Элементы дискретной математики

1 Функции алгебры логики

1.1 Основные определения.

Пусть $E = \{0; 1\}$.

Определение 1 *Булева функция* (*функция алгебры логики*) от n переменных — это отображение $f: E^n \to E$.

Если $(x_1, \ldots, x_n)$ — набор значений переменных, каждая из которых равна либо 0, либо 1, то $f(x_1, \ldots, x_n)$ есть значение функции. Булеву функцию от n переменных можно задавать таблицей из 2^n строк и $n + 1$ столбца.

Пример 1 Рассмотрим следующее задание функции f таблицей.

x_1	x_2	x_3	$f(x_1, x_2, x_3)$
0	0	0	1
0	0	1	0
0	1	0	1
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	0

Обозначим через $\mathcal{P}_2$ систему всех функций алгебры логики. Пусть $p_2(n)$ есть количество функций алгебры логики от n переменных.

Теорема 1

$$p_2(n) = 2^{2^n}$$

Для доказательства заметим, что существует в точности 2^n упорядоченных наборов длиной n из нулей и единиц. Булева функция однозначно задается своими значениями на этих наборах. Так как на данном наборе функция может принимать любое из двух значений, получаем в точности 2^{2^n} функций.

Удобно считать, что при $n = 0$ существуют две функции от пустого множества переменных — константы 0 и 1.

Определение 2 Пусть $f: E^n \to E$ — булева функция от n переменных. Переменная x_i $(1 \le i \le n)$ называется *существенной*, если существуют $\alpha_1, \ldots, \alpha_{i-1}, \alpha_{i+1}, \ldots, \alpha_n \in \{0;1\}$ такие, что $f(\alpha_1, \ldots, \alpha_{i-1}, 0, \alpha_{i+1}, \ldots, \alpha_n) \neq f(\alpha_1, \ldots, \alpha_{i-1}, 1, \alpha_{i+1}, \ldots, \alpha_n)$. В противном случае переменная x_i называется *несущественной* или *фиктивной*. Это означает, что $f(x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_n) = f(x_1, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_n)$ для любых $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n \in \{0;1\}$.

Если x_i — фиктивная переменная для $f(x_1, \ldots, x_n)$, то можно образовать функцию $g(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ от $n-1$ переменной, полагая

$$g(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) = f(x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_n) = f(x_1, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_n).$$

При этом говорят, что g получена из f *удалением фиктивной переменной*, а f получена из g *введением фиктивной переменной*.

Замечание 1 В дальнейшем бу ем рассматривать функции алгебры логики с точностью до фиктивных переменных, т.е. не будем различать функции, получаемые друг из друга путем добавления и изъятия фиктивных переменных.

В силу вышесказанного, если дана **конечная** система функций $f_1, \ldots, f_s$ из $\mathcal{P}_2$, то можно считать, что все они зависят от одних и тех же переменных $x_1, \ldots, x_n$.

1.2 Примеры функций алгебры логики.

1) $f_1(x) = 0$ — *константа* 0.

2) $f_2(x) = 1$ — *константа* 1.

3) $f_3(x) = x$ — *тождественная* функция.

4) $f_4(x) = \bar{x}$ — *отрицание* x.

x	0	1	x	$\bar{x}$
0	0	1	0	1
1	0	1	1	0

5) $f_5(x_1, x_2) = (x_1 \,\&\, x_2)$ — *конъюнкция* x_1 и x_2
(другие обозначения: $x_1 \wedge x_2$, $x_1 \cdot x_2$, $x_1 x_2$).

6) $f_6(x_1, x_2) = (x_1 \vee x_2)$ — *дизъюнкция* x_1 и x_2.

7) $f_7(x_1, x_2) = (x_1 \to x_2)$ — *импликация* x_1 и x_2.

8) $f_8(x_1, x_2) = (x_1 \oplus x_2)$ — *сложение* x_1 и x_2 *по модулю* 2
(другое обозначение: $x_1 + x_2$).

9) $f_9(x_1, x_2) = (x_1 \mid x_2)$ — *штрих Шеффера*.

x_1	x_2	x_1x_2	$x_1 \vee x_2$	$x_1 \to x_2$	$x_1 \oplus x_2$	x_1 | x_2
0	0	0	0	1	0	1
0	1	0	1	1	1	1
1	0	0	1	0	1	1
1	1	1	1	1	0	0

Булевы функции 1 – 9 играют в алгебре логики большую роль и часто называются "элементарными" функциями алгебры логики. Рассмотрим некоторые тождества, справедливые для этих функций. Каждое из них легко может быть проверено непосредственно (если тождество зависит от n переменных, то достаточно рассмотреть 2^n случаев).

Пусть $\circ$ обозначает одну из функций (операций): $\&$, $\vee$, $\oplus$. Каждая из этих операций ассоциативна:

$$((x_1 \circ x_2) \circ x_3) = (x_1 \circ (x_2 \circ x_3))$$

и коммутативна:

$$(x_1 \circ x_2) = (x_2 \circ x_1)$$

(здесь и далее переменные принимают произвольные значения из множества $E = \{0; 1\}$).

Для конъюнкции и дизъюнкции чыполнены дистрибутивные законы:

$$((x_1 \vee x_2) \,\&\, x_3) = ((x_1 \,\&\, x_3) \vee (x_2 \,\&\, x_3)),$$

$$((x_1 \,\&\, x_2) \vee x_3) = ((x_1 \vee x_3) \,\&\, (x_2 \vee x_3)).$$

Между конъюнкцией, дизъюнкцией и отрицанием имеют место следующие соотношения:

$$\bar{\bar{x}} = x \qquad \text{(закон двойного отрицания)},$$

$$\overline{(x_1 \,\&\, x_2)} = \bar{x}_1 \vee \bar{x}_2, \quad \overline{(x_1 \vee x_2)} = \bar{x}_1 \,\&\, \bar{x}_2 \quad \text{(законы де Моргана)}.$$

Отметим еще несколько свойств конъюнкции и дизъюнкции:

$$(x \,\&\, x) = x, \qquad (x \vee x) = x,$$

$$(x \,\&\, \bar{x}) = 0, \qquad (x \vee \bar{x}) = 1,$$

$$(x \,\&\, 0) = 0, \qquad (x \vee 0) = x,$$

$$(x \,\&\, 1) = x, \qquad (x \vee 1) = 1.$$

Замечание 2 С целью упрощения мы будем опускать скобки в записи выражений там, где это не приводит к недоразумениям. Мы будем считать, что конъюнкция *сильнее* дизъюнкции и сложения по модулю 2, т.е. выполняется в первую очередь при отсутствии скобок. Кроме того, внешние скобки часто будут опускаться. Скажем, $x_1 \,\&\, x_2 \vee x_3$ будет сокращенной записью для $((x_1 \,\&\, x_2) \vee x_3)$, а запись $x_1x_2 + x_3$ будет сокращением для $((x_1 \,\&\, x_2) + x_3)$.

С учетом ассоциативности операций конъюнкции, дизъюнкции и отрицания, мы будем использовать бесскобочные выражения вида $x_1 \circ x_2 \circ \cdots \circ x_m$, где $\circ$ — одна из операций $\&$, $\vee$ или $+$. Ясно, что результат применения этих операций не будет зависеть от расстановки скобок.

В дальнейшем мы будем употреблять такие обозначения:

$$\bigwedge_{i=1}^{m} x_i = x_1 \,\&\, x_2 \,\&\, \cdots \,\&\, x_m,$$

$$\bigvee_{i=1}^{m} x_i = x_1 \vee x_2 \vee \cdots \vee x_m,$$

$$\sum_{i=1}^{m} x_i = x_1 \oplus x_2 \oplus \cdots \oplus x_m.$$

Данные обозначения имеют смысл для всех $m \geq 1$. Удобно также считать, что при $m = 0$ первое выражение равно 1, а второе и третье 0.

Введем важное понятие двойственной функции.

Определение 3 Булева функция $f^*(x_1, \ldots, x_n)$ называется *двойственной* для функции $f(x_1, \ldots, x_n)$, если для любых значений переменных выполнено тождество

$$f^*(x_1, \ldots, x_n) = \overline{f(\bar{x}_1, \ldots, \bar{x}_n)}.$$

Легко заметить, что $(f^*)^* = f$ для любой функции f. В этом смысле можно говорить, что функции f и f^* двойственны друг другу. Нетрудно проверить, что константы 0 и 1 двойственны друг другу, функции x и $\bar{x}$ двойственны каждая сама себе (самодвойственны), конъюнкция и дизъюнкция двойственны друг другу.

Если дано выражение, задающее некоторую функцию, образованное с помощью констант 0, 1, отрицания $\bar{x}$, конъюнкции $x_1 \,\&\, x_2$ и дизъюнкции $x_1 \vee x_2$, то можно произвести в нем взаимные замены 0 на 1, $\&$ на $\vee$. При этом получится выражение, задающее двойственную функцию. Это правило называют *принципом двойственности*. Например, если

$$f(x_1, x_2, x_3) = \overline{\overline{x_2 \,\&\, x_3} \vee (\bar{x}_3 \,\&\, x_1)},$$

то двойственная функция задается формулой

$$f^*(x_1, x_2, x_3) = \overline{\overline{x_2 \vee x_3} \,\&\, (\bar{x}_3 \vee x_1)}.$$

Принцип двойственности также можно применять, для получения одних тождеств из других. Например, из справедливости тождества $\overline{x_1 \,\&\, x_2} = \bar{x}_1 \vee \bar{x}_2$ следует справедливость тождества, двойственного данному: $\overline{x_1 \vee x_2} = \bar{x}_1 \,\&\, \bar{x}_2$.

1.3 Разложение булевых функций по переменным. Совершенная дизъюнктивная (конъюнктивная) нормальная форма.

Нам будет удобно использовать следующее обозначение. Пусть $x, \sigma \in \{0;1\}$. Положим

$$x^\sigma = \begin{cases} \bar{x} \text{ при } \sigma = 0, \\ x \text{ при } \sigma = 1. \end{cases}$$

Ясно, что $x^\sigma = 1$ тогда и только тогда, когда $x = \sigma$. Нетрудно также заметить, что имеет место тождество $x^\sigma = x + \sigma + 1$.

Теорема 2 *Всякая функция алгебры логики $f(x_1, \ldots, x_n)$ при любом m $(1 \le m \le n)$ может быть представлена в ф ме*

$$f(x_1, \ldots, x_m, x_{m+1}, \ldots, x_n) = \bigvee_{(\sigma_1, \ldots, \sigma_m)} x_1^{\sigma_1} \& \cdots \& x_m^{\sigma_m} \& f(\sigma_1, \ldots, \sigma_m, x_{m+1}, \ldots, x_n),$$

где дизъюнкция берется по всем наборам значений переменных $\sigma_1, \ldots, \sigma_m$.

Это представление называется *разложением* функции f по переменным $x_1, \ldots, x_m$.

Доказательство. Пусть $(\alpha_1, \ldots, \alpha_n)$ — произвольный набор значений переменных. Проверим, что левая и правая части доказываемого соотношения принимают одинаковые значения на этом наборе. Конъюнкции в правой части равны нулю во всех случаях, кроме $\sigma_i = \alpha_i$ $(1 \le i \le m)$. Отсюда ясно, что

$$\bigvee_{(\sigma_1, \ldots, \sigma_m)} \alpha_1^{\sigma_1} \& \cdot \cdot \& \alpha_m^{\sigma_m} \& f(\sigma_1, \ldots, \sigma_m, \alpha_{m+1}, \ldots, \alpha_n) =$$

$$= \alpha_1^{\alpha_1} \& \cdots \& \alpha_m^{\alpha_m} \& f(\alpha_1, \ldots, \alpha_m, \alpha_{m+1}, \ldots, \alpha_n) = f(\alpha_1, \ldots, \alpha_n).$$

В качестве следствия получаем два случая разложений.

1) Разложение по переменной x_n:

$$f(x_1, \ldots, x_{n-1}, x_n) = x_n \& f(x_1, \ldots, x_{n-1}, 1) \vee \bar{x}_n \& f(x_1, \ldots, x_{n-1}, 0).$$

2) Разложение по всем n переменным:

$$f(x_1, \ldots, x_n) = \bigvee_{(\sigma_1, \ldots, \sigma_n)} x_1^{\sigma_1} \& \cdots \& x_n^{\sigma_n} \& f(\sigma_1, \ldots, \sigma_n).$$

Удаляя члены дизъюнкции, для которых $f(\sigma_1, \ldots, \sigma_n) = 0$, мы получим разложение вида

$$f(x_1,\dots,x_n) = \bigvee_{\substack{(\sigma_1,\dots,\sigma_n)\\ f(\sigma_1,\dots,\sigma_n)=1}} x_1^{\sigma_1}\,\&\cdots\&\,x_n^{\sigma_n}. \qquad (1)$$

Разложение (1) называется *совершенной дизъюнктивной нормальной формой* (сокращенно СДНФ). (Заметим, что если функция f есть константа 0, то в правой части (1) стоит пустая дизъюнкция, равная 0 в силу соглашения, принятого выше.)

Следствие 1 *Каждую функцию алгебры логики можно выразить через конъюнкцию, дизъюнкцию и отрицание.*

Для доказательства заметим, что функция $f(x_1,\dots,x_n) = 0$ может быть выражена как $f(x_1,\dots,x_n) = x_1\,\&\,\bar{x}_1$, а для ненулевой функции формула (1) дает искомое выражение.

Пример 2 Представить в виде СДНФ функцию $x_1 \to x_2$.

Из таблицы видно, что имеется ровно три набора: $(0,0)$, $(0,1)$ и $(1,1)$, на которых функция $x_1 \to x_2$ равна 1. Следовательно,

$$(x_1 \to x_2) = x_1^0\,\&\,x_2^0 \vee x_1^0\,\&\,x_2^1 \vee x_1^1\,\&\,x_2^1 = \bar{x}_1\,\&\,\bar{x}_2 \vee \bar{x}_1\,\&\,x_2 \vee x_1\,\&\,x_2.$$

Принцип двойственности позволяет получить еще одно представление для булевых функций. Пусть дана функция $\in \mathcal{P}_2$. Представим функцию f^* в виде СДНФ:

$$f^*(x_1,\dots,x_n) = \bigvee_{\substack{(\sigma_1,\dots,\sigma_n)\\ f^*(\sigma_1,\dots,\sigma_n)=1}} x_1^{\sigma_1}\,\&\cdots\&\,x_n^{\sigma_n}.$$

По принципу двойственности,

$$f^{**}(x_1,\dots,x_n) = \bigwedge_{\substack{(\sigma_1,\dots,\sigma_n)\\ f^*(\sigma_1,\dots,\sigma_n)=1}} x_1^{\sigma_1} \vee \cdots \vee x_n^{\sigma_n}.$$

Учитывая, что $f^{**} = f$, получаем

$$f(x_1,\dots,x_n) = \bigwedge_{\substack{(\sigma_1,\dots,\sigma_n)\\ f^*(\sigma_1,\dots,\sigma_n)=1}} x_1^{\sigma_1} \vee \cdots \vee x_n^{\sigma_n} = \bigwedge_{\substack{(\sigma_1,\dots,\sigma_n)\\ f(\sigma_1,\dots,\sigma_n)=0}} x_1^{\sigma_1} \vee \cdots \vee x_n^{\sigma_n}.$$

Отсюда легко следует, что f имеет разложение вида

$$f(x_1,\dots,x_n)=\bigwedge_{\substack{(\sigma_1,\dots,\sigma_n)\\ f(\sigma_1,\dots,\sigma_n)=0}} x_1^{\bar\sigma_1}\,\&\cdots\&\,x_n^{\bar\sigma_n}, \qquad (2)$$

называемое *совершенной конъюнктивной нормальной формой* (сокращенно СКНФ). (Заметим, что если функция f есть константа 1, то в правой части (2) стоит пустая конъюнкция, равная 1 в силу принятого выше соглашения.)

Пример 3 Представить в виде СКНФ функцию $f(x_1,x_2,x_3)$ из примера 1.

Данная функция (см. таблицу) принимает значение 0 на трех наборах: $(0,0,1)$, $(1,0,0)$ и $(1,1,1)$. Согласно (2), имеем:

$$f(x_1,x_2,x_3)=(x_1\vee x_2\vee\bar x_3)(\bar x_1\vee x_2\vee x_3)(\bar x_1\vee\bar x_2\vee\bar x_3).$$

Одну и ту же функцию можно представлять как в виде СДНФ, так и в виде СКНФ. Если функция "часто" принимает значение 0 и "редко" принимает значение 1, то ее удобнее представлять в виде СДНФ, а в противном случае — в виде СКНФ. Например, импликация $x_1\to x_2$ всего один раз принимает значение 0 и потому ее удобно представлять в виде СКНФ: $(x_1\to x_2)=\bar x_1\vee x_2$. Ее представление в виде СДНФ более громоздко, так как содержит три дизъюнктивных члена (см. пример 2).

1.4 Суперпозиция булевых функций. Замкнутые классы.

Пусть $f(y_1,\dots,y_s)$ — некоторая булева функция. Пусть также имеется s булевых функций $f_1,\dots,f_s$. Можно считать, что все они зависят от одного и того же набора переменных $x_1,\dots,x_n$. Тогда можно образовать новую функцию

$$\Phi(x_1,\dots,x_n)=f(f_1(x_1,\dots,x_n),\dots,f_s(x_1,\dots,x_n)),$$

называемую *суперпозицией* функции f относительно $f_1,\dots,f_s$.

Пусть дан некоторый класс $\mathcal{K}$ булевых функций, т.е. $\mathcal{K}\subseteq\mathcal{P}_2$. Мы хотим дать точное описание свойства "функция f выражается через функции из класса $\mathcal{K}$". Для любого класса $\mathcal{K}$ функций рассмотрим все такие функции Φ, которые являются суперпозициями некоторой функции $f\in\mathcal{K}$ относительно функций $f_1,\dots,f_s$, где каждая из последних либо есть одна из переменных, либо принадлежит $\mathcal{K}$. При этом получается новый класс функций, который, очевидно, содержит $\mathcal{K}$. Мы обозначим этот класс через $\mathcal{K}'$ и назовем его *простым расширением* класса $\mathcal{K}$. При этом может, конечно, оказаться, что $\mathcal{K}'=\mathcal{K}$. В этом случае класс $\mathcal{K}$ будет называться *замкнутым*. Ввиду важности этого понятия, дадим его определение еще раз более подробно.

