Лек. 6. **Кодирование дискретных источников сигнала.** 2.03.2021г.

Обознач.: $\mathcal{A} = \{a_1 \dots a_m\}$ - алфавит источника (ДИС)

$\mathcal{B} = \{b_1, \dots, b_D\}$ - кодовый алфавит

$\mathcal{B}^* := \bigcup\limits_{n \geq 1} \mathcal{B}^n$ - множество всех конечных последовательностей из алф. $\mathcal{B}$.

ОПР. D-ичное кодирование - это отображение

$\varphi: \mathcal{A} \to \mathcal{B}^*$

$a_i \mapsto \varphi(a_i)$ - кодовое слово

Обозначим $l_i := len(\varphi(a_i))$ - длина кодового слова.

$\varphi(\mathcal{A}) := \{\varphi(a_i) \ \forall a_i \in \mathcal{A}\}$ — D-ичный код для алфавита $\mathcal{A}$.

Код

равномерный: все $l_i = l_j =: l$ - длина кода $\forall i, j$

неравномерный: $\exists i, j: l_i \neq l_j$.

Код

Сингулярный (φ - не инъекция)

Несингулярный: $a_i \neq a_j \Rightarrow \varphi(a_i) \neq \varphi(a_j)$ (φ - инъекция)

ОПР. Кодирование конечного сообщения $a^n = (a_{i_1}, \dots, a_{i_n})$ - это операция применения отображ.

$\varphi^*: \mathcal{A}^* = \bigcup\limits_{n \geq 1} \mathcal{A}^n \longrightarrow \mathcal{B}^*$

$\varphi^*(a_{i_1}, \dots, a_{i_n}) \longmapsto \varphi(a_{i_1}), \dots, \varphi(a_{i_n})$

конкатенация ('сцепление') кодовых слов.

ОПР. Кодир-е φ наз. префиксным (суффиксным), если никакое код. слово $\varphi(a_i)$ не явл. началом (окончанием) некот. другого кодового слова $\varphi(a_j)$, $i \neq j$.

Кодир-е φ наз. однозначно декодируемым, если φ^* - инъекция, т.е. $(a_{i_1}, \dots, a_{i_n}) \neq (a_{j_1}, \dots, a_{j_n}) \Rightarrow \varphi(a_{i_1}, \dots, a_{i_n}) \neq \varphi(a_{j_1}, \dots, a_{j_n})$.

$\forall \ \varphi^*(a_{i_1}, \dots, a_{i_n})$ однозначно разбивается на $\varphi(a_{i_1}) \dots \varphi(a_{i_n})$.

Пример: $\mathcal{A} = \{1, 2, 3, 4\}$, $\mathcal{B} = \{0, 1\}$ (двоичное кодир)

$\mathcal{A}$	$\varphi_1(a_i)$	$\varphi_2(a_i)$	$\varphi_3(a_i)$	$\varphi_4(a_i)$
1	0	0	10	0
2	0	010	00	10
3	0	01	11	110
4	0	10	110	111
	сингуляр. код	несингул, не декодир. однозначно	декодир. однознач.	префикс.

(Т) Если код является префиксным (суффиксным), то он является однозначно декодируемым (но не наоборот).

Д-во: опишем алгоритм декодирования — чтение слева направо.

Пусть $(b_1, b_2, \dots) \in \mathcal{B}^+$ — кодированное сообщение.

$\rceil S = \varnothing$.

Пока $S \neq \varphi(a_{i_1})$: $S \leftarrow b_i$;

a_i — декодир. единственным образом по опред. префикс.

Далее — по индукции.

(Т) Теорема.

1) Если G — $\mathcal{D}$-ичное размеченное корневое дерево, то множество меток всех его листьев образует префиксный $\mathcal{D}$-ичный код.

2) Если $\varphi: \mathcal{A} \to \mathcal{B}^*$ — $\mathcal{D}$-ичное префикс. кодирование, то $\exists$ такое $\mathcal{D}$-ичное размеч. корневое дерево G, для кот.

$\{\text{Метки всех листьев } G\} = \varphi(\mathcal{A})$

Д-во: 1) $\rceil$ дерево G имеет х/б 2 листа u и v.

