ДИСКРЕТНАЯ МАТЕМАТИКА

ИУ5 — 4 семестр

Введение

Развитие классической („непрерывной“) математики было обусловлено прежде всего решением задач естествознания, главным образом физики.

„Дискретная“ же математика развивалась в связи с изучением законов и правил человеческого мышления, что и обусловило ее применение в тех областях техники, которые связаны с моделированием мышления, и в первую очередь в вычислительной технике и программировании.

Курс состоит из следующих разделов:

1. Множества и отношения.
2. Элементы классической общей алгебры.
3. Теория графов.
4. Регулярные языки и конечные автоматы.
5. Булевы функции.

Лекция 1. ЭЛЕМЕНТЫ ТЕОРИИ МНОЖЕСТВ

1.1. Множества

Понятие **множества** является исходным, для него нельзя дать строгого математического определения. Множество состоит из **элементов**.

Основоположник теории множеств Георг Кантор, поясняя интуитивную идею множества, писал:

„Под многообразием или множеством я понимаю вообще все многое, которое возможно мыслить как единое, т.е. такую совокупность определенных элементов, которая посредством одного закона может быть соединена в одно целое.“

Множества будем обозначать большими буквами латинского алфавита, а их элементы — малыми, хотя иногда от этого соглашения придется отступать, так как элементами некоторого множества могут быть другие множества.

Тот факт, что элемент a принадлежит множеству A, записывается в виде $a \in A$.

Для элементов этих множеств мы используем два основных вида обозначений: константы и переменные.

Константа с областью значений A обозначает фиксированный элемент множества A.
Например, обозначения действительных чисел (в десятичной системе счисления): 0; 2; $7{,}34$.

Для двух констант a и b с областью значений A будем писать $a = b$, понимая под этим совпадение обозначаемых ими элементов множества A.

Переменное с областью значений A обозначает произвольный, заранее не определенный элемент множества A.Переменное x пробегает множество A или переменное x принимает произвольные значения на множестве A.

Можно фиксировать значение переменного x, записав $x = a$, где a — константа с той же областью значений, что и x.Вместо переменного x подставлено его конкретное значение a, или произведена подстановка a вместо x, или переменное x приняло значение a.

Равенство переменных $x = y$: всякий раз, когда переменное x принимает произвольное значение a, переменное y принимает то же самое значение a, и наоборот.
Равные переменные „синхронно“ принимают всегда одни и те же значения.

Константы и переменные, область значений которых есть некоторое числовое множество называют:

$\mathbb{N}$ — натуральными.

$\mathbb{Z}$ — целыми.

$\mathbb{Q}$ — рациональными.

$\mathbb{R}$ — действительными.

$\mathbb{C}$ —комплексными.

Для сокращения записи мы будем пользоваться логической символикой, позволяющей коротко, записывать *высказывания*.

Понятие высказывания не определяется. Указывается только, что всякое высказывание может быть истинным или ложным (не одновременно!).

Для образования из уже имеющихся высказываний новых высказываний используются следующие **логические операции** (или **логические связки**).

1. *Дизъюнкция* $\vee$: высказывание $P \vee Q$
„P или Q“
истинно тогда и только тогда, когда истинно хотя бы одно из высказываний P и Q.

2. *Конъюнкция* $\wedge$: высказывание $P \wedge Q$
„P и Q“
истинно тогда и только тогда, когда истинны оба высказывания P и Q.

3. *Отрицание* $\neg$: высказывание $\neg P$
„не P“
истинно тогда и только тогда, когда P ложно.

4. *Импликация* $\Rightarrow$: высказывание $P \Rightarrow Q$
„если P , то Q “ или „ P влечет Q “
истинно тогда и только тогда, когда истинно высказывание Q или оба высказывания ложны.

5. *Эквивалентность* (или **равносильность**) $\Leftrightarrow$: высказывание $P \Leftrightarrow Q$
„ P , если и только если Q “
истинно тогда и только тогда, когда оба высказывания P и Q либо одновременно истинны, либо одновременно ложны.