Определение 4 Пусть $\mathcal{K}$ — некоторый класс функций, т.е. $\mathcal{K}\subseteq\mathcal{P}_2$. Он называется *замкнутым*, если для любой функции $f(x_1,\dots,x_s)\in\mathcal{K}$ и любых функций $f_1,\dots,f_s$, каждая из которых есть либо одна из переменных, либо функция из $\mathcal{K}$, суперпозиция $f(f_1,\dots,f_s)$ также принадлежит $\mathcal{K}$.

Иными словами, класс $\mathcal{K}$ замкнут, если он замкнут относительно суперпозиции, т.е. при подстановке в функции из $\mathcal{K}$ переменных или функций из $\mathcal{K}$ вновь получаются функции из $\mathcal{K}$.

Пусть теперь $\mathcal{K} \subseteq \mathcal{P}_2$ — произвольный класс функций. Легко понять, что его простое расширение $\mathcal{K}'$ также не о язано, вообще говоря, быть замкнутым. В связи с этим определим последовательность классов функций $\mathcal{K}_i$ $(i \geq 0)$ при помощи индукции по i. Положим $\mathcal{K}_0 = \mathcal{K}$, и пусть $\mathcal{K}_{i+1} = \mathcal{K}_i'$ для всех $i \geq 0$. Мы имеем вложенную последовательность классов

$$\mathcal{K}_0 \subseteq \mathcal{K}_1 \subseteq \cdots \subseteq \mathcal{K}_i \subseteq \cdots.$$

Введем обозначение $[\mathcal{K}] = \bigcup\limits_{i \geq 0} \mathcal{K}_i$ и назовем этот класс *замыканием* класса $\mathcal{K}$.

Лемма 1 *Для любого класса $\mathcal{K}$ класс $[\mathcal{K}]$ замкнут. Более того, он является наименьшим замкнутым классом, содержащим $\mathcal{K}$. Класс $\mathcal{K}$ замкнут тогда и только тогда, когда он совпадает со своим замыканием, т.е. $\mathcal{K} = [\mathcal{K}]$.*

Доказательство. Проверим вначале, что $[\mathcal{K}]$ замкнут. Пусть $f \in [\mathcal{K}]$ — функция s переменных, и пусть $f_1, \ldots, f_s$ — либо переменные, либо функции из $[\mathcal{K}]$. Каждая из функций, принадлежащая $[\mathcal{K}]$, лежит в $\mathcal{K}_i$ при некотором i. Так как взятых нами функций конечное число, найдется т кое i, что все функции $f, f_1, \ldots, f_s$, не являющиеся переменными, принадлежат $\mathcal{K}_i$. Следовательно, суперпозиция $\Phi = f(f_1, \ldots, f_s)$ принадлежит классу $\mathcal{K}_i' = \mathcal{K}_{i+1}$, откуда $\Phi \in [\mathcal{K}]$. Таким образом, взятие суперпозиции не выводит за пределы $[\mathcal{K}]$, т.е. $[\mathcal{K}]$ замкнут.

Для доказательства того, что $[\mathcal{K}]$ — наименьший замкнутый класс, содержащий $\mathcal{K}$, достаточно заметить, что всякий замкнутый класс, содержащий $\mathcal{K} = \mathcal{K}_0$, должен содержать и $\mathcal{K}_1 = \mathcal{K}_0'$, а потому и $\mathcal{K}_2 = \mathcal{K}_1'$ и т. д. Это значит, что такой класс должен содержать и $[\mathcal{K}]$.

Если $\mathcal{K} = [\mathcal{K}]$, то $\mathcal{K}$ замкнут. Обратно, если $\mathcal{K}$ замкнут, то $\mathcal{K} = \mathcal{K}'$, т.е. $\mathcal{K}_0 = \mathcal{K}_1$. Отсюда видно, что все классы $\mathcal{K}_i$ совпадают с $\mathcal{K}$, что сразу же дает $\mathcal{K} = [\mathcal{K}]$.

Лемма доказана.

Теперь нетрудно дать точное определение свойства "f можно выразить через функции класса $\mathcal{K}$". А именно, это значит, что $f \in [\mathcal{K}]$.

1.5 Полные системы.

Нас будут интересовать такие системы функций, через которые можно выразить любую булеву функцию. Такие системы называются полными. В этом смысле полной будет система функций, состоящая из конъюнкции, дизъюнкции и отрицания, как утверждает следствие 1.

Определение 5 *Система $\mathcal{K}$ функций алгебры логики называется* полной, *если через функции этой системы можно выразить любую функцию алгебры логики, иными словами, если $[\mathcal{K}] = \mathcal{P}_2$.*

Приведем вначале пример системы функций, которая не будет полной. Пусть $\mathcal{K} = \{\&, \vee\}$ — система, состоящая из конъюнкции и дизъюнкции. Она не будет полной, так как любая функция, выражающаяся через конъюнкцию и дизъюнкцию, принимает значение 0, если все переменные равны нулю. В частности, отрицание нельзя выразить через функции из $\mathcal{K}$, т.е. $\mathcal{K}$ не полна.

Теорема 3 *Следующие системы функций являются полными:*

1) $\{\bar{x}, x_1 \& x_2, x_1 \vee x_2\}$,
2) $\{\bar{x}, x_1 \& x_2\}$,
3) $\{\bar{x}, x_1 \vee x_2\}$,
4) $\{x_1 \mid x_2\}$,
5) $\{1, x_1 x_2, x_1 + x_2\}$.

Доказательство. Пункт 1 доказан в следствии 1. Используя законы де Моргана и закон двойного отрицания, мы получаем, что

$$x_1 \vee x_2 = \overline{\bar{x}_1 \& \bar{x}_2}$$

(дизъюнкция выражается через конъюнкцию и отрицание), что сводит пункт 2 к пункту 1), а также

$$x_1 \& x_2 = \overline{\bar{x}_1 \vee \bar{x}_2}$$

(конъюнкция выражается через дизъюнкцию и отрицание), что доказывает пункт 3. Для доказательства пункта 4 заметим, что штрих Шеффера есть отрицание конъюнкции:

$$x_1 \mid x_2 = \overline{x_1 \& x_2}.$$

Следовательно, через штрих Шеффера можно выразить отрицание ($\bar{x} = \overline{x \& x} = x \mid x$), а затем через отрицание и штрих Шеффера выразить конъюнкцию ($x_1 \& x_2 = \overline{x \mid x}$). Согласно пункту 2, любая функция может быть выражена через конъюнкцию и отрицание, т.е. мы доказали пункт 4.

Так как $\bar{x} = x + 1$, отрицание можно выразить через константу 1 и сложение по модулю 2. Поэтому пункт 5 следует теперь из пункта 2.

Теорема полностью доказана.

Проанализируем отдельно систему из пункта 5). Любая функция из $\mathcal{P}_2$ выражается через константу 1, сложение и умножение по модулю 2. Это значит, что любую булеву функцию можно представить многочленом над полем $\{0, 1\}$ из двух элементов. Более того, ввиду тождества $x^2 = xx = x$ мы можем считать, что одночлены не содержат переменных в степени выше первой. Любой такой многочлен от переменных x_1, x_2, ..., x_n есть линейная комбинация одночленов вида $x_{i_1} \dots x_{i_k}$, где $1 \leq i_1 < \dots < i_k \leq n$. При этом допускается случай $k = 0$, приводящий к "пустому" одночлену 1. Такие многочлены называются *полиномами Жегалкина*. Общий вид такого полинома:

$$\sum_{(i_1, \dots, i_k)} a_{i_1 \dots i_k} x_{i_1} \dots x_{i_k},$$

где $a_{i_1 \ldots i_k}$ равно 0 или 1, а суммирование ведется по всем наборам индексов с условием $1 \le i_1 < \cdots < i_k \le n$ $(k \ge 0)$.

Теорема 4 *Любая булева функция $f(x_1, \ldots, x_n)$ от n переменных представима в виде полинома Жегалкина от $x_1, \ldots, x_n$, причем такое представление единственно.*

Доказательство. Факт представимости булевых функций полиномами Жегалкина доказан выше. Осталось показать единственность. Сравним число булевых функций от n переменных и число полиномов Жегалкина от n переменных. Первое число равно 2^{2^n} по теореме 1. Так как коэффициент при одночлене принимает два значения, второе число равно 2^m, где m — количество одночленов. Ясно, что одночлен однозначно задается подмножеством $\{i_1, \ldots, i_n\}$ множества $\{1, 2, \ldots, n\}$. Поэтому $m = 2^n$, и число полиномов Жегалкина также равно 2^{2^n}. Отсюда ясно, что представление функции полиномом Жегалкина единственно.

Пример 4 Представить дизъюнкцию $x_1 \vee x_2$ полиномом Жегалкина.

Имеем: $x_1 \vee x_2 = \overline{x_1 \,\&\, x_2} = \bar{x}_1 \bar{x}_2 + 1 = (x_1 + 1)(x_2 + 1) + 1 = x_1 + x_2 + x_1 x_2$.

1.6 Важнейшие замкнутые классы.

В предыдущем разделе мы показали, что система, состоящая из конъюнкции и дизъюнкции, не является полной. При этом мы воспользовались тем, что все функции, выражаемые через конъюнкцию и дизъюнкцию, принимают значение 0 на наборах из нулей. Этот пример можно обобщить.

Определение 6 Обозначим через T_0 класс всех булевых функций $f(x_1, \ldots, x_n)$, сохраняющих константу 0, т.е. функций, для которых выполнено равенство

$$f(0, \ldots, 0) = 0.$$

Легко видеть, что функции 0, x, $x_1 \,\&\, x_2$, $x_1 \vee x_2$, $x_1 + x_2$ принадлежат классу T_0, а функции 1, $\bar{x}$, $x_1 \to x_2$, $x_1 \mid x_2$ ему не принадлежат.

Среди всех функций от n переменных классу T_0 принадлежит ровно половина, т.е. таких функций ровно $2^{2^n - 1}$.

Класс T_0 является замкнутым. Проверим это. Пусть $f \in T_0$ — функция от s переменных, и пусть $f_1, \ldots, f_s$ — либо переменные, либо функции из T_0. Ясно, что переменные сами принадлежат T_0, поэтому $f_1, \ldots, f_s \in T_0$. Пусть $x_1, \ldots, x_n$ — все переменные, от которых зависят эти функции. Образуем суперпозицию $\Phi(x_1, \ldots, x_n) = f(f_1(x_1, \ldots, x_n), \ldots, f_s(x_1, \ldots, x_n))$. Ясно, что $\Phi(0, \ldots, 0) = f(f_1(0, \ldots, 0), \ldots, f_s(0, \ldots, 0)) = f(0, \ldots, 0) = 0$, т.е. $\Phi \in T_0$. Это доказывает замкнутость класса T_0.

Аналогично можно рассмотреть класс функций, сохраняющих единицу.

Определение 7 Обозначим через T_1 класс всех булевых функций $f(x_1, \ldots, x_n)$, сохраняющих константу 1, т.е. функций, для которых выполнено равенство

$$f(1, \ldots, 1) = 1.$$

Функции 1, x, $x_1 \& x_2$, $x_1 \vee x_2$, $x_1 \to x_2$ принадлежат T_1, а функции 0, $\bar{x}$, $x_1 + x_2$, $x_1 \,|\, x_2$ — не принадлежат.

Как и в предыдущем случае, среди всех функций от n переменных классу T_1 принадлежит ровно половина, т.е. таких функций также 2^{2^n-1}. Легко также видеть, что при $n \geq 1$ каждая четвертая функция от n переменных принадлежит и тому, и другому классу.

Замкнутость класса T_1 устанавливается точно так же, как и для T_0. Используя класс T_1, можно показать, что система, состоящая из константы 1, конъюнкции, дизъюнкции и импликации не полна, так как все эти функции принадлежат T_1, и потому через них нельзя выразить, скажем, отрицание.

Классов T_0 и T_1 не достаточно, чтобы доказывать неполноту некоторых систем. Рассмотрим, например, систему, состоящую из константы 1, отрицания и сложения по модулю 2. Как мы увидим далее, эта система не полна, однако она не содержится ни в T_0, ни в T_1. Введем еще один класс.

Определение 8 Обозначим через L класс *линейных* функций, т.е. класс всех булевых функций, представимых в виде полинома Жегалкина степени не выше первой. Если $f \in L$ — функция от переменных $x_1, \ldots, x_n$, то f имеет вид

$$f(x_1, \ldots, x_n) = a + a_1 x_1 + \cdots + a_n x_n,$$

где $a, a_1, \ldots, a_n \in \{0, 1\}$.

Из сказанного выше следует, что линейная функция от n переменных определяется $n+1$ коэффициентами, поэтому всего таких функций ровно 2^{n+1}, т.е. относительно мало.

Класс L также замкнут. Так как переменные суть линейные функции, достаточно показать, что суперпозиция линейной функции $f(y_1, \ldots, y_s)$ относительно линейных функций $f_1, \ldots, f_s$ также линейна. Это доказывается прямым вычислением. Пусть

$$f(y_1, \ldots, y_s) = a + a_1 y_1 + \cdots + a_s y_s,$$

$$f_i(x_1, \ldots, x_n) = b_i + b_{i1} x_1 + \cdots + b_{in} x_n \quad (1 \leq i \leq s),$$

где все коэффициенты a, b с индексами или без суть константы 0 или 1. Тогда суперпозиция

$$\Phi(x_1, \ldots, x_n) = f(f_1(x_1, \ldots, x_n), \ldots, f_s(x_1, \ldots, x_n)) =$$

$$= a + a_1 f_1(x_1, \ldots, x_n) + \cdots + a_s f_s(x_1, \ldots, x_n)) =$$

$$= a + a_1(b_1 + b_{11} x_1 + \cdots + b_{1n} x_n) + \cdots + a_s(b_s + b_{s1} x_1 + \cdots + b_{sn} x_n)) =$$

$$= (a + a_1b_1 + \cdots + a_sb_s) + (a_1b_{11} + \cdots + a_sb_{s1})x_1 + \cdots + (a_1b_{1n} + \cdots + a_sb_{sn})x_n$$

есть также линейная функция от $x_1, \ldots, x_n$.

Итак, замкнутость класса L доказана. Легко видеть, что константы 0 и 1, функции x, $\bar{x}$, x_1+x_2 линейны, а конъюнкция, дизъюнкция, импликация и штрих Шеффера не линейны, что устанавливается представлением этих функций в виде полиномов Жегалкина с учетом теоремы 4. Неполнота системы функций $\{1, \bar{x}, x_1 + x_2\}$ следует теперь из того, что через них нельзя выразить, например, конъюнкцию.

Рассмотрим такую систему: $\{0, 1, x_1 \,\&\, x_2, x_1 \vee x_2\}$. Она не содержится ни в одном из построенных классов T_0, T_1, L. Тем не менее, она неполна. Для доказательства рассмотрим новый класс функций.

На множестве E^n наборов длины n из нулей и единиц введем отношение *предшествования* $\preceq$, полагая $(\alpha_1, \ldots, \alpha_n) \preceq (\beta_1, \ldots, \beta_n)$, если $\alpha_1 \leq \beta_1, \ldots, \alpha_n \leq \beta_n$. Тем самым задано отношение (частичного) порядка на множестве E^n. Например, $(1, 0, 1, 0) \preceq (1, 0, 1, 1)$. Заметим, что наборы $(0, 1)$ и $(1, 0)$ не сравнимы относительно $\preceq$, т.е. введенный порядок не является линейным при $n \geq 2$.

Определение 9 Булева функция $f(x_1, \ldots, x_n)$ называется *монотонной*, если для любых двух наборов $(\alpha_1, \ldots, \alpha_n)$ и $(\beta_1, \ldots, \beta_n)$ таких, что $(\alpha_1, \ldots, \alpha_n) \preceq (\beta_1, \ldots, \beta_n)$, имеет место неравенство

$$f(\alpha_1, \ldots, \alpha_n) \leq f(\beta_1, \ldots, \beta_n).$$

Класс всех монотонных функций обозначается через M.

Ясно, что функции 0, 1, x, $x_1 \,\&\, x_2$, $x_1 \vee x_2$ монотонны, в то время как $\bar{x}$, $x_1 \to x_2$, $x_1 + x_2$, $x_1 \mid x_2$ не монотонны.

Проверим замкнутость класса M. Так как переменные суть монотонные функции, достаточно показать, что суперпозиция монотонной функции относительно монотонных функций монотонна. Если $f, f_1, \ldots, f_s$ — монотонные функции, то при увеличении некоторых значений переменных $x_1, \ldots, x_n$ значение каждой из функций $f_1, \ldots, f_s$ может только увеличиться, а потому и значение суперпозиции $\Phi = f(f_1, \ldots, f_s)$ может только увеличиться, что доказывает монотонность суперпозиции.

Теперь мы можем сделать вывод, что отрицание не выражается через константы, конъюнкцию и дизъюнкцию, т.е. рассмотренная выше система не полна.

Нам нужно рассмотреть еще один важный класс.

Определение 10 Булева функция f называется *самодвойственной*, если $f = f^*$, т.е. f удовлетворяет тождеству

$$f(\bar{x}_1, \ldots, \bar{x}_n) = \overline{f(x_1, \ldots, x_n)}. \tag{3}$$

Класс всех самодвойственных функций обозначается через S.

Самодвойственными будут функции x и $\bar{x}$. Все остальные "элементарные" функции не самодвойственны, в чем легко убедиться. Приведем менее тривиальный пример самодвойственной функции. Пусть

$$h(x_1, x_2, x_3) = x_1x_2 \vee x_1x_3 \vee x_2x_3.$$

Ясно, что h принимает значение 0, когда две или три переменных принимают значение 0 и равна 1 в противном случае. Поэтому можно сказать, что h принимает то значение, которое принимает большинство переменных. По этой причине h называют *функцией голосования*. Проверим теперь, что h самодвойственна. Используя принцип двойственности, имеем:

$$h^*(x_1, x_2, x_3) = (x_1 \vee x_2)(x_1 \vee x_3)(x_2 \vee x_3) = (x_1 \vee x_1x_2 \vee x_1x_3 \vee x_2x_3)(x_2 \vee x_3) =$$

$$= x_1x_2 \vee x_1x_3 \vee x_2x_3 = h(x_1, x_2, x_3).$$

Назовем два набора из нулей и единиц *противоположными*, если они имеют вид $(\alpha_1, \ldots, \alpha_n)$ и $(\bar{\alpha}_1, \ldots, \bar{\alpha}_n)$. Функция будет самодвойственной в том и только в том случае, когда она принимает противоположные значения на противоположных наборах. Поскольку при $n \geq 1$ все 2^n наборов разбиваются на 2^{n-1} пар противоположных, то самодвойственная функция полностью определяется своими значениями на 2^{n-1} наборах. Следовательно, при $n \geq 1$ имеется ровно $2^{2^{n-1}}$ самодвойственных функций от n переменных, что составляет ровно $\sqrt{p_2(n)}$, т.е. корень квадратный от числа всех функций от n переменных.