Докажем, что $\mu(v)$ не явл. префиксом $\mu(u)$.

От противного: $\rceil \mu(v)$ — префикс $\mu(u) \Rightarrow u \in G_v \Rightarrow v$ — не лист. противоречие.

2) УПР.

(Т) Нер-во Крафта:

1) Если $\varphi: \mathcal{A} \to \mathcal{B}^*$ — $\mathcal{D}$-ичное префиксное алфавитное кодир-е с длинами кодовых слов $len(\varphi(a_i)) = l_i$, $1 \le i \le m$, то справедливо:

$$\sum_{i=1}^{m} \mathcal{D}^{-l_i} \le 1 \qquad (*)$$

2) Если натур. числа $\mathcal{D}, m, l_1, \dots, l_m$ удовл. $(*)$, то $\exists$ $\mathcal{D}$-ичное префиксное алфавитное кодир. $\varphi: \mathcal{A} \to \mathcal{B}^*$ с длинами кодовых слов $len(\varphi(a_i)) = l_i$, $1 \le i \le m$.

Д-во: Обозн. $L = \max_{i=\overline{1,m}} \{l_i\}$ — длина кода

1) Рассмотрим полное размеч. корневое дерево G высоты L.
(число листьев b^L)

Выделим вершины v_i, для кот. $\mu(v_i)$ – код. слова, рассм. поддеревья G_{v_i}.

Высота G_{v_i} $\to$ $L-l_i$, число листьев b^{L-l_i}, причем $G_{v_i}\cap G_{v_j}=\emptyset$, т. к. код префикс.

$$b^{L-l_1}+\ldots+b^{L-l_m}\le b^L \Rightarrow b^{-l_1}+\ldots+b^{-l_m}\le 1.$$

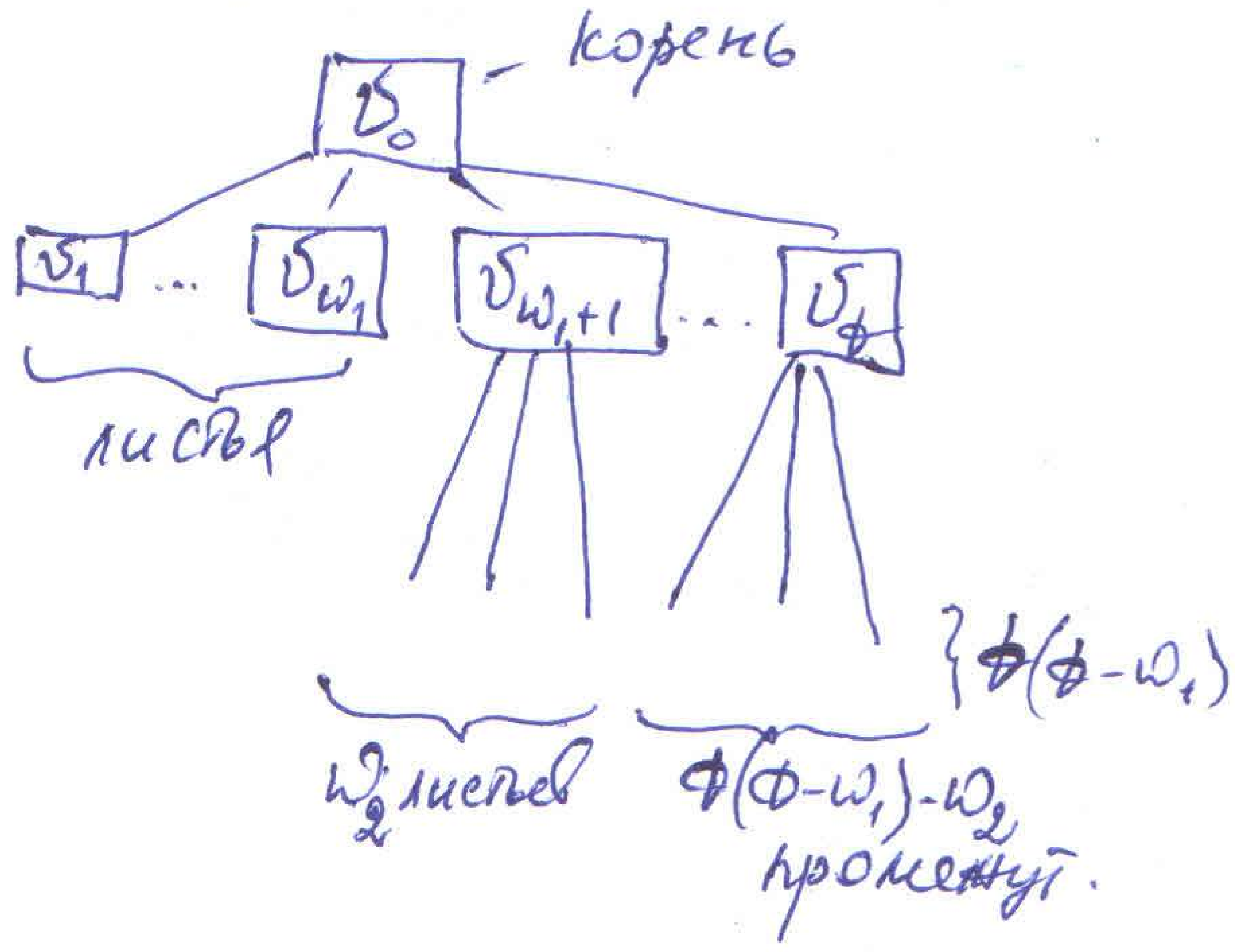
2) Для $l\in\{1,\ldots,L\}$ обозн. w_l – колич. чисел набора $(l_1,\ldots,l_m)$, равных l, тогда:

