

12 дәріс тақырыбы. Кодтау теориясының элементтері. Алфавиттік кодтау. Оңтайлы кодтау. Қателерді табу және түзету кодтары.

1. Код-бұл әртүрлі ақпаратты беру, өңдеу және сақтау (есте сақтау) үшін шартты белгілер (таңбалар) жүйесі. Кодтау теориясын зерттеу тақырыбы 0, 1, ..., $r-1$ сандарынан көптеген тізбектердегі еркін табиғат объектілерінің шектеулі немесе есептелетін жиындарын бейнелеулері болып табылады, мұндағы r – оң бүтін сан (атап айтқанда $r = 2$). Мұндай бейнелеулер кодтау деп аталады.

Кодтау теориясының көптеген міндеттері келесі схемаға сәйкес келеді. Берілген объектілер жиынтығы үшін белгілі бір қасиеттері бар кодтау класы қарастырылады. Қарастырылған класстан кодтауды алдын-ала анықталған мағынада оңтайлысын құру қажет. Әдетте, кодтаудың оңтайлылық критерийі кодтардың ұзындығын азайтумен байланысты, ал қажетті кодтау қасиеттері әр түрлі болуы мүмкін. Осындай қасиеттердің ішінде: бір мәнді кері бейнелеуі болуы (декодтау), әр түрлі қателерді декодтау кезінде түзету мүмкіндігі, кодтау мен декодтаудың қарапайым іске асырылуы (алгоритмнің қарапайымдылығы) және т. б.

А еркін алфавит болсын. А алфавитінің элементтері әріптер деп, ал әріптерден тұратын шектеулі тізбектер (кортеждер) сөздер деп аталады. a сөздерінің ұзындығы (әріптер саны) $l(a)$ арқылы белгіленеді, ал 0 ұзындықты сөзі (бос сөз) L белгісімен белгіленеді. a_1 және a_2 сөздерінің тіркесімі a_1a_2 арқылы, ал бірдей сөздердің n тіркесімі a^n ($a^0 = L$) арқылы белгіленеді.

U – А алфавиттегі еркін сөздер жиынтығы болсын. U^n ($n = 0, 1, \dots$) A^n алфавиттегі еркін сөздер жиынтығы болады. Атап айтқанда A^n ($n = 0, 1, \dots$) А алфавитіндегі n ұзындықты барлық сөздерінің жиынтығын, Ал A^* арқылы – А-дағы барлық сөздердің жиынтығын белгілейміз.

$$U^* = \bigcup_{n=0}^{\infty} U^n$$

А-дағы кейбір сөздердің жиынтығын болады. Оның қуаттылығы U жиынының таңдауына байланысты.

Мысалы, егер $U = \{00, 11\}$ екі сөзден тұрса, онда олардың кез-келген байланысы, біріншіден, жұп ұзындыққа ие, екіншіден, нөлдер мен бірліктер жұп болып орналасады; мысалы, 0110 сөзі U^2 жиынында жатпайды.

a_1 сөзі $a = a_1a_2$ деген сөз бар болса, a сөзінің басы деп аталады; егер $a_2 \neq L$ онда a_1 сөзі a сөзінің меншікті басы деп аталады. a_2 сөзі a сөзінің соңы деп аталады, егер $a = a_1a_2$ сөзі болса; егер $a_1 \neq L$ онда a_2 сөзі a сөзінің меншікті

соңы деп аталады. U - жиынындағы сөздердің басталуы $\vec{U}$ арқылы, ал соңыларды $\overleftarrow{U}$ белгілейміз.

2. $B_{(r)} = \{0, 1, \dots, r-1\}$ алфавитті қарастырайық, мұндағы $r \geq 2$, және еркін жиын G . Атап айтқанда, G -ақырлы алфавит, натурал сандар жиынтығы, белгілі бір типтегі схемалар немесе формулалар жиынтығы, белгілі бір алфавиттегі сөздер жиынтығы және т.б. G жиынын алфавиттегі сөздер жиынына ерікті түрде бейнелеуді G жиынын r -әріптік кодтау деп аталады. Біз тек $r = 2$ жағдайын қарастырамыз, яғни екілік кодтау қарастырамыз.