Легко установить замкнутость класса S. Так как переменные суть функции самодвойственные, достаточно проверить, что суперпозиция самодвойственных функций самодвойственна. Если f, f_1, $\ldots$, f_s — самодвойственные функции, то при замене всех значений переменных x_1, $\ldots$, x_n на противоположные значение каждой из функций f_1, $\ldots$, f_s также изменится на противоположное, а потому и значение суперпозиции $\Phi = f(f_1, \ldots, f_s)$ изменится на противоположное. Это доказывает самодвойственность суперпозиции самодвойственных функций.

Система, состоящая из отрицания и функции голосования, является неполной, так как через них нельзя выразить несамодвойственную функцию (например, конъюнкцию). Однако легко усмотреть, что эта система не содержится ни в одном из классов T_0, T_1, L, M, поэтому для доказательства ее неполноты нам пришлось использовать класс S.

Итак, мы ввели пять важных замкнутых классов. Подытожим сказанное следующей теоремой.

Теорема 5 *Классы T_0, T_1, L, M, S замкнуты, не совпадают с $\mathcal{P}_2$, и ни один из них не содержится в другом.*

Замкнутость всех пяти классов нами доказана выше. Для каждого из классов приведены примеры функций, не принадлежащих этому классу, поэтому все классы — собственные. Более того, ни один из классов не содержится в другом. Этот

вывод можно сделать на основе рассмотренных примеров, однако мы объединим полученную информацию в следующую таблицу.

	0	1	x	$\bar{x}$	$\&$	$\vee$	$\to$	$\oplus$	$\mid$	h
T_0	+	−	+	−	+	+	−	+	−	+
T_1	−	+	+	−	+	+	+	−	−	+
L	+	+	+	+	−	−	−	+	−	−
M	+	+	+	−	+	+	−	−	−	+
S	−	−	+	+	−	−	−	−	−	+

Строки таблицы соответствуют пяти классам: T_0, T_1, L, M, S. Столбцы таблицы соответствуют десяти функциям: константам 0 и 1, тождественной функции, отрицанию, конъюнкции, дизъюнкции, импликации, сложению по модулю 2, штриху Шеффера и функции голосования. На пересечении строки и столбца стоит знак "+" или "−" в зависимости от того, содержит ли класс данную функцию. Тот факт, что ни один из пяти классов не содержится в другом, можно непосредственно усмотреть из таблицы. Скажем, тот факт, что T_1 не содержится в M следует из того, что импликация принадлежит T_1, но не принадлежит M. Функция голосования h дает пример нелинейной самодвойственной функции, откуда следует, что S не содержится в L и т. п.

1.7 Предполные классы. Теорема Поста.

Целью данного раздела является доказательство критерия полноты системы функций. Из теоремы 5 следует, что если система функций содержится полностью в одном из рассмотренных пяти классов, то она неполна. Оказывается, верно и обратное. Это дает простой и удобный критерий проверки системы булевых функций на полноту. Следующая теорема и ее следствия доказаны американским математиком Э. Постом (E.Post).

Теорема 6 *Для того, чтобы система функций $\mathcal{K}$ была полной, необходимо и достаточно, чтобы она не содержалась полностью ни в одном из пяти замкнутых классов T_0, T_1, L, M, S.*

Доказательство. Каждый из пяти классов замкнут. Если система $\mathcal{K}$ содержится в одном из них, то ее замыкание содержится в данном классе, и потому не совпадает с $\mathcal{P}_2$.

Обратное утверждение будет доказано в несколько этапов. При этом нам потребуется несколько вспомогательных утверждений. Итак, предположим, что система функций $\mathcal{K}$ не содержится целиком ни в одном из пяти классов. Тогда можно указать функции g_1, g_2, g_3, g_4 и g_5 из класса $\mathcal{K}$, не принадлежащие классам T_0, T_1, L, M и S, соответственно. (Заметим, что среди выбранных пяти функций возможны

совпадения.) Наша цель — выразить произвольную булеву функцию через g_1, g_2, g_3, g_4, g_5.

1. Для начала покажем, что константы можно выразить через функции $g_1 \notin T_0$, $g_2 \notin T_1$, $g_5 \notin S$. Рассмотрим функцию g_1. Возможны два случая.

а) $g_1(1,\dots,1) = 1$. При этом функция $\phi(x) = g_1(x,\dots,x)$ есть константа 1, так как $\phi(0) = g_1(0,\dots,0) = 1$ ввиду $g_1 \notin T_0$, а $\phi(1) = g_1(1,\dots,1) = 1$ по предположению. Константу 0 можно теперь выразить через 1 и g_2: $g_2(1,\dots,1) = 0$.

б) $g_1(1,\dots,1) = 0$. Тогда мы можем выразить отрицание через g_1: $\phi(x) = g_1(x,\dots,x) = \bar{x}$, так как $\phi(0) = g_1(0,\dots,0) = 1$, $\phi(1) = g_1(1,\dots,1) = 0$. Нам потребуется следующая лемма.

Лемма 2 *Если $f(x_1,\dots,x_n) \notin S$, то из нее путем подстановки функций x и $\bar{x}$ можно получить константу.*

Доказательство. Так как f не самодвойственна, она не удовлетворяет тождеству (3). Поэтому найдутся два противоположных набора $(\alpha_1,\dots,\alpha_n)$ и $(\bar{\alpha}_1,\dots,\bar{\alpha}_n)$, на которых функция принимает одно и то же значение: $f(\alpha_1,\dots,\alpha_n) = f(\bar{\alpha}_1,\dots,\bar{\alpha}_n)$. Положим $\phi_i(x) = x^{\alpha_i}$, и пусть $\phi(x) = f(\phi_1(x),\dots,\phi_n(x))$. Тогда

$$\phi(0) = f(\phi_1(0),\dots,\phi_n(0)) = f(0^{\alpha_1},\dots,0^{\alpha_n}) = f(\bar{\alpha}_1,\dots,\bar{\alpha}_n) =$$

$$= f(\alpha_1,\dots,\alpha_n) = f(1^{\alpha_1},\dots,1^{\alpha_n}) = f(\phi_1(1),\dots,\phi_n(1)) = \phi(1),$$

т.е. $\phi(x)$ — искомая константа.

Лемма доказана.

Применяя лемму 2, мы выразим одну из констант через $g_5 \notin S$ и отрицание. Другая константа выражается как отрицание полученной.

Итак, в обоих случаях мы получаем константы 0 и 1.

2. Теперь покажем, как выразить отрицание через константы и немонотонную функцию g_4.

Лемма 3 *Если $f(x_1,\dots,x_n) \notin M$, то из нее путем подстановки констант и переменной x можно получить функцию $\bar{x}$.*

Доказательство. Будем называть два набора длиной n из нулей и единиц *соседними*, если они отличаются ровно одной координатой. Так как функция f не монотонна, найдутся два набора $\tilde{\alpha} = (\alpha_1,\dots,\alpha_n)$ и $\tilde{\beta} = (\beta_1,\dots,\beta_n)$ такие, что $\tilde{\alpha} \preceq \tilde{\beta}$, но $f(\tilde{\alpha}) = 1 > 0 = f(\tilde{\beta})$. Ясно, что $\tilde{\alpha} \neq \tilde{\beta}$, т.е. эти наборы различаются в нескольких координатах, причем эти координаты равны 0 у $\tilde{\alpha}$ и равны 1 у $\tilde{\beta}$. Меняя последовательно нули на единицы, мы получаем цепочку наборов $\tilde{\alpha} = \tilde{\gamma}_0$, $\tilde{\gamma}_1$, …, $\tilde{\gamma}_t = \tilde{\beta}$ $(t \geq 1)$, в которой стоящие рядом наборы различаются ровно в одной координате, т.е. являются соседними. Поэтому найдется такая пара соседних наборов $(\gamma_1,\dots,\gamma_{i-1},0,\gamma_{i+1},\dots,\gamma_n)$ и $(\gamma_1,\dots,\gamma_{i-1},1,\gamma_{i+1},\dots,\gamma_n)$, что на первом из них функция f равна 1, а на втором — нулю. Функция $\phi(x) = f(\gamma_1,\dots,\gamma_{i-1},x,\gamma_{i+1},\dots,\gamma_n)$, получаемая из f подстановкой констант, очевидно, есть $\bar{x}$.

Лемма доказана.

Итак, в силу леммы 3 мы можем выразить отрицание через функции нашей системы.

3. На последнем этапе мы выразим конъюнкцию через константы, отрицание и нелинейную функцию g_3. Это позволит сделать следующая

Лемма 4 *Если $f(x_1, \ldots, x_n) \notin L$, то из нее путем путем подстановки констант и функций x, $\bar{x}$, а также, возможно, путем навешивания отрицания над f, можно получить функцию $x_1 \& x_2$.*

Доказательство. Представим f полиномом Жегалкина, используя теорему 4:

$$f(x_1, \ldots, x_n) = \sum_{(i_1, \ldots, i_k)} a_{i_1 \ldots i_k} x_{i_1} \ldots x_{i_k}.$$

Так как f не линейна, ее полином Жегалкина содержит одночлен степени выше первой. Без ограничения общности будем считать, что в этот одночлен входят переменные x_1 и x_2. Тогда полином можно представить в виде $f(x_1, \ldots, x_n) = x_1 x_2 \phi_1(x_3, \ldots, x_n) + x_1 \phi_2(x_3, \ldots, x_n) + x_2 \phi_3(x_3, \ldots, x_n) + \phi_4(x_3, \ldots, x_n)$, причем полином ϕ_1 — ненулевой. Так как тождественный нуль задается нулевым полиномом Жегалкина, должны найтись константы α_3, ..., α_n такие, что $\phi_1(\alpha_3, \ldots, \alpha_n) = 1$. Тогда, подставляя эти константы в f, мы получим функцию $\phi(x_1, x_2) = f(x_1, x_2, \alpha_3, \ldots, \alpha_n) = x_1 x_2 + \beta x_1 + \gamma x_2 + \delta$, где β, γ, δ — константы, равные 0 или 1. Легко видеть, что $\phi(x_1 + \gamma, x_2 + \beta) = (x_1 + \gamma)(x_2 + \beta) + \beta(x_1 + \gamma) + \gamma(x_2 + \beta) + \delta = x_1 x_2 + \beta\gamma + \delta$. При этом мы подставили в качестве аргументов функции переменные или их отрицания и получили конъюнкцию или ее отрицание. В последнем случае остается навесить отрицание над f. Итак, мы можем выразить конъюнкцию.

Лемма доказана.

Таким образом, мы выразили отрицание и конъюнкцию через функции системы, а этого достаточно для полноты си емы ввиду теоремы 3.

Теорема полностью доказана.

Пример 5 Система функций $\phi_1 = 0$, $\phi_2 = 1$, $\phi_3 = x_1 x_2$, $\phi_4 = x_1 + x_2 + x_3$ полна.

Действительно, $\phi_2 \notin T_0$, $\phi_1 \notin T_1$, $\phi_3 \notin L$, $\phi_4 \notin M$, $\phi_1 \notin S$, и по теореме Поста данная система полна. Помимо прочего, теорема дает конструктивный способ выражения любой функции через функции системы.

Удаление любой из четырех функций приводит к неполной системе. В самом деле,

$$\{\phi_2, \phi_3, \phi_4\} \subset T_1, \qquad \{\phi_1, \phi_3, \phi_4\} \subset T_0,$$

$$\{\phi_1, \phi_2, \phi_4\} \subset L, \qquad \{\phi_1, \phi_2, \phi_3\} \subset M.$$

Теорема Поста имеет ряд интересных следствий. Пусть имеется замкнутый класс $\mathcal{K}$, не совпадающий с $\mathcal{P}_2$. Из теоремы Поста следует, что $\mathcal{K}$ содержится в одном из пяти классов.

Определение 11 Класс функций $\mathcal{K}$ называется *предполным*, если он не является полным, но для любой функции $f \in \mathcal{P}_2$, не принадлежащей $\mathcal{K}$, класс $\mathcal{K} \cup \{f\}$ является полным.

Следствие 2 *В алгебре логики имеется в точности пять предполных классов: T_0, T_1, L, M и S.*

Для доказательства заметим, что ни один из пяти классов не содержится в другом по теореме 5. Если $\mathcal{K}$ — один из пяти классов, а $f \notin \mathcal{K}$, то объединение $\mathcal{K} \cup \{f\}$ не содержится ни в одном из пяти классов, т.е. является полной системой ввиду теоремы Поста. Итак, каждый из пяти классов — предполный. Обратно, пусть $\mathcal{K}$ — предполный класс. По теореме Поста он должен содержаться в одном из пяти классов. Поэтому $\mathcal{K}$ совпадает с этим классом — иначе к $\mathcal{K}$ можно было бы добавить функцию, ему не принадлежащую, в результате чего получился бы заведомо неполный класс.

Следствие 3 *Из любой полной системы можно выделить полную подсистему, состоящую не более, чем из четырех функций.*

Доказательство. Если бы речь шла о пяти функциях, то это прямо следовало бы из теоремы Поста. Действительно, в данной системе можно указать функции $g_1 \notin T_0$, $g_2 \notin T_1$, $g_3 \notin L$, $g_4 \notin M$, $g_5 \notin S$. Они образуют полную подсистему не более, чем из пяти функций. Как и в доказательстве теоремы Поста, рассмотрим два случая.

1) $g_1(1, \ldots, 1) = 1$. При этом на противоположных наборах $(0, \ldots, 0)$ и $(1, \ldots, 1)$ функция g_1 принимает одно и то же значение 1. Следовательно, $g_1 \notin S$, и функции g_1, g_2, g_3, g_4 образуют полную систему.

2) $g_1(1, \ldots, 1) = 0$. При этом $g_1 \notin T_1$, т.е. функцию g_2 можно не брать. Более того, $g_1(0, \ldots, 0) = 1 > 0 = g_1(1, \ldots, 1)$, т.е. к тому же g_1 еще и не монотонна. Функции g_1, g_3, g_5 образуют полную систему.

Данное следствие нельзя усилить, как показывает пример 5.

В заключение скажем несколько слов о результатах Поста, касающихся описания замкнутых классов. Система $\mathcal{S}$ функций из замкнутого класса $\mathcal{K}$ называется его *базисом*, если замыкание системы $\mathcal{S}$ есть $\mathcal{K}$, но замыкание любого собственного подмножества $\mathcal{S}$ не равно $\mathcal{K}$. Скажем, можно доказать, что система функций $\{0, 1, \&, \vee\}$ образует базис класса монотонных функций. Пост получил такие результаты:

1. Всякий замкнутый класс из $\mathcal{P}_2$ имеет конечный базис, состоящий не более, чем из четырех элементов.

2. Мощность множества замкнутых классов в $\mathcal{P}_2$ — счетная.

2 k-значная логика

2.1 Функции k-значной логики.

Пусть $k \geq 2$, $E_k = \{0, 1, \ldots, k-1\}$.

Определение 12 *Функция k-значной логики* от n переменных — это отображение $f: E_k^n \to E_k$.

При $k = 2$ мы получаем булевы функции. Таким образом, функции k-значной логики можно рассматривать как обобщение булевых функций. Как и в случае булевых функций, функции k-значной логики можно задавать с помощью таблиц, например, таблица

x_1	x_2	$f(x_1, x_2)$
0	0	2
0	1	0
0	2	1
1	0	1
1	1	0
1	2	1
2	0	2
2	1	2
2	2	1

задает некоторую функцию трехзначной логики от двух переменных.

Множество всех функций k-значной логики будет обозначаться через $\mathcal{P}_k$; через $p_k(n)$ мы обозначаем число всех функций k-значной логики от n переменных. По аналогии с теоремой 1, имеет место

Теорема 7

$$p_k(n) = k^{k^n}.$$

Доказательство. Число упорядоченных наборов длиной n из элементов множества E_k равно k^n. Функция от n переменных, в свою очередь, задается упорядоченным набором длиной k^n из элементов того же множества, откуда число функций равно k^{k^n}.

Так же, как и в случае булевых функций, вводятся понятия существенной и несущественной переменной. Функции далее рассматриваются с точностью до фиктивных переменных, как и в предыдущей главе.

Приведем примеры "элементарных" функций k-значной логики.

1) $\bar{x} = x + 1 \pmod{k}$. Это есть одно из обобщений отрицания.

2) $\sim x = k-1-x$. Это еще одно обобщение отрицания — *отрицание Лукасевича*.
3) $\min(x_1, x_2)$ — обобщение конъюнкции. Другое обозначение: $x_1 \& x_2$.
4) $x_1 x_2 \pmod k$ — другое обобщение конъюнкции.
5) $\max(x_1, x_2)$ — обобщение дизъюнкции. Другое обозначение: $x_1 \vee x_2$.
6) $x_1 + x_2 \pmod k$ — сложение по модулю k.

Далее, рассматривая функции k-значной логики, мы будем использовать знак $+$ для обозначения сложения по модулю k.

Введем также обозначение x^σ для $x, \sigma \in E_k$, полагая

$$x^\sigma = \begin{cases} k-1 & \text{при } x = \sigma, \\ 0 & \text{при } x \neq \sigma. \end{cases}$$

Нетрудно видеть, что при $k = 2$ данное определение совпадает с тем, которое рассматривалось для булевых функций. Заметим, что x^σ можно рассматривать как функцию от двух переменных, а при фиксированном $\sigma \in E_k$ — как функцию от x. Нам потребуются обозначения

$$I_0(x),\ I_1(x),\ \ldots, I_{k-1}(x),$$

где $I_\sigma(x) = x^\sigma$ при $\sigma \in \{0, 1, \ldots, k-1\}$.

Отметим несколько свойств "элементарных" функций. Во-первых, операции $\min$, $\max$, сложение по модулю k, умножение по модулю k ассоциативны и коммутативны. Далее, аналог закона двойного отрицания выполнен для отрицания Лукасевича: $\sim(\sim x) = x$, но не выполнен для отрицания $\bar{x}$ при $k \geq 3$: $\bar{\bar{x}} \neq x$ ни при каком x.