$$(*)\Leftrightarrow \sum_{l=1}^{L} w_l\cdot b^{-l}\le 1$$

$w_L>0$ по опред. L $\Rightarrow$ $\sum_{l=1}^{k} w_l\cdot b^{-l}<1$, $k=1,\ldots,L-1$.

$$\Leftrightarrow b^k-\sum_{l=1}^{k} w_l\cdot b^{k-l}>0,\quad k=1,\ldots,L-1 \qquad (**)$$

Построим дерево G с m листьями, метки кот. имеют длины $l_1,\ldots,l_m$.



$k=1$: $(**)\Leftrightarrow b-w_1>0$

$k=2$: $(**)\Leftrightarrow b(b-w_1)-w_2>0$

Т (McMillan) Если алф. кодир. $\varphi:\mathcal{A}\to\mathcal{B}^*$ с длинами код. слов $len(\varphi(a_i))=l_i$, $1\le i\le m$, явл. однозначно декодир. $\Rightarrow$ вып. (*).

Д-во: для $n,l\in\mathbb{N}$ обозн. $C_{n,l}=\{a^n\in\mathcal{A}^n \mid len(\varphi^*(a^n))=l_{i_1}+\ldots+l_{i_n}=l\}$, $c_{n,l}=\#C_{n,l}$.

$\varphi^*:\mathcal{A}^*\to\mathcal{B}^*$ – инъектив. по усл. $\Rightarrow$ $\varphi^*(C_{n,l})\subseteq\mathcal{B}^l\Rightarrow c_{n,l}\le|\mathcal{B}|^l=b^l$.

$$*\left(\sum_{i=1}^{m} b^{-l_i}\right)^n=\sum_{1\le i_1,\ldots,i_n\le m} b^{-(l_{i_1}+\ldots+l_{i_n})}=\sum_{(a_{i_1},\ldots,a_{i_n})\in\mathcal{A}^n} b^{-len(\varphi^*(a_{i_1},\ldots,a_{i_n}))}=$$

$$=\sum_{l=1}^{nL}\underbrace{c_{n,l}b^{-l}}_{\le 1},\quad \text{где } L=\max\{l_1,\ldots,l_m\}$$

$$\left(\sum_{i=1}^{m} b^{-l_i}\right)^n\le nL\quad \forall n\ge 1\Rightarrow \sum_{i=1}^{m} b^{-l_i}\le 1,\ \text{т. к. иначе}\ \lim_{n\to\infty}\frac{\left(\sum_{i=1}^{m} b^{-l_i}\right)^n}{nL}=+\infty$$