Мысалы, Екілік жүйдегі санадардың жазылуы.

$G = \{0, 1, 2, 3, 4, 5, 6, 7\}$

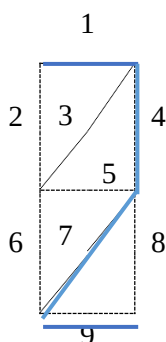
a_i		v_i
0	0	000
1	1	001
2	10	010
3	11	011
4	100	100
5	101	101
6	110	110
7	111	111

$1_10=12, 11_0 = 30, 110 = 6$

001 100 110 111 010

1 4 6 7 2 = 14672

Почта индексiнiң кодтауы



0 1 2 3 4 5 6 7 8 9

$2 \Rightarrow 100100101$

$3 \Rightarrow 101010100$

$7 \Rightarrow 101001000$

$5 \Rightarrow 110010011$

Шектеулі жиындарды екілік сөздердің кейбір ішкі жиындарына бейнелеу көбінесе осы жиындардың элементтер санын есептеу немесе бағалау үшін ыңғайлы құрал болып табылады. Мысалы, $F_{n,m}$ – n санының m бүтін теріс емес қосындысына реттелген бөліністерінің жиынын қарастырайық. Әрбір

$n = n_1 + n_2 + \dots + n_m$ бөлінуіне $0^{n_1}10^{n_2}1\dots10^{n_m}$ сөзді сәйкестендіреміз. Оның ұзындығы $(n + m - 1)$, өйткені оның құрамында нөлдер сериясын бөлетін n нөлдер және $(m-1)$ бірліктер бар. Сонымен, 11 санын 4 қосындысына бөлу

$11 = 5 + 0 + 2 + 4 \Rightarrow 00000110010000$ сөзімен кодталған,

бөлім $11 = 2 + 7 + 2 + 0$ – сөз $\Rightarrow 00100000001001$

және бөлу $11 = 1 + 1 + 6 + 3 \Rightarrow 01010000001000$ сөзі.

Мұндай кодтау $F_{n,m}$ жиынының біркелкі $(m - 1)$ бірліктері бар ұзындықтың $(n + m-1)$ екілік сөздерінің жиынтығына бір-біріне ұқсамайтын бейнелеуі.

Сондықтан $F_{n,m}$ жиынының элементтерінің саны тең C_{n+m-1}^{m-1} .

Кодтау теориясында белгілі бір A алфавиттегі сөздердің барлығынан (немесе таңдалған бөлігінен) тұратын G жиынтығын кодтау ерекше рөл атқарады. Кодталған жиынның элементтері (яғни A -дағы арнайы сөздер) байланыс арнасы арқылы кодтың берілуін мүмкіндігін ескере отырып, **хабарламалар** деп аталады. Жалпы жағдайда, оған сәйкес екілік сөзді хабарлау үшін есептеу процесіне (алгоритмнің өзі, оның тиімділігі) ешқандай жағдай жасалмайды. Алайда, көптеген сұрақтар үшін әріптік кодтау деп аталатын сөздерді кодтаудың тар кластарын қарастырумен шектелуі жеткілікті.

$A = \{a_i\}$ әріптері натурал сандармен нөмірленген шектеулі алфавит болсын. Бұл жағдайда A алфавитінің әріптерін кодтауды екілік сөздермен орнатуға болады $V = \{v_i\}$, мұндағы v_i – A алфавиттегі a_i әрібінің бейнесі. $V = \{v_i\}$

сөздерін кодтар деп атаймыз (A алфавитінің). Әр $a_{i_1}a_{i_2}\dots a_{i_k}$ сөзге

(хабарламаға) сәйкес келетін A алфавитіндегі сөздерді $v_{i_1}v_{i_2}\dots v_{i_k}$ кодтау

әріптік кодтау деп аталады және $K_{v_i}^{a_i}$ (немесе K_V^A) арқылы белгіленеді.