Любые два высказывания P и Q , такие, что истинно $P \Leftrightarrow Q$, называют **логически эквивалентными** или **равносильными**.

Очередность выполнения всех операций определяется расстановкой скобок.При отсутствии скобок порядок выполнения операций определяется „соглашением о приоритетах“.

Операция отрицания всегда имеет высший приоритет, т.е. выполняется первой (ее в скобки обычно не заключают).
Второй выполняется операция конъюнкции, затем дизъюнкции. Операции импликации и эквивалентности имеют равный приоритет и выполняются в последнюю очередь.

Примеры.
Высказывание $(\neg P) \vee Q$ обычно записывают без скобок: $\neg P \vee Q$.
Это высказывание есть дизъюнкция двух высказываний: первое является отрицанием P, а второе — Q.
Высказывание $\neg(P \vee Q)$ есть отрицание дизъюнкции высказываний P и Q.

Высказывание

$$\neg P \wedge Q \vee \neg Q \wedge P \Rightarrow \neg Q$$

после расстановки скобок в соответствии с приоритетами примет вид

$$(((\neg P) \wedge Q) \vee ((\neg Q) \wedge P)) \Rightarrow (\neg Q).$$

Для определения истинности или ложности сложного высказывания в зависимости от истинности или ложности входящих в него высказываний используют **таблицы истинности**.

В первых двух столбцах таблицы записывают все возможные наборы значений, которые могут принимать высказывания P и Q.

Истинность высказывания обозначают буквой „И“, а ложность — буквой „Л“. Остальные столбцы заполняют слева направо.

Так для каждого набора значений P и Q находят соответствующие значения высказываний.

Таблицы истинности логических операций

P	Q	$P \vee Q$
Л	Л	Л
Л	И	И
И	Л	И
И	И	И

P	Q	$P \wedge Q$
Л	Л	Л
Л	И	Л
И	Л	Л
И	И	И

P	Q	$P \Rightarrow Q$
Л	Л	И
Л	И	И
И	Л	Л
И	И	И

P	Q	$P \Leftrightarrow Q$
Л	Л	И
Л	И	Л
И	Л	Л
И	И	И

P	$\neg P$
Л	И
И	Л

Рассмотрим сложное высказывание

$$(\neg P \wedge Q) \Rightarrow (\neg Q \wedge P).$$

Обозначим высказывание $\neg P \wedge Q$ через A, высказывание $\neg Q \wedge P$ через B, а исходное высказывание запишем в виде $A \Rightarrow B$.
Таблица истинности этого высказывания состоит из следующих столбцов. P, Q, $\neg P$, $\neg Q$, A, B и $A \Rightarrow B$.

P	Q	$\neg P$	$\neg Q$	A	B	$A \Rightarrow B$
Л	Л	И	И	Л	Л	И
Л	И	И	Л	И	Л	Л
И	Л	Л	И	Л	И	И
И	И	Л	Л	Л	Л	И

Сложные высказывания образуются не только посредством логических связок, но и с помощью предикатов и кванторов.
Предикат есть высказывание, содержащее одно или несколько переменных.

Например, „x есть четное число“ или
„x есть студент МГТУ им. Баумана, поступивший в 2012 г.“.

В первом предикате x есть целочисленное переменное, во втором — переменное, пробегающее множество „человеческих индивидов“.

Предикаты, содержащие несколько переменных: „x делится на y“„x меньше y“.

Предикаты записывают в виде $P(x)$, $Q(x,y)$, $R(x,y,z)$, в скобках перечисляют все переменные, входящие в данный предикат.

Подставляя вместо каждого переменного, входящего в предикат $P(x_1,\ldots,x_n)$, конкретное значение, т.е. фиксируя значения $x_1 = a_1,\ldots,x_n = a_n$, где $a_1,\ldots,a_n$ — некоторые константы с соответствующей областью значений, получаем высказывание, не содержащее переменных.