Также справедливы аналоги законов де Моргана:

$$\sim \min(x_1, x_2) = \max(\sim x_1, \sim x_2),$$

$$\sim \max(x_1, x_2) = \min(\sim x_1, \sim x_2),$$

но они не имеют места, если вместо отрицания Лукасевича взять отрицание $\bar{x}$.

С учетом ассоциативного закона, условимся использовать обозначения

$$\bigwedge_{i=1}^{m} x_i = \min_{1 \leq i \leq m} x_i = x_1 \& x_2 \& \cdots \& x_m,$$

$$\bigvee_{i=1}^{m} x_i = \max_{1 \leq i \leq m} x_i = x_1 \vee x_2 \vee \cdots \vee x_m,$$

$$\sum_{i=1}^{m} x_i = x_1 + x_2 + \cdots + x_m.$$

Как и ранее, эти обозначения имеют смысл для всех $m \geq 1$. При $m = 0$ считаем, что первое выражение равно $k-1$, а второе и третье 0.

Приведем теперь тождество, представляющее собой аналог СДНФ.

Теорема 8 *Для любой функции k-значной логики $f(x_1, \ldots, x_n)$ имеет место тождество*

$$f(x_1, \ldots, x_n) = \bigvee_{(\sigma_1, \ldots, \sigma_n)} x_1^{\sigma_1} \,\&\, \cdots \,\&\, x_n^{\sigma_n} \,\&\, f(\sigma_1, \ldots, \sigma_n).$$

Доказательство. Доказательство представляет собой прямую проверку равенства правой и левой части на наборе значений переменных $(\alpha_1, \ldots, \alpha_n)$. Заметим, что все дизъюнктивные члены в правой части равны нулю, за исключением, может быть, одного члена, для которого $\sigma_i = \alpha_i$ при всех i от 1 до n. Это следует из того, что конъюнкция $\&$ представляет собой минимум, и если хотя бы один из членов конъюнкции обратился в ноль (что имеет место, если $\sigma_i \neq \alpha_i$ при некотором i), то и вся конъюнкция обратилась в ноль. Так как дизъюнкция есть максимум, можно игнорировать нулевые дизъюнктивные члены. В результате в правой части остается один дизъюнктивный член: $\alpha_1^{\sigma_1} \,\&\, \cdots \,\&\, \alpha_n^{\sigma_n} \,\&\, f(\alpha_1, \ldots, \alpha_n)$, который равен $f(\alpha_1, \ldots, \alpha_n)$, так как конъюнкция есть минимум, а все остальные члены данной конъюнкции принимают максимальное значение $k-1$. Таким образом, значения правой и левой частей равны.

Теорема доказана.

2.2 Полные системы в k-значной логике.

Для любого класса $\mathcal{K} \subseteq \mathcal{P}_k$ можно определить его замыкание $[\mathcal{K}]$ таким же образом, как и для случая булевых функций. Аналогичным образом дается и определение полной системы функций. Напомним, что система функций $\mathcal{K}$ называется полной, если все функции можно выразить через функции данной системы, т.е. $[\mathcal{K}] = \mathcal{P}_k$. Далее мы построим несколько примеров полных систем функций k-значной логики. Один такой пример уже фактически имеется.

Теорема 9 *Система функций k-значной логики*

$$\mathcal{K} = \{0, 1, \ldots, k-1, I_0(x), I_1(x), \ldots, I_{k-1}(x), \min(x_1, x_2), \max(x_1, x_2)\}$$

является полной.

Этот факт прямо следует из теоремы 8. Далее мы построим полную систему из двух функций.

Теорема 10 *Система $\mathcal{K} = \{\bar{x}, \max(x_1, x_2)\}$ полна в $\mathcal{P}_k$.*

Доказательство. Доказательство проводится в несколько этапов. Мы последовательно будем выражать функции предыдущей (полной) системы через максимум и отрицание.

1. Построение констант.

Из функции $\bar{x} = x + 1$ последовательно получаем функции $x + 2 = (x + 1) + 1, \ldots,$ $x + (k - 1) = (x + k - 2) + 1$, $x = x + k = (x + (k - 1)) + 1$. Так как при любом x выполнено равенство

$$\max(x, x + 1, \ldots, x + (k - 1)) = k - 1,$$

мы имеем константу $k - 1$. Из нее при помощи $\bar{x} = x + 1$ получаем все остальные константы.

2. Построение функций одной переменной.

Наша первая цель — построить функции $I_\sigma(x)$ $(0 \leq \sigma \leq k - 1)$. Пусть $\alpha \in E^k$. Рассмотрим функцию

$$g_\alpha(x) = \max_{i \neq \alpha}(x + i).$$

Среди чисел $x + i$ $(i \neq \alpha)$ либо имеется $k - 1$, и тогда g_α принимает это значение, либо $k - 1$ отсутствует. В последнем случае $g_\alpha(x)$ равно $k - 2$, и этот случай имеет место при $x + \alpha = k - 1$. Таким образом,

$$g_\alpha(x) = \begin{cases} k - 2 & \text{при } x = k - 1 - \alpha, \\ k - 1 & \text{при } x \neq k - 1 - \alpha. \end{cases}$$

Отсюда очевидно, что $g_\alpha(x) + 1 = I_{k-1-\alpha}(x)$. Следовательно, функции $I_\sigma(x) = g_{k-1-\sigma}(x) + 1$ можно выразить через функции нашей системы $(0 \leq \sigma \leq k - 1)$.

Построим теперь для любых $\alpha, \beta \in E^k$ такую функцию $f_{\alpha,\beta}(x)$, которая в точке α равна β, а в остальных точках равна нулю. Для этого проверим равенство

$$f_{\alpha,\beta}(x) = \max(I_\alpha(x), k - 1 - \beta) + \beta + 1.$$

Правая часть в точке α равна $k - 1 + \beta + 1 = \beta$, так как максимум в данной точке принимает значение $k - 1$. Во всех остальных точках максимум равен $k - 1 - \beta$, и потому правая часть равна нулю.

Пусть теперь $f(x)$ — произвольная функция от одной переменной. Легко видеть, что

$$f(x) = \max(f_{0,f(0)}(x), f_{1,f(1)}(x), \ldots, f_{k-1,f(k-1)}(x)).$$

Итак, любую функцию одного переменного мы можем выразить. В частности, мы можем выразить отрицание Лукасевича:

$$\sim x = \max(f_{0,k-1}(x), f_{1,k-2}(x), \ldots, f_{k-1,0}(x)).$$

3. Построение $\min(x_1, x_2)$.

Достаточно воспользоваться аналогами законов де Моргана и двойного отрицания. Исходя из этого, имеем

$$\min(x_1, x_2) = \sim \max(\sim x_1, \sim x_2).$$

Таким образом, мы можем выразить через максимум и отрицание любую функцию, принадлежащую полной системе из предыдущей теоремы. Этим доказано, что наша система полна.

Замечание 3 Заметим, что система $\{\sim x, \max(x_1, x_2)\}$ при $k \geq 3$ не является полной. Дело в том, что обе функции данной системы сохраняют множество $\{0, k-1\}$. Следовательно, все функции, выразимые через функции данной системы, на наборах из 0 и $k-1$ принимают значения 0 или $k-1$. Так как $1 \notin \{0, k-1\}$ при $k \geq 3$, функцию $\bar{x}$ выразить нельзя.

Доказанная нами теорема легко позволяет построить полную систему, состоящую из одной функции. Положим $W_k(x_1, x_2) = \max(x_1, x_2) + 1$. Функция $W_k(x_1, x_2)$ называется *функцией Вебба*.

Следствие 4 *Система функций k-значной логики $\{W_k(x_1, x_2)\}$, состоящая из функции Вебба, является полной.*

Для доказательства достаточно заметить, что $\bar{x} = \max(x, x) + 1 = W_k(x, x)$, т.е. отрицание выразимо через функцию Вебба. Имея отрицание, мы строим последовательно функции $x + 2, \ldots, x + (k-1)$ и выражаем максимум: $\max(x_1, x_2) = W_k(x_1, x_2) + (k-1)$. Далее все следует из теоремы 10.

2.3 Алгоритм для распознавания полноты.

Теорема Поста дает алгоритм, который выясняет, будет ли данная конечная система булевых функций полной. В этом разделе мы докажем, что алгоритм распознавания полноты имеется для функций k-значной логики при любом $k \geq 2$. Для начала введем одно новое понятие и используем его для еще одного описания замыкания класса функций. (Отметим, что определение замыкания в k-значной логике дается точно так же, как и в предыдущей главе.)

Определение 13 Пусть $\mathcal{K}$, $\mathcal{L}$ — два класса функций k-значной логики. Назовем *суперпозицией классов* $\mathcal{K}$ и $\mathcal{L}$ класс, обозначаемый $\mathcal{K}(\mathcal{L})$ или $\mathcal{K} \circ \mathcal{L}$, состоящий из всех функций, представимых в виде суперпозиции некоторой функции из $\mathcal{K}$ относительно функций, каждая из которых есть либо функция из $\mathcal{L}$, либо одна из переменных.

Из определения очевидно, что $\mathcal{K} \subseteq \mathcal{K} \circ \mathcal{L}$. Также ясно, что из условий $\mathcal{K} \subseteq \mathcal{L}$ и $\mathcal{M} \subseteq \mathcal{N}$ следует, что $\mathcal{K} \circ \mathcal{M} \subseteq \mathcal{L} \circ \mathcal{N}$.

Лемма 5 *Для любых классов функций $\mathcal{K}$, $\mathcal{L}$, $\mathcal{M}$ имеет место включение*

$$(\mathcal{K} \circ \mathcal{L}) \circ \mathcal{M} \supseteq \mathcal{K} \circ (\mathcal{L} \circ \mathcal{M}). \tag{4}$$

Если $\mathcal{M} \subseteq \mathcal{L} \circ \mathcal{M}$, то имеет место и обратное включение.

Доказательство. Пусть функция Φ принадлежит классу $\mathcal{K} \circ (\mathcal{L} \circ \mathcal{M})$. Это означает, что Φ представима в виде суперпозиции: $\Phi = f(f_1, \ldots, f_s)$, где $f \in \mathcal{K}$, а функции

$f_1, \ldots, f_s$ — либо переменные, либо функции из $\mathcal{L} \circ \mathcal{M}$. Мы можем представить каждую из этих функций в виде суперпозиции следующим образом: $f_i = g_i(h_1, \ldots, h_t)$ $(1 \le i \le s)$, где $g_1, \ldots, g_s$ — либо переменные, либо функции из $\mathcal{L}$, а $h_1, \ldots, h_t$ — либо переменные, либо функции из $\mathcal{M}$ (если f_i — переменная, то ее представляем как суперпозицию двух переменных). Тогда

$$\Phi = f(f_1, \ldots, f_s) = f(g_1(h_1, \ldots, h_t), \ldots, g_s(h_1, \ldots, h_t)) = g(h_1, \ldots, h_t),$$

где $g = f(g_1, \ldots, g_s)$ — функция из $\mathcal{K} \circ \mathcal{L}$. В итоге оказывается, что $\Phi \in (\mathcal{K} \circ \mathcal{L}) \circ \mathcal{M}$. Это доказывает включение (4).

Предположим теперь, что $\mathcal{M} \subseteq \mathcal{L} \circ \mathcal{M}$ и докажем, что имеет место обратное включение. Пусть функция Φ принадлежит классу $(\mathcal{K} \circ \mathcal{L}) \circ \mathcal{M}$. Это означает, что Φ представима в виде суперпозиции: $\Phi = f(f_1, \ldots, f_s)$, где $f \in \mathcal{K} \circ \mathcal{L}$, а функции $f_1, \ldots, f_s$ — либо переменные, либо функции из $\mathcal{M}$. В свою очередь, f можно представить в виде суперпозиции: $f = g(g_1, \ldots, g_t)$, где $g \in \mathcal{K}$, а функции $g_1, \ldots, g_t$ — либо переменные, либо функции из $\mathcal{L}$. Имеет место равенство

$$\Phi = f(f_1, \ldots, f_s) = g(g_1(f_1, \ldots, f_s), \ldots, g_t(f_1, \ldots, f_s)) = g(h_1, \ldots, h_t),$$

где $h_i = g_i(f_1, \ldots, f_s)$ для всех i от 1 до t. Если g_i — функция из $\mathcal{L}$, то $h_i \in \mathcal{L} \circ \mathcal{M}$. Пусть g_i — переменная. Тогда h_i есть одна из функций $f_1, \ldots, f_s$, т.е. либо одна из переменных, либо функция из $\mathcal{M}$. В силу условия $\mathcal{M} \subseteq \mathcal{L} \circ \mathcal{M}$, во всех случаях имеем, что список функций $h_1, \ldots, h_t$ состоит либо из переменных, либо из функций класса $\mathcal{L} \circ \mathcal{M}$. Следовательно, $\Phi \in \mathcal{K} \circ (\mathcal{L} \circ \mathcal{M})$. Это доказывает обратное включение.

Пример 6 Рассмотрим такой случай: $\mathcal{K} = \{x\}$, $\mathcal{L} = \{0\}$, $\mathcal{M} = \{1\}$. Тогда $\mathcal{K} \circ \mathcal{L} = \{x, 0\}$, $(\mathcal{K} \circ \mathcal{L}) \circ \mathcal{M} = \{x, 0, 1\}$, однако $\mathcal{L} \circ \mathcal{M} = \{0\}$, $\mathcal{K} \circ (\mathcal{L} \circ \mathcal{M}) = \{x, 0\}$. Таким образом, включение (4) нельзя в общем случае заменить на равенство.

Следующая лемма дает еще одно описание замыкания класса функций.

Лемма 6 *Пусть $\mathcal{K}$ — некоторый класс функций k-значной логики. Определим по индукции последовательность классов функций $\mathcal{K}^{(i)}$, $i \ge 1$, полагая $\mathcal{K}^{(1)} = \mathcal{K}$, $\mathcal{K}^{(i+1)} = \mathcal{K} \circ \mathcal{K}^{(i)}$ для $i \ge 1$. Тогда*

1) *Последовательность $\mathcal{K}^{(i)}$ $(i \ge 1)$ удовлетворяет условию*

$$\mathcal{K}^{(1)} \subseteq \mathcal{K}^{(2)} \subseteq \ldots \subseteq \mathcal{K}^{(i)} \subseteq \ldots; \tag{5}$$

2) *для любых $i, j \ge 1$ имеет место равенство*

$$\mathcal{K}^{(i)} \circ \mathcal{K}^{(j)} = \mathcal{K}^{(i+j)}; \tag{6}$$

3) *объединение членов последовательности (5) совпадает с замыканием класса $\mathcal{K}$, т.е.*

$$\bigcup_{i=1}^{\infty} \mathcal{K}^{(i)} = [\mathcal{K}].$$

Доказательство. 1) Из элементарных свойств суперпозиции классов следует, что $\mathcal{K}^{(1)} = \mathcal{K} \subseteq \mathcal{K} \circ \mathcal{K} = \mathcal{K}^{(2)}$, а также тот факт, что для любого $i \geq 1$ из $\mathcal{K}^i \subseteq \mathcal{K}^{(i+1)}$ следует $\mathcal{K}^{i+1} = \mathcal{K} \circ \mathcal{K}^{(i)} \subseteq \mathcal{K} \circ \mathcal{K}^{(i+1)} = \mathcal{K}^{(i+2)}$. Это доказывает (5).

2) Проведем индукцию по i. При $i = 1$ обе части равенства (6) совпадают по определению. Допустим, что для некоторого $i \geq 1$ и произвольного $j \geq 1$ справедливо (6). Докажем, что (6) справедливо, если i заменить на $i + 1$. Имеем: $\mathcal{K}^{(i+1)} \circ \mathcal{K}^{(j)} = (\mathcal{K} \circ \mathcal{K}^{(i)}) \circ \mathcal{K}^{(j)}$. Заметим, что $\mathcal{K}^{(i)} \circ \mathcal{K}^{(j)} = \mathcal{K}^{(i+j)} \supseteq \mathcal{K}^{(j)}$ по предположению индукции с учетом пункта 1. Поэтому, применяя лемму 5 для $\mathcal{L} = \mathcal{K}^{(i)}$, $\mathcal{M} = \mathcal{K}^{(j)}$, заключаем, что (4) превращается в равенство, т.е.

$$\mathcal{K}^{(i+1)} \circ \mathcal{K}^{(j)} = (\mathcal{K} \circ \mathcal{K}^{(i)}) \circ \mathcal{K}^{(j)} = \mathcal{K} \circ (\mathcal{K}^{(i)} \circ \mathcal{K}^{(j)}) = \mathcal{K} \circ \mathcal{K}^{(i+j)} = \mathcal{K}^{((i+1)+j)},$$

что и требовалось доказать.

3) Напомним, как определялось замыкание класса $\mathcal{K}$. Оно является объединением последовательности множеств

$$\mathcal{K}_0 \subseteq \mathcal{K}_1 \subseteq \cdots \subseteq \mathcal{K}_r \subseteq \cdots, \tag{7}$$

где $\mathcal{K}_0 = \mathcal{K}$, $\mathcal{K}_{i+1} = \mathcal{K}_i' = \mathcal{K}_i \circ \mathcal{K}_i$. Проводя индукцию по i, докажем, что при любом $i \geq 0$ имеет место равенство $\mathcal{K}_i = \mathcal{K}^{(2^i)}$. При $i = 0$ оно справедливо, так как $\mathcal{K}_0 = \mathcal{K} = \mathcal{K}^{(1)}$. Предположим, что оно имеет место для данного $i \geq 0$ и докажем, что оно верно, если i заменить на $i+1$. Имеем: $\mathcal{K}_{i+1} = \mathcal{K}_i \circ \mathcal{K}_i = \mathcal{K}^{(2^i)} \circ \mathcal{K}^{(2^i)} = \mathcal{K}^{(2^{i+1})}$ согласно пункту 2).

Поскольку $2^i \geq i + 1$ для всех $i \geq 0$, мы имеем также включения $\mathcal{K}_i \supseteq \mathcal{K}^{(i+1)}$ $(i \geq 0)$. Таким образом, любой член последовательности (5) содержится в объединении членов последовательности (7) и обратно, любой член последовательности (7) содержится в объединении членов последовательности (5). Следовательно, объединение классов из (5) равно объединению членов (7), т.е. равно $[\mathcal{K}]$.

Лемма доказана.