Мысал. Алайда $A = \{a, b, c, d\}$ болсын, оның екілік коды болсын –

$a \rightarrow 01, b \rightarrow 100, c \rightarrow 101, d \rightarrow 0.$

0100 сөзін db немесе add; 0010100 сөзін ddcdd, daadd немесе dadb деп түсінуге болады.

$V = \{v_i\}$ кодында (сөздер жиынтығы) тек әртүрлі сөздер бар деп болжауға болады. $V = \{v_i\}$ кодын бөлінетін деп атаймыз, егер әр

$$V_{i_1} V_{i_2} \dots V_{i_k} = V_{j_1} V_{j_2} \dots V_{j_l}$$

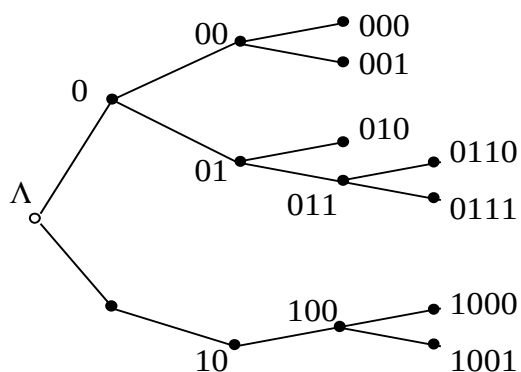
теңдіктен $V = \{0, 1\}$ алфавитінде $k = l$ және $i_t = j_t$, $t = 1, 2, \dots, k$ болады

Бұл әріптік кодтау бірімәнді болып табылады, сонда және тек сонда ғана, егер ол бөлінетін болса. Анықтамадан бөлінетін кодтың барлық сөздері әр түрлі және бос емес екендігі де шығады.

$V = \{v_i\}$ коды префиксті деп аталады, егер v_k сөзі v_j сөзінің басы болмаса, $j \neq k$. Префиксті коды бөлінетіні анық, яғни префиксті бөлінудің жеткілікті шарты болып табылады. Егер префиксті кодының әр сөзі басқа кодтық сөздердің басы емес, оның ең кіші басталуымен алмастырылса, онда алынған код префиксті болады. Бұл операция префиксті кодының қысқарту деп аталады.

Префиксті кодының қарапайым мысалы-біркелкі код. Ол бөлгіштерді қажет етпейді, өйткені әр жіберілген әріптің коды аяқталады және келесі әріптің коды белгілі бір бірдей кодтау таңбаларынан басталады.

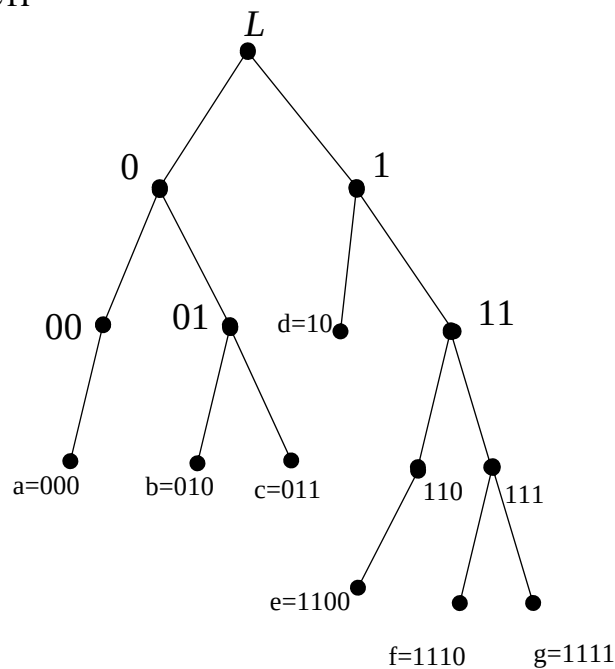
i	l_i	v_i
0	3	000
1	3	001
2	3	010
3	4	0110
4	4	0111
5	4	1000
6	4	1001