Например,
„2 есть четное число“,

„Исаак Ньютон есть студент МГТУ им. Баумана, поступивший в 2012 г.“,

„5 делится на 7“.

В зависимости от того, истинно или ложно полученное таким образом высказывание, говорят, что предикат P выполняется или не выполняется на наборе значений переменных $x_1 = a_1, \ldots, x_n = a_n$.

Предикат, выполняющийся на любом наборе входящих в него переменных, называют **тождественно истинным**.

Предикат, не выполняющийся ни на одном наборе значений входящих в него переменных, — **тождественно ложным**.

Высказывание из предиката можно получать не только подстановкой значений его переменных, но и посредством кванторов. Вводят два *квантора — существования* и *всеобщности*, обозначаемые $\exists$ и $\forall$ соответственно.

Высказывание $(\forall x \in A)P(x)$ истинно, тогда и только тогда, когда предикат $P(x)$ выполняется для каждого значения переменного x .(читается: „для каждого элемента x , принадлежащего множеству A („для всех $x \in A$) истинно $P(x)$ “).

Высказывание $(\exists x \in A)P(x)$ истинно, тогда и только тогда, когда на некоторых значениях переменного x выполняется предикат $P(x)$.(читается: „существует (найдется) такой элемент x множества A , что истинно $P(x)$ “).

При образовании высказывания из предиката посредством квантора говорят, что переменное предиката связывается квантором.В общем случае используют формы высказываний вида

$$(Q_1 x_1 \in A_1)(Q_2 x_2 \in A_2) \dots (Q_n x_n \in A_n) P(x_1, x_2, \dots, x_n),$$

где вместо каждой буквы Q с индексом может быть подставлен любой из кванторов $\forall$ или $\exists$.

Например, высказывание $(\forall x \in A)(\exists y \in B) P(x, y)$ читается так: „для всякого $x \in A$ существует $y \in B$, такой, что истинно $P(x, y)$“.

Математические теоремы можно записать в подобной форме:

$$(\exists \lim_{x \to a} f(x) = b_1) \wedge (\exists \lim_{x \to a} g(x) = b_2) \Rightarrow (\exists \lim_{x \to a} f(x) g(x) = b_1 b_2)$$

Множество полностью определяется своими элементами.

Способы задания множеств

Для конечного множества, число элементов которого относительно невелико, может быть использован способ непосредственного перечисления элементов.

Элементы конечного множества перечисляют в фигурных скобках в произвольном фиксированном порядке $\{1, 3, 5\}$.

Поскольку множество полностью определено своими элементами, то при задании конечного множества порядок, в котором перечислены его элементы, не имеет значения.

Записи $\{1, 3, 5\}$, $\{3, 1, 5\}$, $\{5, 3, 1\}$ и т.д. все задают одно и то же множество.

В общем случае для конечного множества используют форму записи $\{a_1, \ldots, a_n\}$.
Тогда конечное множество, заданное записью $\{a_1, \ldots, a_n\}$, состоит из n элементов. Его называют также **n -элементным множеством**.

Способ задания множества путем непосредственного перечисления его элементов применим в весьма узком диапазоне конечных множеств.

Наиболее общим способом задания конкретных множеств является указание некоторого свойства, которым должны обладать все элементы описываемого множества, и только они.

Пусть переменное x пробегает некоторое множество U, называемое *универсальным множеством*.
Свойство, которым обладают исключительно элементы данного множества A, может быть выражено посредством предиката $P(x)$, выполняющегося тогда и только тогда, когда переменное x принимает произвольное значение из множества A.
$P(x)$ истинно тогда и только тогда, когда вместо x подставляется константа $a \in A$.