Теорема 11 *Существует алгоритм, который по данной конечной системе функций k-значной логики определяет, будет ли эта система полной.*

Доказательство. Пусть дана конечная система функций $\mathcal{K} = \{f_1, \ldots, f_m\}$. Можно считать, что все они зависят от n переменных. Построим по индукции последовательность (конечных) множеств $\mathcal{M}_i$ $(i \geq 0)$, состоящих из функций от двух переменных x_1, x_2. Сразу отметим, что всего имеется конечное число функций от x_1, x_2, а именно, k^{k^2} функций.

Положим $\mathcal{M}_0 = \emptyset$. Предположим теперь, что множества $\mathcal{M}_0, \ldots, \mathcal{M}_r$ $(r \geq 0)$ уже построены. Покажем, как определяется множество $\mathcal{M}_{r+1}$. Для каждого i от 1 до m рассмотрим всевозможные функции от x_1, x_2 вида

$$f_i(h_1(x_1, x_2), \ldots, h_n(x_1, x_2)),$$

где для всех j от 1 до n функция $h_j(x_1, x_2)$ либо принадлежит $\mathcal{M}_r$, либо совпадает с одной из переменных x_1 или x_2. Проще говоря, в функции из $\mathcal{K}$ разрешается

подставлять либо функции из $\mathcal{M}_r$, либо переменные x_1, x_2. Нам придется при этом рассмотреть $m(s_r+2)^n$ функций, где s_r обозначает количество функций в $\mathcal{M}_r$. Обозначим через $\mathcal{N}_r$ множество рассмотренных функций, полагая далее $\mathcal{M}_{r+1} = \mathcal{M}_r \cup \mathcal{N}_r$.

В итоге имеем бесконечную последовательность множеств

$$\mathcal{M}_0 \subseteq \mathcal{M}_1 \subseteq \cdots \subseteq \mathcal{M}_r \subseteq \cdots,$$

причем каждое из построенных множеств состоит не более, чем из k^{k^2} элементов, так как оно состоит из функций от x_1, x_2. Следовательно, должен найтись такой номер t, при котором $\mathcal{M}_t = \mathcal{M}_{t+1}$. Очевидно, что последовательность тем самым стабилизируется, начиная с t-го члена. Более того, ясно, что наименьший номер t, начиная с которого последовательность стабилизируется, не превосходит k^{k^2}. Это означает, что, строя последовательно множества $\mathcal{M}_r$, мы за конечное число шагов обнаружим момент, при котором $\mathcal{M}_t = \mathcal{M}_{t+1}$. Введем обозначение $\mathcal{M}_* = \mathcal{M}_t$. Возможны два случая.

1) $\mathcal{M}_*$ содержит функцию Вебба $W_k(x_1, x_2)$. Тогда можно сделать вывод, что система $\mathcal{K}$ полна. Действительно, функция Вебба, как и любая функция из $\mathcal{M}_r$ ($r \geq 0$), выражается через функции системы $\mathcal{K}$ (т.е. принадлежит замыканию класса $\mathcal{K}$), а следствие 4 утверждает, что любая функция выражается через функцию Вебба.

2) $\mathcal{M}_*$ не содержит функцию Вебба $W_k(x_1, x_2)$. Докажем, что в этом случае система $\mathcal{K}$ не полна. Достаточно показать, что $W_k(x_1, x_2) \notin [\mathcal{K}]$. Предположим противное. Воспользуемся описанием замыкания из леммы 6. Мы заключаем, что $W_k(x_1, x_2) \in \mathcal{K}^{(i)}$ для некоторого $i \geq 1$. Достаточно доказать, что любая функция от x_1, x_2 из $\mathcal{K}^{(i)}$ принадлежит $\mathcal{M}_i$. Это сразу приводит к противоречию, так как получается, что W_k принадлежит $\mathcal{M}_i$, а значит и $\mathcal{M}_*$.

При помощи индукции по i мы докажем более общий факт, а именно, что если у функции из $\mathcal{K}^{(i)}$ ее переменные заменить на x_1, x_2, то получится функция из $\mathcal{M}_i$. При $i = 1$ это прямо следует из определения $\mathcal{M}_1$. Допустим, что наше утверждение справедливо для данного значения i и докажем его для значения $i+1$. Рассмотрим произвольную функцию $f \in \mathcal{K}^{(i+1)} = \mathcal{K} \circ \mathcal{K}^{(i)}$. По определению, f есть суперпозиция некоторой функции $f_j \in \mathcal{K}$ ($1 \leq j \leq m$) относительно функций из $\mathcal{K}^{(i)}$ или переменных: $f(y_1, \ldots, y_t) = f_j(h_1(y_1, \ldots, y_t \ \ldots, h_n(y_1, \ldots, y_t))$, где h_1, $\ldots$, h_n — функции из $\mathcal{K}^{(i)}$ или переменные. Если теперь заменить переменные y_1, $\ldots$, y_t на x_1, x_2, то функции h_1, $\ldots$, h_n после замен станут либо переменными x_1, x_2, либо будут принадлежать $\mathcal{M}_i$ в силу предположения индукции. Левая часть станет тем самым функцией из $\mathcal{M}_{i+1}$, что и требовалось доказать.

Мы доказывали (следствие 3), что из любой полной системы булевых функций можно выделить конечную подсистему. Это верно для функций k-значной логики при любом $k \geq 2$. Докажем вспомогательное утверждение, из которого нужный нам факт легко вытекает.

Лемма 7 *Если $\mathcal{K}$ — произвольный класс функций, и $f \in [\mathcal{K}]$, то существует конечное подмножество $\mathcal{L} = \mathcal{L}_f \subseteq \mathcal{K}$ такое, что $f \in [\mathcal{L}]$.*

Подчеркнем, что $\mathcal{L}$ зависит от f.

Доказательство. Поскольку $f \in [\mathcal{K}]$, согласно пункту 3 леммы 6 найдется $i \geq 1$ такое, что $f \in \mathcal{K}^{(i)}$. Это позволяет применить индукцию по i. При $i = 1$ функция f принадлежит $\mathcal{K}$, и можно положить $\mathcal{L}_f = \{f\}$. Допустим, что для некоторого $i \geq 1$ всякая функция $f \in \mathcal{K}^{(i)}$ принадлежит замыканию некоторого конечного подмножества $\mathcal{L}_f \subseteq \mathcal{K}$. Докажем, что утверждение остается верным, если от i перейти к $i+1$.

Пусть $f \in \mathcal{K}^{(i+1)} = \mathcal{K} \circ \mathcal{K}^{(i)}$. Представим f в виде суперпозиции: $f = g(h_1, \ldots, h_n)$, где $g \in \mathcal{K}$, а функции h_j $(1 \leq j \leq n)$ либо принадлежат $\mathcal{K}^{(i)}$, либо являются переменными. Пусть J — множество всех j, для которых $h_j \in \mathcal{K}^{(i)}$. Для каждого $j \in J$ ввиду предположения индукции существует конечное подмножество $\mathcal{L}_j$ в $\mathcal{K}$ такое, что $h_j \in [\mathcal{L}_j]$. Положим $\mathcal{L} = \bigcup\limits_{j \in J} \mathcal{L}_j \cup \{g\}$. Ясно, что $\mathcal{L}$ конечно как объединение конечного числа конечных множеств. При этом $\mathcal{L} \subseteq \mathcal{K}$, и из определения замыкания очевидно, что $f \in [\mathcal{L}]$.

Лемма доказана.

Следствие 5 *Из любой полной системы функций k-значной логики можно выделить полную конечную подсистему.*

Доказательство. Пусть $\mathcal{K}$ — полная система. Тогда функция Вебба выражается через функции системы $\mathcal{K}$. По лемме 7, функция Вебба принадлежит замыканию некоторой конечной подсистемы $\mathcal{L}$ в $\mathcal{K}$, т.е. W_k выражается через конечную подсистему функций из $\mathcal{K}$. Любая функция выражается через W_k согласно следствию 4, поэтому $\mathcal{L}$ — полная конечная подсистема.

2.4 Предполные классы. Теорема Кузнецова.

В предыдущей главе было доказано, что имеется в точности пять предполных классов для булевых функций. Мы установим, что число предполных классов k-значной логики конечно при любом $k \geq 2$.

Для начала дадим эквивалентное определение предполного класса и укажем, как выразить свойство полноты системы на языке предполных классов.

Рассмотрим множество всех *собственных* замкнутых подклассов в $\mathcal{P}_k$ (подкласс $\mathcal{K}$ называется собственным, если он не совпадает с классом всех функций, т.е. $\mathcal{K} \subset \mathcal{P}_k$). Замкнутый собственный подкласс $\mathcal{K} \subset \mathcal{P}_k$ называется *максимальным*, если не существует замкнутого подкласса $\mathcal{L}$ такого, что $\mathcal{K} \subset \mathcal{L} \subset \mathcal{P}_k$. Нетрудно видеть, что класс является предполным (в смысле определения, данного в предыдущей главе) тогда и только тогда, когда он максимален. Обозначим через $\mathcal{A}_k$ множество замкнутых собственных подклассов в $\mathcal{P}_k$. Оно непусто, так как содержит, например, класс констант. На множестве $\mathcal{A}_k$ имеется (нестрогий) частичный порядок $\subseteq$.

Лемма 8 *Любой замкнутый собственный класс $\mathcal{S} \subset \mathcal{P}_k$ содержится в некотором максимальном классе. В частности, предполные классы существуют.*

Доказательство. Рассмотрим в $\mathcal{A}_k$ подмножество $\mathcal{B}$, состоящее из всех классов, содержащих $\mathcal{S}$. Покажем, что в частично упорядоченном множестве $(\mathcal{B}, \subseteq)$ имеются максимальные элементы. Воспользуемся теоретико-множественной **леммой Цорна**: *если любое линейно упорядоченное подмножество частично упорядоченного множества $(\mathcal{A}, \preceq)$ имеет в $\mathcal{A}$ верхнюю грань, то в $\mathcal{A}$ имеется максимальный элемент*[1]. Рассмотрим произвольное линейно упорядоченное подмножество X в $\mathcal{B}$, и пусть $\mathcal{K}$ — объединение всех классов, этому подмножеству принадлежащих. Достаточно проверить, что $\mathcal{K} \in \mathcal{B}$, т.е. $\mathcal{K}$ — замкнутый собственный подкласс, содержащий $\mathcal{S}$.

Проверим, что $\mathcal{K}$ совпадает со своим замыканием. Возьмем произвольную функцию $f \in [\mathcal{K}]$ и докажем, что $f \in \mathcal{K}$. По лемме 7, имеется конечное подмножество $\mathcal{L}$ в $\mathcal{K}$ такое, что $f \in [\mathcal{L}]$. Так как $\mathcal{L}$ содержится в объединении классов, образующих X, являясь при этом конечным, то $\mathcal{L}$ содержится в объединении конечного числа классов, принадлежащих X. Поскольку X линейно упорядочено, любое его непустое конечное подмножество имеет наибольший (в смысле порядка $\subseteq$) элемент. Отсюда следует, что в X имеется класс $\mathcal{M}$, которому принадлежат все элементы $\mathcal{L}$, т.е. $\mathcal{L} \subseteq \mathcal{M}$. Класс $\mathcal{M}$ замкнут, поэтому $[\mathcal{L}] \subseteq \mathcal{M}$. Таким образом, $f \in \mathcal{M}$, а потому $f \in \mathcal{K}$.

Теперь осталось проверить, что $\mathcal{K}$ — собственный класс. Это следует из того, что функция Вебба не принадлежит ни одному из собственных замкнутых классов $\mathcal{M}$, следовательно, не принадлежит объединению всех собственных классов. В частности, W_k не принадлежит $\mathcal{K}$, а потому $\mathcal{K}$ — собственный.

Включение $\mathcal{S} \subseteq \mathcal{K}$ очевидно.

Лемма доказана.

Итак, доказано, что предполные классы существуют. Охарактеризуем на языке предполных классов условие полноты. Оказывается, *система функций $\mathcal{S}$ полна тогда и только тогда, когда она не содержится целиком ни в одном из предполных классов*. Действительно, если $\mathcal{S}$ содержится в некотором предполном классе $\mathcal{K}$, то и $[\mathcal{S}]$ содержится в $\mathcal{K}$ ввиду замкнутости $\mathcal{K}$. Поэтому $[\mathcal{S}] \neq \mathcal{P}_k$, т.е. $\mathcal{S}$ не полна. Обратно, пусть $\mathcal{S}$ не содержится целиком ни в одном из предполных классов. Рассмотрим класс $[\mathcal{S}]$. Если он является собственным, то по лемме 8 он содержится в некотором максимальном, т.е. предполном классе. Следовательно, $[\mathcal{S}] = \mathcal{P}_k$, т.е. $\mathcal{S}$ — полная система.

Рассуждение, приведенное выше показывает большую важность вопроса о предполных классах. Следующая теорема, устанавливающая конечность числа предполных классов, была доказана А. В. Кузнецовым.

Теорема 12 *Число предполных классов в $\mathcal{P}_k$ конечно.*

[1] Элемент $u \in \mathcal{A}$ называется верхней гранью подмножества $X \subseteq \mathcal{A}$, если $x \preceq u$ для любого $x \in X$.

Доказательство. Для каждого предполного класса рассмотрим те функции от двух переменных x_1, x_2, которые ему принадлежат. Покажем, что этот набор функций полностью определяет данный предполный класс. Так как число функций от двух переменных равно k^{k^2}, существует не более $2^{k^{k^2}}$ множеств, состоящих из таких функций. Отсюда будет следовать, что число предполных классов конечно.

Введем обозначение. Если $\mathcal{K}$ — некоторый класс функций, то через $\mathcal{K}_{x_1x_2}$ будем обозначать его подмножество, состоящее из всех функций от x_1, x_2, принадлежащих $\mathcal{K}$. Нам нужно доказать, что если $\mathcal{K}$ — предполный класс, то он полностью определяется своим подмножеством $\mathcal{K}_{x_1x_2}$.

Удобно ввести еще один термин. Пусть $\mathcal{M}$ — некоторый класс функций от переменных $y_1, \ldots, y_m$, содержащий переменные y_i для всех i от 1 до m. Скажем, что функция $f(x_1, \ldots, x_n)$ *сохраняет* класс $\mathcal{M}$, если для любых функций $h_1, \ldots, h_n \in \mathcal{M}$ суперпозиция $\Phi = f(h_1, \ldots, h_n)$ также принадлежит $\mathcal{M}$. Например, функция x сохраняет любой класс, константа сохраняет класс тогда и только тогда, когда она ему принадлежит и т. п.

Сохраняя обозначения из предыдущего абзаца, докажем, что класс $\mathcal{N}$, состоящий из всех функций f, сохраняющих $\mathcal{M}$, является замкнутым. Ясно, что тождественная функция принадлежит $\mathcal{N}$, поэтому достаточно доказать, что для любой функции $f \in \mathcal{N}$ от s переменных и любых функций $g_1, \ldots, g_s \in \mathcal{N}$ суперпозиция $\Psi = f(g_1, \ldots, g_s)$ также принадлежит $\mathcal{N}$. Пусть функции g_i $(1 \leq i \leq s)$ зависят от n переменных. Рассмотрим произвольные n функций $h_1, \ldots, h_n \in \mathcal{M}$ и образуем суперпозицию $\Phi = \Psi(h_1, \ldots, h_n)$. Введем обозначения $\phi_i = g_i(h_1, \ldots, h_n)$ $(1 \leq i \leq s)$. Так как каждая из функций g_i сохраняет $\mathcal{M}$, заключаем, что $\phi_i \in \mathcal{M}$ $(1 \leq i \leq s)$. Далее, f сохраняет $\mathcal{M}$, поэтому суперпозиция $\Phi = \Psi(h_1, \ldots, h_n) = f(g_1(h_1, \ldots, h_n), \ldots, g_s(h_1, \ldots, h_n)) = f(\phi_1, \ldots, \phi_s)$ также принадлежит $\mathcal{M}$. Следовательно, Ψ сохраняет $\mathcal{M}$, т.е. принадлежит $\mathcal{N}$, что и требовалось доказать.

Нетрудно убедиться в том, что предполный класс содержит тождественную функцию, так как для любого класса $\mathcal{K}$ выполнено равенство $[\mathcal{K} \cup \{x\}] = [\mathcal{K}] \cup \{x\}$.

Нам достаточно доказать, что любой предполный класс $\mathcal{K}$ состоит в точности из функций, сохраняющих $\mathcal{K}_{x_1x_2}$. (В силу сказанного выше, переменные x_1, x_2 входят в $\mathcal{K}_{x_1x_2}$.)

Если $f \in \mathcal{K}$, то суперпозиция функции f относительно функций из $\mathcal{K}_{x_1x_2}$ также принадлежит $\mathcal{K}$, так как $\mathcal{K}$ замкнут. Ясно, что эта суперпозиция также принадлежит $\mathcal{K}_{x_1x_2}$. Таким образом, f сохраняет класс $\mathcal{K}_{x_1x_2}$.

Обратно, пусть $f \notin \mathcal{K}$. Тогда замыкание класса $\mathcal{K} \cup \{f\}$ есть $\mathcal{P}_k$. Рассуждая от противного, допустим, что f сохраняет класс $\mathcal{K}_{x_1x_2}$. Так как и функции из $\mathcal{K}$, и f, сохраняют $\mathcal{K}_{x_1x_2}$, а класс всех функций, сохраняющих $\mathcal{K}_{x_1x_2}$, замкнут, мы получаем, что все функции из $\mathcal{P}_k = [\mathcal{K} \cup \{f\}]$ сохраняют $\mathcal{K}_{x_1x_2}$. Это заведомо невозможно, так как функция Вебба $W_k(x_1, x_2)$ не принадлежит $\mathcal{K}_{x_1x_2}$, а потому и не может сохранять этот класс. Противоречие показывает, что f не сохраняет $\mathcal{K}_{x_1x_2}$.

Мы доказали, что любой предполный класс $\mathcal{K}$ состоит в точности из функций, сохраняющих $\mathcal{K}_{x_1x_2}$. Таким образом, предполных классов в k-значной логике имеется

не более $2^{k^{k^2}}$.

Теорема доказана.

Теорема Кузнецова дает оценку числа предполных классов, очень быстро растущую с ростом k. Точное количество предполных классов известно только для небольших значений k. Как мы видели, в двузначной логике имеется 5 предполных классов. Предполные классы трехзначной логики описаны С. В. Яблонским; их имеется 18. С ростом k описание значительно усложняется.

В заключение приведем без доказательства формулировку теоремы Слупецкого, которая дает еще один критерий полноты системы функций k-значной логики.