Жаттығу. Кодтық ағашын салу

a) a = 000	б) a = 0000	в) a = 00
b = 010	b = 0001	b = 0101
c = 011	c = 0011	c = 011
d = 10	d = 01	d = 1100
e = 1100	e = 100	e = 1101
f = 1110	f = 101	f = 111
g = 1111	g = 111	

а) Префиксті



Бұдан әрі $V = \{v_0, v_1, \dots, v_{m-1}\}$ шектеулі код үшін ($m \geq 2$)

$$\lambda_{\max} = \max_{0 \leq i < m-1} \lambda(v_i)$$

белгіні қолданамыз.

Теорема 1. $l_0, l_1, \dots, l_{m-1}$ – натурал сандардың еркін жиынтығы болсын ($m \geq 2$).
Берілген $\lambda(v_i) = l_i$, $i = 0, 1, \dots, m-1$ ұзындықты $V = \{v_0, v_1, \dots, v_{m-1}\}$ код бөлінетін болуы үшін **бөлінетін кодтар үшін Крафт-Макмилланның теңсіздігін** орындалуы қажет және жеткілікті:

$$\sum_{i=0}^{m-1} \frac{1}{2^{l_i}} \leq 1.$$

Кестеде көрсетілген Код үшін:

$$3 \cdot \frac{1}{2^3} + 4 \cdot \frac{1}{2^4} = \frac{3}{8} + \frac{4}{16} = \frac{5}{8} < 1$$

Сал. Кез келген бөлінетін код үшін $V = \{v_0, v_1, \dots, v_{m-1}\}$ V кодымен бірдей кодтық сөз ұзындығының жиынтығы бар префиксті коды табылады.

Өңтайлы кодтау

1. Мұнда біз хабарлама көзінің статистикалық қасиеттеріне қатысты кейбір қарапайым болжамдармен тиімді бір мәнді кодтаудың құрылысын талдаймыз. Сонымен қатар, кодтау тиімдірек деп саналады, онда хабардың орташа әрпінде екілік сандардың аз саны болады.

$A = \{a_0, a_1, \dots, a_{m-1}\}$ алфавитінің әріптерін кездейсоқ түрде (яғни кездейсоқ ретпен) шығаратын дереккөз қарастырылады. A алфавитінің әріптерінің дәйекті пайда болуы статистикалық тұрғыдан тәуелсіз және

$$P = \{p_0, p_1, \dots, p_{m-1}\} \quad (p_i \geq 0, \sum_{i=0}^{m-1} p_i = 1).$$

ықтималдылықты бөлуге бағынады деп болжанады

Әрбір код $B = \{0, 1\}$ алфавитіндегі $V = \{v_0, v_1, \dots, v_{m-1}\}$ келесі орташа санмен сипатталады

$$L_V(P) = \sum_{i=0}^{m-1} p_i \cdot \lambda(v_i)$$

әріптік кодтау кезінде A алфавитінің бір әрпіне сәйкес келетін екілік

$K_{v_0}^{a_0}, K_{v_1}^{a_1}, \dots, K_{v_{m-1}}^{a_{m-1}}$ сандар. P бөлу кезінде V кодының мәні деп аталады

$L_V(P)$, мұндағы сөздерінен тұратын V префикс кодтарының жиынтығы бойынша алынады. Префикс коды $V = \{v_0, v_1, \dots, v_{m-1}\}$ бөлу кезінде оңтайлы

деп атаймыз $P = \{p_0, p_1, \dots, p_{m-1}\}$, егер $L(P) = \inf L_V(P)$. Мұнда туындайтын негізгі міндет-оңтайлы немесе оларға жақын кодтарды құру әдістерін табу және олардың құнын бағалау.

Фано коды.

