1 Представление информации

Направления развития:

1. Сжатие
2. Избыточное кодирование
3. Криптографическое кодирование

Определение. **Алфавитом** Σ называется непустое конечное множество. Множество из n элементов Σ обозначается Σ^n.

$$\bigcup_{i=0}^{\infty} \Sigma^i = \Sigma^* \quad \Sigma^0 = \{\varepsilon\}$$

Определение. Конкатенация:

$$\alpha \in \Sigma^* \quad \beta \in \Sigma^* \mapsto \alpha\beta \in \Sigma^*$$

Конкатенация транзитивна $(\alpha\beta)\gamma = \alpha(\beta\gamma) \Rightarrow$ алфавит — полугруппа.

$\alpha\varepsilon = \varepsilon\alpha = \alpha \Rightarrow$ алфавит — моноид.

Т.к. алфавит — полугруппа и моноид, алфавит — свободный моноид.

Определение. **Гомоморфизм** $\varphi : \Sigma^* \to \Pi^*$

$\varphi(\alpha\beta) = \varphi(\alpha)\varphi(\beta)$

Пример.

$$0 \to a, 1 \to ab$$
$$\varphi : \{0,1\}^* \to \{a,b\}^*$$
$$\varphi(001) = aaab$$

φ — гомоморфизм $\Rightarrow \varphi(c_1, c_2 \dots c_n) = \varphi(c_1)\varphi(c_2)\dots\varphi(c_n)$

Определение. Отображение из произвольного Σ^* в Π^* называется **кодом**.

Если φ — гомоморфизм, φ — **разделяемый**.

Если $\Pi = \mathbb{B}$, φ — **бинарный/двоичный**.

Пример.

$$\Sigma = \{a,b,c\} \quad \varphi(a) = 0, \varphi(b) = 01, \varphi(c) = 1$$
$$\varphi(abc) = 0011 \quad \varphi(aacc) = 0011$$

Определение. Код называется **однозначно декодируемым**, если $\forall x, y \in \Sigma^* \quad \varphi(x) = \varphi(y) \Rightarrow x = y$

Определение. Кодом **постоянной длины** называется код, если $\varphi : \Sigma \to \Pi^k, k = const$

Лемма 1. φ — код постоянной длины

$\forall c \neq d \in \Sigma \quad \varphi(c) \neq \varphi(d)$

Тогда φ — однозначно декодируемый.

Теорема 1. $\Sigma, \Pi, |\Sigma = s|, |\Pi| = p, \Sigma \to \Pi^k$

$k = \lceil \log_p s \rceil$

$p^k < s$

Теорема 2. Крафта, Мак-Милана.

$\exists$ двоичных разделяемый однозначно декодируемый код переменной длины с длинами кодовых слов $l_1, l_2 \dots l_s \Leftrightarrow \sum_{i=1}^{s} 2^{-l_i} \leq 1, S \geq 2$

Доказательство. Докажем "$\Rightarrow$".

Пусть ab, abb, ab — все члены Σ

$$(ab + abb + bb)^2 = abab + ababb + abbb + \dots$$

$$(ab + abb + bb)^k - S^k \text{ слов, при этом все слова разные}$$

$$]a = \frac{1}{2}, b = \frac{1}{2}$$

$$ab + abb + bb = \sum 2^{-l_i}$$

$$(\sum 2^{-l_i})^k = \sum_{j=0}^{k \max l_i} (2^{-j} + 2^{-j} + 2^{-j}) - \text{всего} \leq 2^j \text{ слов}$$

$$\sum_{j=0}^{k \max l_i} (2^{-j} + 2^{-j} + 2^{-j}) \leq k \max l_i$$

$$\forall k : x^k \leq k \max l_i \to x \leq 1$$

$\square$

Определение. **Префиксный** код: $\forall c \neq d \quad \varphi(c)$ — не префикс $\varphi(d)$

Лемма 2. Префиксный код — однозначно декодируем.

Доказательство. Докажем "$\Leftarrow$"

$\sum 2^{-l_i} \leq 1 \Rightarrow \exists$ префиксный код с длинами $l_1 \dots l_s$

$$l_1 \leq l_2 \leq \ldots \leq l_s$$

2^{-l_1}

$2^{-l_1} + 2^{-l_2}$

$\vdots$

$2^{-l_1} + 2^{-l_2} + \ldots + 2^{-l_s}$

$$S = 2 \quad 2^{-l_1} + 2^{-l_2} \leq 1$$

□

Тут автор сдох.

Следствие 1. $\exists$ однозначно декодируемый код с длинами $l_1 \dots l_s \Rightarrow \exists$ префиксный код с длинами $l_1 \dots l_n$

1.1 Код Хаффмана

Дано: $f_1, f_2 \dots f_s$ — как часто встречаются соответствующие слова. Найти $l_1 \dots l_s$, такие что $\sum 2^{-l_i}$ и $\sum l_i f_i \to \min$

$S = 2 \Rightarrow l_1 = l_2 = 1$

$S > 2$ Возьмём два символа x и y, такие что f_x и $f_y \to \min$ *(x, y – самые редкие)*. Заменим их на z, $f_z = f_x + f_y$.

Возьмём в качестве кодового слова для x слово для $z + 0$, а для y возьмём $z + 1$.