Теорема 13 *Пусть система S функций из P_k, где $k \geq 3$, содержит все функции одной переменной. Тогда для полноты системы S необходимо и достаточно, чтобы S содержала функцию, существенно зависящую не менее чем от двух переменных, принимающую все k значений.*

Заметим, что при $k = 2$ это утверждение неверно, так как система $S = \{0, 1, x, \bar{x}, x_1 + x_2\}$ удовлетворяет условиям теоремы, но неполна.

2.5 Особенности k-значных логик.

В предыдущих разделах мы показали, что многие результаты, относящиеся к двузначной логике, обобщаются на случай k-значной логики для любого $k \geq 2$. В настоящем разделе мы укажем некоторые особенности двузначной логики.

Вспомним некоторые результаты предыдущей главы (первые два приводились без доказательства).

1. Любой замкнутый класс в P_2 имеет конечный базис.

2. Множество всех замкнутых классов в P_2 счетно.

3. Любая функция из P_2 может быть представлена полиномом по модулю 2.

Напомним, что подсистема S класса K называется его базисом, если замыкание S равно K, но замыкание любого собственного подмножества системы S не равно K.

Приведем несколько результатов о свойствах замкнутых классов в P_k при $k \geq 3$. Следующий результат принадлежит Ю. И. Янову.

Теорема 14 *Для любого $k \geq 3$ можно указать в P_k замкнутый класс, не имеющий базиса.*

Доказательство. Рассмотрим последовательность функций

$$f_0 = 0, \quad f_n(x_1, \ldots, x_n) = \begin{cases} 1 & \text{при } x_1 = \cdots = x_n = 2, \\ 0 & \text{в остальных случаях} \end{cases} \qquad (n \geq 1).$$

Обозначим через $\mathcal{M}_k$ класс функций от переменных $x_1, \ldots, x_n, \ldots$, получающихся из f_n $(n \geq 0)$ путем переименования переменных (без их отождествления). Легко проверить, что если в некоторую функцию из $\mathcal{M}_k$ подставить вместо ее переменных либо функции из $\mathcal{M}_k$, либо переменные, то мы не выйдем за пределы $\mathcal{M}_k$. В самом деле, если f — функция из $\mathcal{M}_k$, существенно зависящая от n переменных, а $g_1, \ldots, g_n$ — либо переменные, либо функции из $\mathcal{M}_k$, то суперпозиция $\Phi = f(g_1, \ldots, g_n)$ будет тождественно нулевой при условии, что $g_i \in \mathcal{M}_k$ хотя бы для одного i от 1 до n. Если же все $g_1, \ldots, g_n$ суть переменные, среди которых ровно $m \leq n$ различных, то Φ получается из функции f_m путем переименования переменных.

Предположим, что $\mathcal{M}_k$ имеет базис. Рассмотрим минимальное число m такое, что базис содержит функцию f, получающуюся из f_m путем переименования переменных. Очевидно, что базис не может содержать еще одну функцию, которая получается переименованием переменных из f_m, так как в противном случае ее можно было бы исключить из базиса. Поэтому возможны два случая.

1. Базис содержит функцию g, которая получается переименованием переменных из f_n, где $n > m$. Очевидно, что f_m получается из f_n путем подстановки переменных: $f_m(x_1, \ldots, x_m) = f_n(x_1, \ldots, x_m, x_m, \ldots, x_m)$. Поэтому f выражается через g, что противоречит определению базиса

2. Базис состоит только из f. Легко понять, что при $n > m$ никакая функция f_n не выражается через f, так как через f можно выразить только функции f_s при $0 \leq s \leq m$, а также функции, получаемые из них переименованием переменных. Это также противоречит определению базиса.

Итак, наше рассуждение показывает, что $\mathcal{M}_k$ не имеет базиса.

Приведем еще один результат, полученный А. А. Мучником.

Теорема 15 *Для любого $k \geq 3$ можно указать в $\mathcal{P}_k$ замкнутый класс, имеющий счетный базис.*

Доказательство. Для любого $n \geq 1$ рассмотрим функцию f_n от $n+1$ переменных, принимающую значение 1 на всех наборах, в которые входят n двоек и одна единица и принимающую значение 0 на всех остальных наборах. Обозначим через $\mathcal{L}_k$ замыкание системы функций $\mathcal{S}_i = \{f_1, f_2, \ldots, f_n, \ldots\}$. Для каждого $m \geq 1$ рассмотрим подсистему $\mathcal{S}_{k,m}$ в $\mathcal{S}_k$, получающуюся из $\mathcal{S}_k$ удалением функции f_m. Нашей целью является доказательство того, что $f_m \notin [\mathcal{S}_{k,m}]$ для любого $m \geq 1$. Из этого будет следовать, что $\mathcal{S}_k$ является базисом в $\mathcal{L}_k$.

Допустим, что при некотором $m \geq 1$ функция f_m выражается через функции системы $\mathcal{S}_{k,m}$. Тогда она является суперпозицией некоторой функции f_r $(r \neq m)$

относительно некоторых функций $g_1, \ldots, g_{r+1}$, причем последние суть либо переменные, либо принадлежат замыканию системы $\mathcal{S}_{k,m}$. Можно считать, что все они являются функциями от переменных $x_1, \ldots, x_{m+1}$, и имеет место равенство

$$f_m(x_1, \ldots, x_{m+1}) = f_r\left(g_1(x_1, \ldots, x_{m+1}), \ldots, g_{r+1}(x_1, \ldots, x_{m+1})\right). \qquad (8)$$

При этом мы учитываем, что если бы функции g_i $(1 \le i \le r+1)$ зависели от каких-либо других переменных, отличных от $x_1, \ldots, x_{m+1}$, то их можно было заменить, скажем, на переменную x_1 с сохранением всех интересующих нас свойств.

Рассмотрим несколько случаев.

1. Все $g_1, \ldots, g_{r+1}$ — переменные. Так как все переменные функции f_m являются существенными, правая часть равенства (8) содержит вхождения всех переменных $x_1, \ldots, x_{m+1}$. Следовательно, $r \ge m$, и ввиду $r \ne m$ имеем $r > m$. Тогда среди функций $g_1, \ldots, g_{r+1}$, принимающих $m+1$ значение, имеются одинаковые. Пусть переменная x_i $(1 \le i \le m+1)$ входит дважды в правую часть. Выберем значения переменных $x_1, \ldots, x_{m+1}$: пусть x_i принимает значение 1, а все остальные переменные — значение 2. Тогда левая часть (8) становится равной 1, а правая часть принимает значение 0, так как f_r на наборе, в который входят по крайней мере две единицы, равна нулю.

2. Среди $g_1, \ldots, g_{r+1}$ имеется ровно одна функция g_j, не являющаяся переменной. При этом среди этих функций имеется и некоторая переменная x_i. Ясно, что g_j на любых наборах значений переменных не может принимать значение 2. Приладим переменной x_i значение 1, а всем остальным переменным — значение 2. Тогда на данном наборе значений левая часть (8) равна 1, а правая — нулю, так как среди значений функций $g_1, \ldots, g_{r+1}$ имеется не менее двух, отличных от 2.

3. Среди $g_1, \ldots, g_{r+1}$ имеется не менее двух функций, не являющихся переменными. Тогда правая часть (8) тождественно равна нулю, а левая — нет.

Теорема доказана.

Следствие 6 *Для любого $k \ge 3$ множество замкнутых классов в $\mathcal{P}_k$ имеет мощность континуума.*

Доказательство. Рассмотрим класс $\mathcal{L}_k$, имеющий счетный базис f_n $(n \ge 1)$ из предыдущей теоремы. Для каждого непустого подмножества X множества натуральных чисел рассмотрим класс $\mathcal{L}_{k,X}$, являющийся замыканием системы функций $\{ f_i \mid i \in X \}$. Если X, Y — различные непустые множества, то классы $\mathcal{L}_{k,X}$, $\mathcal{L}_{k,Y}$ также различны. Действительно, предположим, что $\mathcal{L}_{k,X} = \mathcal{L}_{k,Y}$. Для любого $i \in X$ функция f_i принадлежит классу $\mathcal{L}_{k,Y}$. Поскольку f_i не выражается через функции системы $\{ f_j \mid j \ne i \}$, функция f_i должна входить в систему $\{ f_i \mid i \in Y \}$, откуда $i \in Y$. Таким образом, $X \subseteq Y$. Аналогично, $Y \subseteq X$, т.е. $X = Y$.

Множество всех (непустых) подмножеств множества натуральных чисел имеет мощность континуума. Следовательно, множество замкнутых классов в $\mathcal{P}_k$ имеет не меньшую мощность. Но само множество $\mathcal{P}_k$ счетно, и поэтому множество всех его

подмножеств имеет мощность континуума. Отсюда можно заключить[2], что множество замкнутых классов в $\mathcal{P}_k$ имеет мощность, в точности равную мощности континуума.

Нам осталось рассмотреть вопрос о представимости функций k-значной логики полиномами по модулю k. На этот счет имеется следующий результат.

Теорема 16 *Все функции из $\mathcal{P}_k$ ($k \geq 2$) представимы полиномами по модулю k тогда и только тогда, когда k — простое число.*

Доказательство. Для каждого $i \in \{0, 1, \ldots, k-1\}$ введем функцию

$$\chi_i(x) = \begin{cases} 1 & \text{при } x = i, \\ 0 & \text{при } x \neq i. \end{cases}$$

Очевидно, что для любой функции $f(x_1, \ldots, x_n)$ имеет место тождество

$$f(x_1, \ldots, x_n) = \sum_{(\sigma_1, \ldots, \sigma_n)} \chi_{\sigma_1}(x_1) \cdots \chi_{\sigma_n}(x_n) f(\sigma_1, \ldots, \sigma_n),$$

где суммирование ведется по всем наборам значений переменных. Поэтому вопрос о представимости всех функций полиномами сводится к вопросу о представимости полиномами функций $\chi_i(x)$ ($0 \leq i \leq k-1$). Ясно, что $\chi_i(x) = \chi_0(x - i)$ для всех i от 1 до $k-1$, поэтому все сводится к вопросу о представимости функции $\chi_0(x)$.

Допустим, что функция $\chi_0(x)$ представима полиномом по модулю k. Можно, конечно, считать, что этот полином зависит только от переменной x (если это не так, то все входящие в него переменные заменим на x). Мы имеем равенство по модулю k вида

$$\chi_0(x) = a_0 + a_1 x + \cdots + a_m x^m, \tag{9}$$

где все коэффициенты a_i ($0 \leq i \leq m$) суть константы. Ясно, что $1 = \chi_0(0) = a_0$. Предположим, что число k не простое. Тогда оно имеет делитель ℓ, где $1 < \ell < k$. Так как $\ell \neq 0$, имеем $\chi_0(\ell) = 0$. С другой стороны, при $x = \ell$ равенство (9) дает $0 = \chi(\ell) \equiv 1 \pmod{\ell}$, что невозможно.

Итак, пусть теперь k простое. Воспользуемся Малой теоремой Ферма, утверждающей, что при a не делящемся на k имеет место сравнение $a^{k-1} \equiv 1 \pmod{k}$. Легко проверить, что по модулю k имеет место равенство $\chi_0(x) = 1 - x^{k-1}$, которое и дает нужное представление полиномом.

Теорема доказана.

В дополнение к доказанной теореме заметим, что в случае, когда число k простое, мы можем представить любую функцию от n переменных в виде полинома

$$f(x_1, \ldots, x_n) = \sum_{0 \leq d_1, \ldots, d_n \leq k-1} a_{d_1 \ldots d_n} x_1^{d_1} \ldots x_n^{d_n}, \tag{10}$$

[2]При этом используется теорема Кантора – Бернштейна: если мощности двух множеств не превосходят друг друга, то они равны.

так как $x^k \equiv x \,(\mathrm{mod} k)$ для всех x в силу Малой теоремы Ферма. Оказывается, что такое представление единственно. Чтобы убедиться в этом, достаточно подсчитать число таких полиномов. Одночленов вида $x_1^{d_1} \dots x_n^{d_n}$, где $0 \leq d_1, \dots, d_n \leq k-1$, имеется в точности k^n. При каждом из этих одночленов можно выбрать коэффициент k способами, что дает k^{k^n} полиномов. В силу теоремы 7 это число совпадает с числом всех функций от n переменных. Это доказывает единственность представления функции от n переменных в виде (10). Таким образом, для простых значений k имеется аналог полиномов Жегалкина.

Подведем итог результатам, полученным в этом разделе.

1. При $k \geq 3$ существуют замкнутые классы в $\mathcal{P}_k$, не имеющие конечного базиса (можно указать замкнутые классы, не имеющие базиса, а также имеющие счетный базис).

2. При $k \geq 3$ множество всех замкнутых классов в $\mathcal{P}_k$ имеет мощность континуума.

3. Все функции из $\mathcal{P}_k$ могут быть представлена полиномами по модулю k тогда и только тогда, когда k простое.

3 Кодирование

3.1 Понятие кодирования.

В данной главе речь будет идти об алфавитном кодировании. Мы будем везде опускать слово "алфавитное", говоря просто о кодировании. Вначале условимся о некоторой терминологии, касающейся слов над заданным алфавитом.

Непустое множество будем называть *алфавитом*, а его элементы — *буквами*. Конечные последовательности букв алфавита будут называться *словами* (над этим алфавитом). Мы будем также допускать пустую последовательность букв, называемую *пустым словом*, обозначаемым через Λ. Однобуквенное слово будем отождествлять с соответствующей буквой. На множестве слов определена бинарная операция *умножения* слов (иногда ее называют также *конкатенацией*). Если x, y — некоторые слова, то их произведение (результат приписывания справа к слову x слова y) обозначается $x \cdot y$ или просто xy. Легко видеть, что данная операция ассоциативна, причем пустое слово служит нейтральным элементом. Множество всех непустых слов над алфавитом A, обозначаемое A^+, является таким образом полугруппой относительно умножения. Эту полугруппу называют *свободной полугруппой* над алфавитом A. Множество всех слов над A, включая пустое, обозначается через A^*. Оно является полугруппой с единицей (моноидом) относительно умножения и называется *свободным моноидом* над алфавитом A. Любое слово $x \in A^+$ представимо, причем единственным образом, в виде произведения элементов из A, т.е. букв: $x = x_1 \dots x_n$, где $x_i \in A$ $(1 \leq i \leq n)$. Число n (количество букв в слове x) называется *длиной* слова x и обозначается через $|x|$. Длину пустого слова естественно считать равной нулю. Очевидно, для любых $x, y \in A^*$ справедливо равенство $|xy| = |x| + |y|$.

Слово y называется *началом* или *префиксом* слова x, если существует слово z такое, что $x = yz$. Если при этом z непусто, то y называют *собственным началом слова* x. Пустое слово является началом любого слова. Аналогично, слово y называется *концом* или *суффиксом* слова x, если существует слово z такое, что $x = zy$. Если при этом z непусто, то y называют *собственным концом слова* x. Пустое слово также является концом любого слова. Слово y называют *подсловом* слова x, если найдутся слова z, w такие, что $x = zyw$.

Напомним, что гомоморфизмом полугрупп называют отображение $\phi\colon S \to T$ полугруппы S в полугруппу T, сохраняющее операцию, т.е. $\phi(xy) = \phi(x)\phi(y)$ для любых $x, y \in S$. Пусть A, B — два алфавита, которые для простоты будем считать конечными. *Кодированием* мы будем называть гомоморфизм свободных полугрупп $\phi\colon A^+ \to B^+$. Если x — слово над A, то слово $\phi(x)$ над B называется *кодом* слова x. Легко понять, что всякий гомоморфизм $\phi\colon A^+ \to B^+$ свободных полугрупп однозначно определяется заданием образов букв алфавита A. Более того, если произвольным образом сопоставить буквам из A некоторые слова над B, то полученное отображение единственным образом продолжается до гомоморфизма из A^+ в B^+. Пусть $A = \{a_1, \dots, a_m\}$, $B = \{b_1, \dots, b_n\}$. Гомоморфизм свободных полугрупп $\phi\colon A^+ \to B^+$

(т.е. кодирование) мы будем задавать *схемой кодирования* вида

$$\begin{array}{ccc} a_1 & \mapsto & B_1 \\ a_2 & \mapsto & B_2 \\ \vdots & \vdots & \vdots \\ a_m & \mapsto & B_m, \end{array} \qquad (11)$$

указывающей образы букв алфавита A, где $B_1, \ldots, B_m$ — произвольные непустые слова над алфавитом B, называемые *элементарными кодами* или *кодовыми словами*. Данная схема задает кодирование ϕ, при котором код $\phi(x)$ слова x получается в результате одновременной замены букв слова x на соответствующие им кодовые слова.

3.2 Однозначно декодируемые коды.

Определение 14 Кодирование $\phi: A^+ \to B^+$ называется *однозначным* или *однозначно декодируемым*, если отображение ϕ инъективно, т.е. различные слова имеют различные коды.

Приведем простое достаточное условие однозначности кодирования.

Определение 15 Кодирование называется *префиксным*, если ни одно из кодовых слов не является началом другого. Кодирование называется *суффиксным*, если ни одно из кодовых слов не является концом другого.

Пример 7 Кодирование, заданное схемой

$$\begin{array}{ccc} a & \mapsto & xyzxz \\ b & \mapsto & xyz \\ c & \mapsto & yx \\ d & \mapsto & zzyy \end{array}$$

является суффиксным, но не является префиксным.

Теорема 17 *Кодирование, являющееся префиксным или суффиксным, однозначно.*

Доказательство. Предположим, что дано префиксное кодирование $\phi: A^+ \to B^+$. Проверим, что оно однозначно. Возьмем произвольное слово w, лежащее в образе отображения ϕ. Покажем, как восстановить его прообраз и убедимся, что он единствен. Легко видеть, что существует в точности одно кодовое слово, являющееся началом w, так как в противном случае одно из кодовых слов было бы началом другого. Тем самым мы находим первую из букв прообраза слова w. Отсекая от w найденное кодовое слово, мы продолжаем описанный процесс до тех пор, пока не останется пустое слово.

Случай суффиксного кодирования отличается лишь тем, что рассматриваются концы слова.

Пример 8 Кодирование в примере 7 позволяет однозначно разложить слово $w = zzyyxyzxzxyzxyzyx$ в произведение кодовых слов, двигаясь справа налево: $w = zzyy \cdot xyzxz \cdot xyz \cdot xyz \cdot yx = \phi(dabbc)$. Таким образом, w есть код слова $dabbc$.

Можно показать, что кодирование может быть однозначным, не будучи ни префиксным, ни суффиксным.