Әріп	ықтималдығы	қадамдар					Фано коды	Ұзындығы	Бәр келкі коды	Ұзындығы
		1	2	3	4	5				
d	0,21	0,41	0,21				00	2	000	3
e	0,2		0,20				01	2	001	3
a	0,19	0,59	0,32	0,19			100	3	010	3
b	0,13			0,13			101	3	011	3
h	0,1		0,27	0,10			110	3	100	3
c	0,08			0,17	0,08		1110	4	101	3
g	0,07				0,09	0,07	11110	5	110	3
f	0,02					0,02	11110	5	111	3
	1,0						Құны	2,85	Құны	3

Әріп	ықтималдығы	біркелкі коды
a	0,10	000
b	0,15	001
c	0,08	010
d	0,21	011
e	0,11	100
f	0,15	101
g	0,07	110
h	0,13	111
	1	

Әріп	ықтималдығы	Қадамдар				бір өлшемді код	ұзындығы	коды	ұзындығы
		1	2	3	4				
d	0,21	0.51	0.21			000	3	00	2
b	0,15		0.30	0.15		001	3	010	3
f	0,15			0.15		010	3	011	3
h	0,13	0.49	0.24	0.13		011	3	100	3
e	0,11			0.11		100	3	101	3
a	0,10		0.25	0.10		101	3	110	3
c	0,08			0.15	0.08	110	3	1110	4
g	0,07				0.07	111	3	1111	4
	1					Құны = 3		Құны = 2.94	

$$2*0,21+3*(0,15+0,15+0,13+0,11+0,10)+4*(0,08+0,07) =$$

$$= 0,42+1,92+0,60=2,94<3,0$$

Әріп	Ықтималдығы	Біркелкі коды
a	0,28	000
b	0,04	001
c	0,12	010
d	0,12	011
e	0,33	100
f	0,11	101
	1	

Әріп	Ықтималдығы	қадамдар			коды	ұзындығы
		1	2	3		
e	0,33	0,61	0,33		00	2
a	0,28		0,28		01	2
c	0,12	0,39	0,24	0,12	100	3
d	0,12			0,12	101	3
f	0,11		0,15	0,11	110	3
b	0,04			0,04	111	3

$$L_v(P) = 2*0.61+3*0.39 = 1.22+1.17 = 2.39 < 3$$

Қателерді анықтау және түзету кодтары

1. $V=\{0,1\}$ алфавитіндегі n ұзындықты сөздердің V^n жиынын қарастырамыз. Әр $V=\{0,1\}$ алфавиттегі $x_1x_2...x_n$ сөзді n -өлшемді $(x_1, x_2, ..., x_n)$ вектормен сәйкестіндіреміз. V^n дағы $(x_1, x_2, ..., x_n)$ және $(y_1, y_2, ..., y_n)$ векторлардың қосындысы деп $(x_1 \oplus y_1, x_2 \oplus y_2, ..., x_n \oplus y_n)$ векторды атаймыз. Естеріңізге сала кетейік, $\oplus$ белгісі 2-лік модулінің қосындысын білдіреді. V^n жиынында 0 немесе 1 тең болатын $b \in V$ тұрақтыға $X = (x_1, x_2, ..., x_n)$ векторды көбейтіндіні келесідей еңгіземіз: $bX = (bx_1, bx_2, ..., bx_n)$. Олай болса V^n n өлшемді векторлық кеңістік ретінде қарастыруға болады.

