящего только из букв А, Б, В, Г, решили использовать неравномерный код:
- $$A = 0, B = 100, V = 101.$$

Как нужно закодировать букву Г, чтобы длина кода была минимальной и допускалось однозначное декодирование?

Постфиксные коды

Постфиксный код — это код, в котором ни одно кодовое слово не совпадает с **окончанием** другого кодового слова. Сообщения декодируются однозначно (**с конца!**).

А	Б	В	Г	Д
000	01	10	011	100

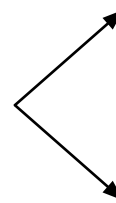
011000110110
Б Д **1**Г Б В

Неоднозначное декодирование

А	Б	В	Г	Д
01	010	011	11	101

? Выполняются ли условия Фано?

Декодирование *может быть* неоднозначным...

010100111101  **АБАГД**
АВВГ

! Может быть, что условия Фано не выполнены, а декодирование однозначно (см. учебник)!

Задача

*Докажите, что все сообщения, закодированные этим кодом, декодируются однозначно.

А	Б	В
0	11	010

01000011001011110000100

Кодирование информации

§ 7. Алфавитный подход к измерению количества информации

Алфавитный подход

Количество информации в битах определяется длиной сообщения в двоичном коде.

10101100

8 битов

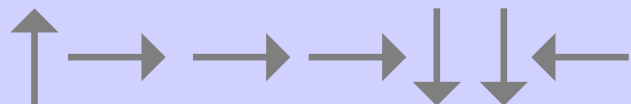
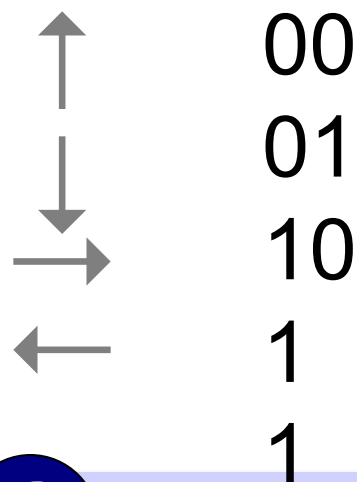


вперёд

назад

вправо

влево



00101010010111



Сколько битов?

14 битов

Другие единицы измерения

1 байт (*byte*) = 8 бит

1 Кбайт (килобайт) = 1024 байта

1 Мбайт (мегабайт) = 1024 Кбайт

1 Гбайт (гигабайт) = 1024 Мбайт

1 Тбайт (терабайт) = 1024 Гбайт

1 Пбайт (петабайт) = 1024 Тбайт



Алфавитный подход

- 1) определяем мощность алфавита M ;
- 2) определяем количество битов информации i , приходящихся на один символ, — информационную ёмкость (объём) символа:

M , символов	2	4	8	16	32	64	128	256	512	1024
i , битов информации	1	2	3	4	5	6	7	8	9	10

- 3) количество информации в сообщении:

$$I = L \cdot i$$

где L — количество символов в сообщении.

Алфавитный подход

- каждый символ несёт одинаковое количество информации
- частота появления разных символов (и сочетаний символов) не учитывается
- количество информации определяется только длиной сообщения и мощностью алфавита
- смысл сообщения не учитывается

Задача

Определить количество информации в 10 страницах текста (на каждой странице 32 строки по 64 символа) при использовании алфавита из 256 символов.

1) информационная ёмкость символа:

$$256 = 2^8 \Rightarrow i = 8 \text{ бит} = 1 \text{ байт}$$

2) количество символов на странице:

$$32 \cdot 64 = 2^5 \cdot 2^6 = 2^{11}$$

3) общее количество символов:

$$L = 10 \cdot 2^{11}$$

4) информационный объём сообщения:

$$I = L \cdot i = 10 \cdot 2^{11} \cdot 1 \text{ байтов} = 20 \text{ Кбайт}$$

Задача

Пароль длиной не более **11** символов (**цифры и 12 различных букв, как строчные, так и прописные**). Посимвольное равномерное кодирование, для хранения пароля отводится минимально возможное целое количество байт. Сколько байт нужно для **60** паролей?

1) мощность алфавита $M = 10 + 12 + 12 = 34$

2) информационная ёмкость символа:

$$2^5 < 34 \leq 2^6 \Rightarrow i = 6 \text{ бит}$$

2) на один пароль:

$$6 \cdot 11 = 66 \text{ бит} = 8, \dots \rightarrow 9 \text{ байт}$$

4) на 60 паролей:

$$I = 60 \cdot 9 \text{ байтов} = \mathbf{540 \text{ байт}}$$

округление
вверх

Конец фильма

ПОЛЯКОВ Константин Юрьевич

д.т.н., учитель информатики

ГБОУ СОШ № 163, г. Санкт-Петербург

kpolyakov@mail.ru

ЕРЕМИН Евгений Александрович

к.ф.-м.н., доцент кафедры мультимедийной

дидактики и ИТО ПГГПУ, г. Пермь

eremin@pspu.ac.ru

Источники иллюстраций

1. <http://overhealth.ru>
2. <https://ufhealth.org>
3. <http://wmposters.com>
4. <http://www.ulmart.ru>
5. <http://all-graphic.net>
6. <http://123rf.com>
7. <http://made-in-china.com>
8. <http://megamaster.biz>
9. <http://evrobass.ru>
10. <http://blendercontest.com>
11. <http://ru.wikipedia.org>
12. авторские материалы