Пример 9 Рассмотрим кодирование ϕ, заданное схемой

$$\begin{aligned} a_1 &\mapsto b_1 \\ a_2 &\mapsto b_1 b_2 \\ a_3 &\mapsto b_3 b_1. \end{aligned}$$

Очевидно, что ϕ не является ни префиксным, ни суффиксным. Тем не менее оно однозначно. Действительно, если w — код некоторого слова, то любому вхождению буквы b_2 в слово w предшествует вхождение буквы b_1, а за вхождением буквы b_3 следует буква b_1. Выделим все вхождения вида $b_1 b_2$ и $b_3 b_1$. Очевидно, что они попарно не пересекаются, а остальные буквы слова w равны b_1. Это доказывает однозначность. Скажем, если $w = b_1 b_1 b_1 b_2 b_3 b_1 b_1 b_1 b_2$, то $w = b_1 b_1 (b_1 b_2)(b_3 b_1) b_1 (b_1 b_2) = \phi(a_1 a_1 a_2 a_3 a_1 a_2)$.

Пусть кодирование задано своей схемой. Возникает естественный вопрос, можно ли алгоритмически определить, будет ли это кодирование однозначным. Ответ положителен. Сформулируем соответствующий факт.

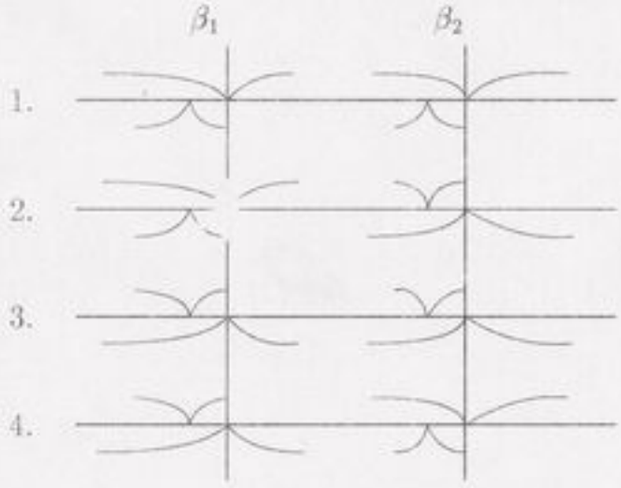
Теорема 18 *Пусть дано кодирование $\phi: A^+ \to B^+$, где A — алфавит из m букв, а M — максимум длин кодовых слов. Оно является однозначным тогда и только тогда, когда однозначно декодируемо любое слово над алфавитом B, имеющее длину не более $m(M-1)^2 + M$.*

Таким образом, перебирая конечное множество слов, мы можем проверить, однозначно ли кодирование.

Доказательство. Если кодирование однозначно, то любое слово над B однозначно декодируемо по определению. Предположим теперь, что кодирование ϕ, заданное схемой (11), не однозначно. Тогда найдутся два различных слова $x, y \in A^+$, имеющие один и тот же образ $w \in B^+$. Выберем такие слова, для которых w имеет как можно меньшую длину. Рассмотрим две различные расшифровки слова w. С ними связаны два разбиения слова w на элементарные коды: верхнее и нижнее (см. иллюстрацию).

Ввиду минимальности длины слова w можно сделать вывод, что внутренние точки одного из разбиений не будут точками другого из разбиений. Рассмотрим объединение двух разбиений. Оно определяет разложение слова w в произведение подслов (мы будем называть их *кусками*), которые бывают двух видов. К первому виду относятся те куски, которые получаются как кодовые слова одного разбиения, содержащиеся в некотором кодовом слове другого разбиения. Ко второму виду отнесем остальные куски. Последние являются собственными началами некоторого кодового слова одного из разбиений и собственными концами некоторого кодового слова другого разбиения.

Мы покажем, что все слова второго вида, встречающиеся в разбиении, попарно различны. Допустим, что одно из таких слов β_1 совпало с другим словом β_2. Слово w можно представить в виде $w = w'\beta_1 w''\beta_2 w'''$. Возможны 4 случая, в зависимости от того, началами или концами кодовых слов будут β_1, β_2. Эти случаи изображены ниже.



Легко видеть, что третий и четвертый случай аналогичны первому и второму случаям соответственно (нужно просто поменять разбиения ролями). В первом случае удалим из слова w его подслово $w''\beta_2$. Получим новое (более короткое, нежели w) слово, для которого существуют две различных расшифровки: нужно просто склеить части верхнего и нижнего разбиения после удаления $w''\beta_2$. Во втором случае также удалим подслово $w''\beta_2$, но теперь верхнее разбиение нового слова получается склейкой верхнего разбиения для $w'\beta_1$ и нижнего разбиения для w''', а нижнее разбиение — склейкой нижнего разбиения для w'_{+1} и верхнего разбиения для w'''. Таким образом, в каждом из случаев мы приходим к слову, более короткому, нежели w, допускающему при этом различные расшифровки. Это противоречит выбору слова w. Итак, куски второго вида попарно различны.

Оценим число N кусков второго вида. Каждый из них является собственным началом одного из кодовых слов, что дает не более $(|B_1|-1)+(|B_2|-1)+\ldots+(|B_m|-1) \leq m(M-1)$ слов. Таким образом, $N \leq m(M-1)$. Теперь можно оценить сверху длину слова w. Обозначим куски второго вида через $D_1, \ldots, D_N$, нумеруя их слева направо. Тогда w можно представить в виде $w = D_0U_0D_1U_1 \ldots D_NU_ND_{N+1}$, где слова

D_0, D_{N+1} пусты по определению. Каждое из слов $D_iU_iD_{i+1}$ является кодовым словом одного из разбиений, откуда имеем неравенства $|D_i|+|U_i|+|D_{i+1}| \le M$ $(0 \le i \le N)$. Суммируя все эти неравенства по i от 0 до N, получаем $2(|D_1|+\ldots+|D_N|)+(|U_1|+\ldots+|U_N|) \le M(N+1)$. Следовательно, $|w| = (|D_1|+\ldots+|D_N|)+(|U_1|+\ldots+|U_N|) \le M(N+1)-(|D_1|+\ldots+|D_N|) \le M(N+1)-N = N(M-1)+M \le m(M-1)^2+M$.

Теорема доказана.

Алгоритм проверки кодирования на однозначность, изложенный выше, не является очень эффективным практически, так как он связан с просмотром большого числа слов. Ниже мы покажем, как получить более наглядную и практически более эффективную версию этого алгоритма.

Пусть Σ — схема кодирования (11). Назовем слово γ *двусторонним*, если существуют кодовые слова B_i, B_j $(1 \le i,j \le m$, возможно, $i=j)$ такие, что γ является собственным началом слова B_i и одновременно собственным концом слова B_j (при этом γ может быть пустым).

Рассмотрим все разложения кодовых слов следующего вида:

$$B_i = \gamma' B_{i_1} \ldots B_{i_d} \gamma'', \tag{12}$$

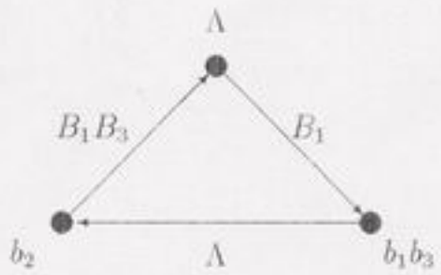
где $d \ge 0$, $1 \le i, i_1, \ldots, i_d \le m$, γ' и γ'' — двусторонние слова. При этом допускается, что одно или оба слова γ', γ'' могут быть пустыми, но мы исключаем случай *тривиального разложения*, при котором $\gamma' = \gamma'' = \Lambda$, $d=1$, $i=i_1$. Разложений вида (12) имеется конечное число. Рассмотрим ориентированный граф[3], вершинами которого служат двусторонние слова. Для каждого разложения вида (12) проведем в графе ориентированное ребро, снабженное *меткой* $B_{i_1}\ldots B_{i_d}$. Полученный граф будем обозначать через $\Gamma = \Gamma(\Sigma)$. Заметим, что можно брать не все двусторонние слова, а только те, которые встречаются в разложениях вида (12) и в качестве γ', и в качестве γ''. Далее, для наших целей будет достаточно рассматривать лишь те вершины графа, в которые можно прийти по ребрам графа из вершины Λ, двигаясь по направлениям стрелок. Для данной схемы кодирования Σ граф Γ легко построить.

Пример 10 Зададим кодирование Σ_1 схемой

$$\begin{aligned} a_1 &\mapsto b_1b_2 \\ a_2 &\mapsto b_1b_3b_2 \\ a_3 &\mapsto b_2b_3 \\ a_4 &\mapsto b_1b_2b_1b_3 \\ a_5 &\mapsto b_2b_1b_2b_2b_3. \end{aligned}$$

Через B_i будем обозначать код буквы a_i $(1 \le i \le 5)$. Двусторонними словами будут в точности слова $\gamma_0 = \Lambda$, $\gamma_1 = b_2$, $\gamma_2 = b_1b_3$. Все разложения вида (12) таковы: $B_2 = (b_1b_3)(b_2) = \gamma_2\Lambda\gamma_1$, $B_4 = (b_1b_2)(b_1b_3) = \gamma_0B_1\gamma_2$, $B_5 = (b_2)(b_1b_2)(b_2b_3) = \gamma_1B_1B_3\gamma_0$. Имеем следующий граф с тремя вершинами и тремя ориентированными ребрами:

[3]Точное определение графа будет дано в следующих главах. Пока мы ограничиваемся наглядным представлением.



Рассмотрим замкнутый путь из вершины Λ в вершину Λ, состоящий из трех ребер. Пусть w — произведение меток вершин и ребер, последовательно проходимых при этом: $w = \Lambda B_1 b_1 b_3 \Lambda b_2 B_1 B_3 \Lambda = b_1 b_2 b_1 b_3 b_2 b_1 b_2 b_2 b_3$. Слово w можно естественным образом представить в виде произведения кодовых слов двумя способами: $w = B_1 B_2 B_1 B_3 = B_4 B_5$. Таким образом, данное кодирование не является однозначным. Причиной является наличие цикла в графе.

Теорема 19 *Пусть Σ — схема кодирования, и пусть $\Gamma = \Gamma(\Sigma)$ — соответствующий граф. Кодирование является однозначным тогда и только тогда, когда в Γ нет ориентированных замкнутых циклов с началом и концом в вершине Λ.*

Доказательство. Для начала предположим, что в Γ имеется ориентированный замкнутый путь с началом и концом в вершине Λ. Перечислим последовательно метки всех вершин и ребер данного пути: γ_0, u_0, γ_1, u_1, …, γ_n, u_n, γ_{n+1}, где γ_i $(0 \le i \le n+1)$ — метки вершин, $\gamma_0 = \gamma_{n+1} = \Lambda$, а u_i $(0 \le i \le n)$ — метки ребер, каждая из которых есть произведение кодовых слов. Кроме того, $c_i = \gamma_i u_i \gamma_{i+1}$ есть кодовое слово для всех i от 1 до n. Тогда имеем для слова $w = \gamma_0 u_0 \dots u_n \gamma_{n+1}$ два разложения в произведение кодовых слов: $w = c_0 u_1 c_2 u_3 \dots$ и $w = u_0 c_1 u_2 \dots$. Эти разложения различны ввиду нетривиальности представления кодового слова c_0 в виде $\gamma_0 u_0 \gamma_1$.

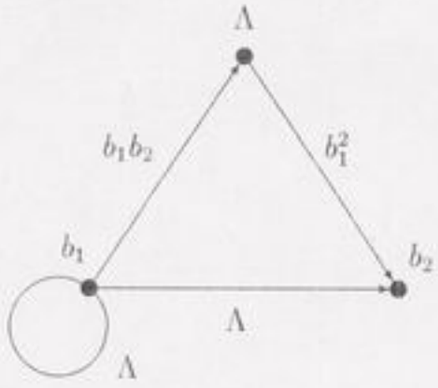
Обратно, предположим, что кодирование не является однозначным. Выберем кратчайшее по длине слово w, которое декодируется неоднозначно. Так же, как и в доказательстве теоремы 18, рассмотрим два разбиения слова w, не имеющие общих точек (см. иллюстрацию на стр. 36). При этом слово w представляется в виде $w = D_0 U_0 D_1 U_1 \dots D_N U_N D_{N+1}$, где слова D_0, D_{N+1} пусты, каждое из слов $D_i U_i D_{i+1}$ — кодовое слово, а каждое из слов U_i $(0 \le i \le N)$ — произведение кодовых слов. При этом в графе Γ возникает путь с началом и концом в вершине $\Lambda = D_0 = D_{N+1}$, последовательно проходимыми вершинами которого являются D_i $(0 \le i \le N+1)$, а последовательно проходимые ребра имеют метки U_i $(0 \le i \le N)$.

Теорема доказана.

Пример 11 Рассмотрим кодирование, заданное схемой

$$\begin{aligned} a_1 &\mapsto b_1^2 \\ a_2 &\mapsto b_1 b_2 \\ a_3 &\mapsto b_1^2 b_2 \\ a_4 &\mapsto b_2^2 b_1 \end{aligned}$$

Двусторонними словами будут Λ, b_1, b_2. Обозначим через B_i код буквы a_i $(1 \le i \le 4)$. Нетривиальные разложения таковы: $B_1 = b_1\Lambda b_1$, $B_2 = b_1\Lambda b_2$, $B_3 = \Lambda B_1 b_2$, $B_3 = b_1 B_2 \Lambda$. Граф Γ состоит из трех вершин и четырех ребер:



Ясно, что в нем нет ориентированных циклов с началом и концом в Λ. Следовательно, по теореме 19 данное кодирование однозначно.

3.3 Коды с минимальной избыточностью.

Рассмотрим следующий естественный вопрос: каким условиям должны удовлетворять длины кодовых слов, чтобы кодирование при этом было однозначным? Необходимое условие дается следующей теоремой.

Теорема 20 *Если кодирование $\phi: A \to B$ с длинами кодовых слов $\ell_1, \ldots, \ell_m$ однозначно, где A — алфавит из m букв, а B — алфавит из q букв, то выполняется следующее неравенство (неравенство Макмиллана):*

$$\frac{1}{q^{l_1}} + \ldots + \frac{1}{q^{l_m}} \le 1.$$

Доказательство. Через B_i обозначим i-е кодовое слово, имеющее длину ℓ_i $(1 \le i \le m)$. Рассмотрим числовое тождество

$$\left(\frac{1}{q^{l_1}} + \ldots + \frac{1}{q^{l_m}}\right)^n = \sum_{(i_1,\ldots,i_n)} \frac{1}{q^{l_{i_1}+\ldots+l_{i_n}}},$$

где n — произвольное натуральное число, а суммирование в правой части ведется по всем последовательностям длиной n, члены которых независимо принимают значения от 1 до m. Через $\nu(n,t)$ обозначим число наборов вида $(i_1,\ldots,i_n)$, для которых слово $B_{i_1}\ldots B_{i_n}$ имеет длину t. Можно записать

$$\sum_{(i_1,\ldots,i_n)} \frac{1}{q^{l_{i_1}+\ldots+l_{i_n}}} = \sum_{t=1}^{n\ell} \frac{\nu(n,t)}{q^t},$$

где ℓ — максимум длин кодовых слов.

Однозначность кодирования позволяет заключить, что $\nu(n,l) \le q^l$. Следовательно, правая часть последнего равенства не превосходит $n\ell$. Отсюда выводим, что

$$\frac{1}{q^{l_1}} + \ldots + \frac{1}{q^{l_m}} \le \sqrt[n]{n\ell}$$

для любого натурального n. Его левая часть не зависит от n, а правая стремится к 1 при $n \to \infty$. Переходя к пределу в неравенстве, получаем искомую оценку.

Оказывается, неравенство Макмиллана является также достаточным условием существования однозначного кодирования с заданными длинами кодовых слов. Справедлив даже более сильный результат.

Теорема 21 *Пусть натуральные числа q, $\ell_1, \ldots, \ell_m$ удовлетворяют неравенству*

$$\frac{1}{q^{l_1}} + \ldots + \frac{1}{q^{l_m}} \le 1.$$

Тогда существует префиксное кодирование $\phi\colon A \to B$, где A — алфавит из m букв, B — алфавит из q букв, а кодовые слова имеют длины $\ell_1, \ldots, \ell_m$.

Доказательство. Пусть среди чисел $\ell_1, \ldots, \ell_m$ имеется в точности μ попарно различных, принимающих значения $\lambda_1 < \cdots < \lambda_\mu$, причем кодовых слов длиной λ_i имеется ν_i $(1 \le i \le \mu)$. Неравенство из условия теоремы можно записать в виде

$$\sum_{i=1}^{\mu} \frac{\nu_i}{q^{\lambda_i}} \le 1.$$

Из него выводим в качестве следствий следующие вспомогательные неравенства: $\frac{\nu_1}{q^{\lambda_1}} \le 1$ или $\nu_1 \le q^{\lambda_1}$, $\frac{\nu_1}{q^{\lambda_1}} + \frac{\nu_2}{q^{\lambda_2}} \le 1$ или $\nu_2 \le q^{\lambda_2} - \nu_1 q^{\lambda_2-\lambda_1}$, $\ldots$, $\frac{\nu_1}{q^{\lambda_1}} + \frac{\nu_2}{q^{\lambda_2}} + \ldots + \frac{\nu_\mu}{q^{\lambda_\mu}} \le 1$ или $\nu_\mu \le q^{\lambda_\mu} - \nu_1 q^{\lambda_\mu-\lambda_1} - \nu_2^{\lambda_\mu-\lambda_2} - \cdots - \nu_{\mu-1} q^{\lambda_\mu-\lambda_{\mu-1}}$.

Ввиду неравенства $\nu_1 \le q^{\lambda_1}$, можно выбрать ν_1 слов над алфавитом B, которые имеют длину λ_1. Исключим из рассмотрения все слова, начинающиеся с хотя бы одного из выбранных слов. При этом в нашем распоряжении останется не менее $q^{\lambda_2} - \nu_1 q^{\lambda_2-\lambda_1}$ слов длиной λ_2, и ввиду неравенства $\nu_2 \le q^{\lambda_2} - \nu_1 q^{\lambda_2-\lambda_1}$ мы можем выбрать ν_2 слов над алфавитом B, которые имеют длину λ_2, и ни одно из них не начинается ни с одного из выбранных ранее слов. Продолжая таким образом, мы на последнем шаге сможем выбрать ν_μ слов над B, имеющих длину λ_μ, и при этом ни одно из выбранных нами слов не является началом никакого другого. Выбранные нами слова определяют префиксное кодирование и имеют заданные длины $\ell_1, \ldots, \ell_m$.