V^n векторлық кеңістігінде X және Y векторлары арасындағы $d(X, Y)$ Хемминг қашықтығын X және Y векторларының сәйкес келмейтін

компоненттерінің саны және X векторының нормасы X және $(0,0,...,0)$ -нөлдік n өлшемді вектор арасындағы қашықтық ретінде анықтауға болады. Кез келген $X = (x_1, x_2, ..., x_n)$ үшін, оның нормасы тең болады:

$$\|X\| = \sum_{j=1}^n x_j$$

сондай ақ кез келген $X = (x_1, x_2, ..., x_n)$ және $Y = (y_1, y_2, ..., y_n)$ векторлар үшін

$$d(X, Y) = \|X \oplus Y\|$$

B^n метрикалық кеңістігі қарапайым геометриялық түсіндіруге мүмкіндік береді: B^n жиынтығы n -өлшемді бірлік кубтың төбелер жиынына сәйкес келеді, ал B^n -дағы екі элемент арасындағы қашықтық кубтың тиісті төбелерді қосатын ең қысқа тізбектегі қабырғалар санына тең.

Еркін $X = (x_1 x_2 ... x_n) \in B^n$ сөз үшін оның $N(X)$ сандық мәнін анықтаймыз:

$$N(X) = \sum_{j=1}^n x_j \cdot 2^{n-j}$$

Мысалдар. $X = 0110011$ және $Y = 0001101$ болсын. Содан кейін $\|X\| = 4$, $\|Y\| = 3$, $X \oplus Y = 0111110$, $d(X, Y) = 5$, $N(X) = 1 + 2 + 16 + 32 = 51$, $N(Y) = 1 + 4 + 8 = 13$. Сондай-ақ, мысалы, 19 сан үшін $e(19) = 10011$, $e_7(19) = 0010011$.

Болашақта бұл бөлімде тек $m \geq 2$ жолдан тұратын кодтар қарастырылады, ал код сөздерінің нөмірленуі маңызды емес жағдайларда біз оны көрсетпейміз.

Еркін $V = \{v_0, v_1, ..., v_{m-1}\} \subseteq B^n$ код үшін келісеміз:

$$d(V) = \min_{i \neq j} d(v_i, v_j)$$

Бұл $d(V)$ сан кодтық қашықтық деп аталады.

Жаттығу. $V = \{a_1: 0100, a_2: 1101, a_3: 1110, a_4: 0010\}$ код үшін кодтық қашықтықты санаймыз. Ол үшін $d(a_i, a_j)$, $i \neq j$ сандарды есептеп арасынан максималдысын таңдаймыз.

	a_1	a_2	a_3	a_4
a_1	x	2	2	2
a_2	2	x	2	4
a_3	2	2	x	2
a_4	2	4	2	x

$$d(V) = 2.$$

2. Қателер деп аталатын екілік сөздерді түрлендірудің кейбір түрлерін қарастырамыз.

$0 \rightarrow 1$ ($1 \rightarrow 0$) түріндегі жалғыз қателік X сөзіндегі 0-дің орнына 1 (сәйкесінше 1-дің орнына 0) таңбамен ауыстыру нәтижесі деп аталады. Бұл түрдегі жалғыз қателер таңбаларды ауыстыру қателер деп те аталады.

$0 \rightarrow L$ ($1 \rightarrow L$) түріндегі жалғыз қателік X сөзіндегі 0 (сәйкесінше 1) таңбасы жойылып қалу нәтижесі деп аталады. Бұл түрдегі жалғыз қателер таңбалардың жоғалуы деп аталады.

$L \rightarrow 0$ ($L \rightarrow 1$) түріндегі жалғыз қателік X сөзіндегі кейбір символдың алдында немесе оның соңғы символынан кейін 0 (сәйкесінше 1) таңбасын енгізу нәтижесі деп аталады; бұл жағдайда сөздің ұзындығы бірлікке артады.

$X \in B^n$ сөзіндегі $+2^i$ (-2^i), мұндағы $0 \leq i < n$, түріндегі жалғыз қателік X сөзін Y сөзге түрлендіру деп аталады, оның сандық мәні X сөзінің сандық мәнінен 2^i үлкен (сәйкесінше аз). Бұл түрдегі жалғыз қателер арифметикалық қателер деп аталады.

Болашақта $0 \rightarrow 1$ ($1 \rightarrow 0$) қателіктерге, яғни таңбаларды ауыстыру қателерге, ерекше назар аударылады

Берілген типтегі қателерді анықтау және түзетудің әртүрлі міндеттерін қою кезінде білім беру заңындағы қателіктердің бір немесе басқа болжамынан туындайды. Бұл заңдар ықтималды да, комбинаторлық сипатта да болуы мүмкін. Қателіктердің пайда болуының белгілі бір заңы екілік сөздерді беру (немесе сақтау) арнасы арқылы берілген деп айтылады. Арналардың келесі екі түрі жиі қарастырылады:

а) p ($0 < p < 1/2$) ықтималдығы бар кез-келген таңбада берілген типтегі қате пайда болады, әр түрлі таңбалардағы қателер статистикалық тұрғыдан тәуелсіз;

ә) n ұзындығының әрбір сөзінде s -ден аспайтын жиіліктің берілген түрінің кез келген қатесі орын алуы мүмкін.

Шуылға төзімді кодтаудың негізгі идеясы келесідей. $X, Y, Z, \dots$, хабарламаларының орнына $X', Y', Z', \dots$, кодталған олардың бейнелері байланыс арнасы арқылы белгілі бір жолмен жіберіледі. Аталған бейнелер қателерді анықтауға немесе дұрыс декодтауға мүмкіндік береді, егер рұқсат етілген типтегі қателер болса да.

$$X = 01101$$

$s=1, n=5$. Әр символды $2*s+1$ рет қайталаймыз. Яғни X' сөздің ұзындығы тең $n(2*s+1) = 15$

$X' = 000111111000111$ ақпаратты жібердік.

$X'' = 000101111000111$ қабылдаушы осында хабарламаны қабылдады.

$X''' = 000 \underline{111} 111 000 111$ кері кодтайды

Хемминг коды

алғашқы ақпарат $X = 01101001101$

01001001101

орын	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
типі	т	т	а	т	а	а	а	т	а	а	а	а	а	а	а
символ			0		1	1	0		1	0	0	1	1	0	1

а – ақпараттық орындар

т – тексеру орны (2-нің дәрежесіне тең орындар)

$$2^0=1, 2^1=2, 2^2=4, 2^3=8, \dots$$

0	0	0	0	0	типі	X	толықтыруы	X'
1	0	0	0	1	т		0	0
2	0	0	1	0	т		1	1
3	0	0	1	1	а	0		0
4	0	1	0	0	т		0	0
5	0	1	0	1	а	1		1



6	0	1	1	0	a	0		0
7	0	1	1	1	a	0		0
8	1	0	0	0	т		0	0
9	1	0	0	1	a	1		1
10	1	0	1	0	a	0		0
11	1	0	1	1	a	0		0
12	1	1	0	0	a	1		1
13	1	1	0	1	a	1		1
14	1	1	1	0	a	0		0
15	1	1	1	1	a	1		1

0	1	0	1	a	1
1	0	0	1	a	1
1	1	0	0	a	1
1	1	0	1	a	1
1	1	1	1	a	1
0	0	1	0		



$X' = \underline{0}\underline{1}00\underline{1}000\underline{1}001101$

$X'' = \underline{0}\underline{0}\underline{0}\underline{1}1100\underline{1}000\underline{0}101$

0	0	0	0	0	типі	X''
1	0	0	0	1	т	0
2	0	0	1	0	т	0
3	0	0	1	1	a	0
4	0	1	0	0	т	1

5	0	1	0	1	a	1
6	0	1	1	0	a	1
7	0	1	1	1	a	0
8	1	0	0	0	т	0
9	1	0	0	1	a	1
10	1	0	1	0	a	0
11	1	0	1	1	a	0
12	1	1	0	0	a	0 1
13	1	1	0	1	a	1
14	1	1	1	0	a	0
15	1	1	1	1	a	1

0	1	0	0	т	1
0	1	0	1	a	1
0	1	1	0	a	1
1	0	0	1	a	1
1	1	0	1	a	1
1	1	1	1	a	1
1	1	0	0		