Теорема доказана.

Следствие 7 *Если существует однозначное кодирование $\phi\colon A \to B$ с длинами кодовых слов $\ell_1, \ldots, \ell_m$, то существует префиксное кодирование $\psi\colon A \to B$ с теми же длинами кодовых слов.*

Действительно, числа q, ℓ_1, ..., ℓ_m удовлетворяют неравенству Макмиллана по теореме 20, а потому имеется и префиксное кодирование с теми же длинами кодовых слов в силу теоремы 21.

Рассмотрим теперь постановку задачи о построении кодирования с минимальной избыточностью. Пусть $A = \{a_1, \ldots, a_r\}$ — алфавит из r букв. Известны средние частоты (вероятности) появления этих букв. Именно, пусть $p_1, \ldots, p_r$ — положительные числа, сумма которых равна 1, где p_i есть вероятность появления буквы a_i $(1 \le i \le r)$. Пусть $B = \{b_1, \ldots, b_q\}$ — другой алфавит из q букв. Требуется построить однозначное кодирование $\phi\colon A \to B$ с длинами кодовых слов $\ell_1, \ldots, \ell_r$ соответственно, таким образом, чтобы средняя длина передаваемых сообщений была минимальна. В качестве характеристики кодирования мы берем число, равное

$$\sum_{i=1}^{r} p_i \ell_i = p_1\ell_1 + \ldots + p_r\ell_r,$$

называемое *ценой* или *избыточностью* данного кодирования. Итак, при заданных числах $p_1, \ldots, p_r$ и q требуется построить (однозначное) кодирование с минимальной избыточностью.

Убедимся сначала в том, что оптимальное кодирование, т.е. однозначное кодирование, имеющее минимально возможную избыточность, существует. Действительно, если ℓ — такое натуральное число, что $q^\ell \ge r$, то можно построить однозначное кодирование, взяв в качестве кодовых слов любые r слов над B, которые имеют длину ℓ. Данное кодирование имеет избыточность ℓ и не будет, вообще говоря, оптимальным. Пусть p_α — наименьшее из чисел $p_1, \ldots, p_r$. Ясно, что наличие хотя бы одного кодового слова длиной $> \ell/p_\alpha$ привело бы к кодированию с ценой, большей ℓ. Таким образом, имеет смысл рассматривать лишь кодирования, у которых длины кодовых слов не превосходят величины ℓ/p_α. Таких кодирований имеется конечное число, и поэтому среди них существует кодирование, имеющее минимальную цену. Оно и будет оптимальным.

Далее будем всюду без ограничения общности предполагать, что вероятности упорядочены следующим образом: $0 < p_1 \le p_2 \le \cdots \le p_r$.

Лемма 9 *Предположим, что кодовые слова оптимального кодирования имеют длины $\ell_1, \ldots, \ell_r$ соответственно. Тогда из условия $p_i < p_j$ следует, что $\ell_i \ge \ell_j$.*

Доказательство. Допустим, что $\ell_i < \ell_j$. При этом $p_i\ell_j + p_j\ell_i < p_i\ell_i + p_j\ell_j$, так как $(p_i\ell_j + p_j\ell_i) - (p_i\ell_i + p_j\ell_j) = (p_i - p_j)(\ell_j - \ell_i) < 0$. Это означает, что перестановка местами i-го и j-го кодовых слов приводит к кодированию с меньшей избыточностью.

Лемма доказана.

В силу леммы, можно считать, что длины кодовых слов оптимального кодирования для списка вероятностей $0 < p_1 \le p_2 \le \cdots \le p_r$ удовлетворяют условиям $\ell_1 \ge \ell_2 \ge \cdots \ge \ell_r$, где ℓ_i — длина i-го кодового слова.

Сейчас мы покажем, как построить оптимальный код для случая *двоичного кодирования*, т.е. при $q = 2$. Будем считать, что $B = \{0; 1\}$. Вначале опишем построение такого кодирования, используя индукцию по r, а затем докажем его оптимальность.

При $r = 1$ просто берем в качестве кодового слова однобуквенное слово 0. Отдельно также рассматриваем случай $r = 2$: в качестве кодовых слов берем однобуквенные слова 0 и 1. Предположим теперь, что $r > 2$, а кодирование для алфавита из $r - 1$ букв уже построено. Положим $p = p_1 + p_2$. Упорядочим теперь по неубыванию числа p, p_3, ..., p_r. Построим кодирование с кодовыми словами w, w_3, ..., w_r. Теперь i-й букве исходного r-буквенного алфавита сопоставим слово w_i, если $i \geq 3$, а первой и второй буквам сопоставим кодовые слова $w0$ и $w1$ соответственно.

Пример 12 Пусть $r = 4$, а вероятности равны $p_1 = 0,15$, $p_2 = 0,2$, $p_3 = 0,25$, $p_4 = 0,4$. Вначале сводим задачу к построению кодирования для списка вероятностей $p_3 = 0,25$, $p_1 + p_2 = 0,35$, $p_4 = 0,4$ (мы сложили два наименьших числа прежнего списка и упорядочили полученный новый список). Далее, вновь складывая два наименьших числа, приходим к списку $p_4 = 0,4$, $p_3 + (p_1 + p_2) = 0,6$. Для данного списка кодовые слова суть 0 и 1. Идя теперь в обратном направлении согласно описанному выше, мы получаем, что для списка $p_3 = 0,25$, $p_1 + p_2 = 0,35$, $p_4 = 0,4$ кодовые слова равны 10, 11, 0, а для исходного списка $p_1 = 0,15$, $p_2 = 0,2$, $p_3 = 0,25$, $p_4 = 0,4$ приходим к кодовым словам 110, 111, 10, 0. Легко видеть, что соответствующее кодирование будет префиксным, а потому однозначным. Избыточность составляет $3 \cdot 0,15 + 3 \cdot 0,2 + 2 \cdot 0,25 + 1 \cdot 0,4 = 0,45 + 0,6 + 0,5 + 0,4 = 1,95$. Заметим, что если бы мы вместо этого использовали кодовые слова 00, 01, 10, 11, то получили бы кодирование, имеющее избыточность $2 > 1,95$.

Докажем лемму, из которой будет прямо вытекать оптимальность кодирования, описанного выше.

Лемма 10 *Пусть p, $p_1 \leq \ldots \leq p_m$ ($m \geq 1$) — положительные числа, причем $p + p_1 + \cdots + p_m = 1$. Предположим, что кодовые слова w, w_1, ..., w_m задают префиксное двоичное кодирование, оптимальное для данного списка вероятностей. Пусть p', p'' — положительные числа, $p = p' + p''$, причем $p' \leq p'' \leq p_1 \leq \ldots \leq p_m$. Тогда кодовые слова $w0$, $w1$, w_1, ..., w_m задают префиксное кодирование для списка вероятностей p', p'', p_1, ..., p_m, и оно является оптимальным.*

Доказательство. Очевидно, что если кодовые слова w, w_1, ..., w_m задают префиксное двоичное кодирование, то и слова $w0$, $w1$, w_1, ..., w_m также задают префиксное кодирование. Рассуждая от противного, предположим, что кодирование, заданное словами $w0$, $w1$, w_1, ..., w_m, не является оптимальным. Пусть ℓ — длина слова w, ℓ_i — длина слова w_i ($1 \leq i \leq m$). Рассмотрим оптимальное (для списка вероятностей p', p'', p_1, ..., p_m) кодирование с кодовыми словами v', v'', v_1, ..., v_m, длины которых равны d', d'', d_1, ..., d_m соответственно. Следствие 7 позволяет считать, что это кодирование является префиксным. По предположению, имеем неравенство

$$p'd' + p''d'' + \sum_{i=1}^{m} p_i d_i < p'(\ell + 1) + p''(\ell + 1) + \sum_{i=1}^{m} p_i \ell_i = p(\ell + 1) + \sum_{i=1}^{m} p_i \ell_i. \qquad (13)$$

Лемма 9 позволяет считать, что $d' \geq d'' \geq d_1 \geq \cdots \geq d_m$. Проанализируем, как выглядят кодовые слова, имеющие максимальную длину d'. Возьмем одно из таких слов. Без ограничения общности предположим, что это слово имеет вид $v0$ для некоторого слова v. Ясно, что v непусто: иначе $d' = d'' = d_1 = 1$, но тогда кодирование содержит два одинаковых (однобуквенных) кодовых слова, что противоречит его свойству быть префиксным. Если среди кодовых слов не встречается слово $v1$, то замена кодового слова $v0$ на v не нарушает префиксность кода, так как $v0$ — кодовое слово максимальной длины. Ввиду оптимальности рассматриваемого кодирования, заключаем, что кодовых слов максимальной длины не менее двух, т.е. $d' = d''$, причем можно считать, что $v' = v0$ и $v'' = v1$. Кодовые слова $v, v_1, \ldots, v_m$ определяют, очевидно, префиксное кодирование. Действительно, слово v не может быть префиксом ни одного из слов v_i $(1 \leq i \leq m)$, так как иначе оказалось бы, что либо v' начинается с v_i (в случае $v = v_i$), либо v_i начинается с одного из слов v', v'' (в случае, когда v — собственное начало v_i). Для списка вероятностей $p, p_1, \ldots, p_m$ рассмотрим кодовые слова $v, v_1, \ldots, v_m$. Цена получаемого при этом кодирования не меньше минимальной, что приводит к неравенству

$$p\ell + \sum_{i=1}^{m} p_i \ell_i \leq pd + \sum_{i=1}^{m} p_i d_i,$$

где d — длина слова v, т.е. $d' = d'' = d + 1$. Тогда

$$p(\ell + 1) + \sum_{i=1}^{m} p_i \ell_i \leq p(d + 1) + \sum_{i=1}^{m} p_i d_i = p'd' + p''d'' + \sum_{i=1}^{m} p_i d_i,$$

что противоречит неравенству (13).

Лемма доказана.

Данная лемма позволяет обосновать, почему построенные нами кодирования обладают свойством минимальной избыточности. При $q > 2$ также имеется алгоритм, позволяющий строить коды с минимальной избыточностью, но мы здесь этот алгоритм не приводим.

Коды с минимальной избыточностью, о которых шла речь, называют также *кодами Хаффмана*.

3.4 Коды с исправлением ошибок.

Так как на практике при передаче информации возможны помехи, приводящие к искажению передаваемой информации, имеет смысл задача построения кодов, при использовании которых исходную информацию можно восстановить. Такие коды называются *кодами с исправлением ошибок* или *самокорректирующимися кодами*. Мы рассмотрим в качестве простейшего случая так называемые коды с исправлением одной ошибки. Уточним постановку задачи.

Пусть нам требуется передавать сообщения, которые представляют собой слова над алфавитом $B = \{0; 1\}$. При этом за счет помех некоторые передаваемые нами символы могут искажаться. Допустим, что нам априори известно, что при передаче N бит возможна ошибка не более, чем в одном символе. Простейший способ избежать искажений — передать нужное сообщение три раза подряд. Таким образом, мы можем передать сообщение вида www, где w — слово над B, длина которого не превышает $N/3$. При таком способе передачи сообщений мы втрое увеличиваем объем передаваемой информации, что весьма неэкономно. Рассмотрим значительно более экономный способ, предложенный Хэммингом.

Пусть нам требуется передать k бит информации, т.е. слово длиной k из нулей и единиц. Для того, чтобы можно было распознать возникшую из-за помех ошибку, нам нужно добавить несколько дополнительных бит информации. Пусть мы добавляем m бит, получая ℓ-битное слово, где $\ell = k + m$. (Для способа, рассмотренного выше, $m = 2k$, $\ell = 3k$.) Будем считать, что при передаче ℓ-битных сообщений возможно искажение не более одного из передаваемых символов. Отсюда следует, что посланное нами ℓ-битное сообщение может принять один из $\ell + 1$ видов: оно может либо не претерпеть изменений, либо один из его ℓ бит изменится на противоположный. Для того, чтобы дополнительных m бит хватило для различения этих $\ell + 1$ случаев, необходимо выполнение неравенства $2^m \geq \ell + 1$. С учетом $\ell = k + m$, данное неравенство можно переписать в виде $2^{\ell - k} \geq \ell + 1$, т.е.

$$\frac{2^\ell}{\ell + 1} \geq 2^k. \tag{14}$$

При фиксированном ℓ выберем наибольшее k, удовлетворяющее неравенству (14). Легко видеть, что такое k имеется. При этом число $m = \ell - k$ является наименьшим, для которого выполнено условие $2^m \geq \ell + 1$. Это означает, что $2^{m-1} \leq \ell$.

Опишем построение кодов Хэмминга, предполагая, что числа k и ℓ удовлетворяют неравенству (14), $\ell = k + m$, $2^{m-1} \leq \ell$.

Нам требуется определить способ построения ℓ-битного слова $\beta_1\beta_2\ldots\beta_\ell$ по заданному k-битному слову $\alpha_1\ldots\alpha_k$. Рассмотрим m чисел, являющихся степенями двойки: $1 = 2^0$, $2 = 2^1$, $\ldots$, 2^{m-1}, последнее из которых не превосходит ℓ. Разряды β_i передаваемого сообщения ($1 \leq i \leq \ell$), для которых $i \in \{1, 2, \ldots, 2^{m-1}\}$, будем называть *контрольными*, а остальные разряды — *информационными*. Ясно, что информационных разрядов имеется в точности $\ell - m = k$. Заполним их членами последовательности $\alpha_1\ldots\alpha_k$:

$$\beta_3 = \alpha_1,\ \beta_5 = \alpha_2,\ \beta_6 = \alpha_3,\ \beta_7 = \alpha_4,\ \ldots$$

Любое число s такое, что $0 \leq s \leq \ell \leq 2^m - 1$, можно записать в виде m-разрядного двоичного числа (разряды нумеруем справа налево числами от 0 до $m-1$). Обозначим i-й разряд числа s через $\delta_i(s)$, где $0 \leq i \leq m - 1$. При этом двоичная запись числа s имеет вид $\delta_{m-1}(s)\ldots\delta_1(s)\delta_0(s)$. Пусть множество M_i ($0 \leq i \leq m-1$) состоит из всех s таких, для которых $\delta_i(s) = 1$ ($1 \leq s \leq \ell$). Мы имеем

$$M_0 = \{1, 3, 5, 7, 9, 11, \ldots\},$$

$$
\begin{aligned}
M_1 &= \{2,3,6,7,10,11,\ldots\},\\
M_2 &= \{4,5,6,7,12,13,\ldots\}
\end{aligned}
$$

и так далее. Ясно, что число 2^i есть наименьший элемент множества M_i для всех i от 0 до $m-1$. Наша цель — задать контрольные разряды таким образом, чтобы для любого i от 0 до $m-1$ сумма по модулю 2 всех разрядов ℓ-битного сообщения с номерами, принадлежащими M_i, равнялась нулю. Именно, для этого мы полагаем

$$
\begin{aligned}
\beta_1 &= \beta_3+\beta_5+\beta_7+\beta_9+\beta_{11}+\ldots,\\
\beta_2 &= \beta_3+\beta_6+\beta_7+\beta_{10}+\beta_{11}+\ldots,\\
\beta_4 &= \beta_5+\beta_6+\beta_7+\beta_{12}+\beta_{13}+\ldots,
\end{aligned}
$$

и так далее. Суммирование везде происходит по модулю 2.

Итак, коды Хэмминга построены. Покажем теперь, как обнаружить возможную ошибку. Пусть отправлено сообщение $\beta_1\beta_2\ldots\beta_\ell$, а получено сообщение $\beta'_1\beta'_2\ldots\beta'_\ell$. Через s обозначим номер разряда, в котором при передаче допущена ошибка. Если ошибки нет, что полагаем $s=0$. Таким образом, $0\le s\le\ell$. Для каждого i от 0 до $m-1$ вычислим сумму S_i по всем разрядам с номерами, принадлежащими M_i:

$$
\begin{aligned}
S_0 &= \beta'_1+\beta'_3+\beta'_5+\beta'_7+\beta'_9+\beta'_{11}+\ldots,\\
S_1 &= \beta'_2+\beta'_3+\beta'_6+\beta'_7+\beta'_{10}+\beta'_{11}+\ldots,\\
S_2 &= \beta'_4+\beta'_5+\beta'_6+\beta'_7+\beta'_{12}+\beta'_{13}+\ldots
\end{aligned}
$$

и так далее, суммируя по модулю 2. Соответствующие суммы для передаваемого сообщения $\beta_1\beta_2\ldots\beta_\ell$ равны нулю. Если $\delta_{m-1}\ldots\delta_1\delta_0$ — двоичная запись числа s, то число S_i будет равно 1 тогда и только тогда, когда $s\in M_i$, т.е. $\delta_i=1$. Это означает, что, вычисляя суммы S_0, S_1, …, S_{m-1}, мы можем узнать число s — его двоичная запись равна $S_{m-1}\ldots S_1S_0$, т.е. $s=S_0+2S_1+\cdots+2^{m-1}S_{m-1}$.

Теперь остается исправить s-й разряд на противоположный (если $s=0$, то ничего исправлять не нужно). Далее нужно прочитать подряд все информационные разряды полученного ℓ-битного слова, что даст нам исходное сообщение.

Пример 13 Пусть $\ell=7$. Нетрудно видеть, что наибольшее k, при котором выполнено неравенство (14), равно 4. При этом $m=3$. Допустим, что требуется передать сообщение $\alpha_1\alpha_2\alpha_3\alpha_4$. При этом передается 7-битное сообщение $\beta_1\ldots\beta_7$ и принимается сообщение $\beta'_1\ldots\beta'_7$. Пусть принято сообщение 1001011. Вычислим суммы S_i $(0\le i\le 2)$:

$$
\begin{aligned}
S_0 &= \beta'_1+\beta'_3+\beta'_5+\beta'_7=1+0+0+1=0,\\
S_1 &= \beta'_2+\beta'_3+\beta'_6+\beta'_7=0+0+1+1=0,\\
S_2 &= \beta'_4+\beta'_5+\beta'_6+\beta'_7=1+0+1+1=1.
\end{aligned}
$$

Следовательно, искажение произошло в разряде, двоичная запись которого есть 100, т.е. в четвертом разряде. Итак, передано было сообщение $\beta_1\ldots\beta_7=1000011$, а исходное сообщение было равно $\beta_3\beta_5\beta_6\beta_7=0011$.