Предикат P называют в этом случае **характеристическим предикатом** множества A, а свойство, выражаемое с помощью этого предиката, — **характеристическим свойством** или **коллективизирующим свойством**. Множество, заданное через характеристический предикат, записывается в следующей форме:

$$A = \{x\colon P(x)\}\,.$$

Например,

$$A = \{x\colon x \text{ есть четное натуральное число}\}$$

означает, что „ A есть множество, состоящее из всех таких элементов x , что каждое из них есть четное натуральное число“

Предикат, задающий коллективизирующее свойство, может быть тождественно ложным.
Множество, определенное таким образом, не будет иметь ни одного элемента.
Его называют *пустым множеством* и обозначают $\varnothing$.

Тождественно истинный характеристический предикат задает универсальное множество.

Конкретное содержание понятия универсального множества определяется решаемой задачей .

Зафиксировав универсальное множество, мы тем самым фиксируем область значений всех фигурирующих в наших математических рассуждениях переменных и констант.

В этом случае можно не указывать в кванторах то множество, которое пробегает связываемое квантором переменное.

Операции над множествами

Для любых двух множеств A и B определены новые множества, называемые **объединением**, **пересечением**, **разностью** и **симметрической разностью**.

- $A \cup B = \{x \,|\, x \in A \vee x \in B\}$ — объединение A и B есть множество всех таких x, что x является элементом хотя бы одного из множеств A, B;
- $A \cap B = \{x \,|\, x \in A \wedge x \in B\}$ — пересечение A и B есть множество всех таких x, что x — одновременно элемент A и элемент B;

- $A \setminus B = \{x \mid x \in A \wedge x \notin B\}$ — разность A и B есть множество всех таких x, что x — элемент A, но не элемент B ($x \notin B$);
- $A \triangle B = (A \setminus B) \cup (B \setminus A)$, а симметрическая разность A и B — множество всех таких x, что x — элемент A, но не элемент B или x — элемент B, но не элемент A.

Фиксируя универсальное множество U, мы можем определить **дополнение** $\overline{A}$ *множества* A.Дополнение множества A — это множество всех элементов универсального множества, не принадлежащих A.
$\overline{A} = U \setminus A$.

Пусть множество A задано посредством характеристического предиката P

$A = \{x\colon P(x)\}$,

а множество B — посредством характеристического предиката Q

$B = \{x\colon Q(x)\}$.

Тогда

$$A \cup B = \{x\colon P(x) \vee Q(x)\}.$$

$$A \cap B = \{x\colon P(x) \wedge Q(x)\}.$$

$$A \setminus B = \{x\colon P(x) \wedge \neg Q(x)\}.$$

Подмножества

Множество B есть подмножество множества A, если всякий элемент B есть элемент A:

$$B \subseteq A$$

Говорят также, что B содержится в A, B включено в A, A включает B (имеет место **включение** $B \subseteq A$).

Пустое множество есть подмножество любого множества.
Если фиксировано некоторое универсальное множество, каждое рассматриваемое множество есть его подмножество.

Если $A = \{x\colon P(x)\}$, $B = \{x\colon Q(x)\}$, то $B \subseteq A$ тогда и только тогда, когда высказывание $Q(x) \Rightarrow P(x)$ тождественно истинно.

Два множества A и B считают **равными**, если любой элемент x множества A является элементом множества B и наоборот.
Множество A равно множеству B тогда и только тогда, когда A есть подмножество B и наоборот, т.е.

$$A = B \Leftrightarrow ((A \subseteq B) \wedge (B \subseteq A)). \tag{1.1}$$

Формула (1.1) является основой для построения доказательств о равенстве множеств.

Чтобы доказать равенство двух множеств X и Y, т.е. что $X = Y$, достаточно доказать два включения $X \subseteq Y$ и $Y \subseteq X$.
1. Доказать, что из предположения $x \in X$ (для произвольного x) следует, что $x \in Y$.
2. Доказать, что из предположения $x \in Y$ следует, что $x \in X$.

Такой метод доказательства теоретико-множественных равенств называют **методом двух включений**.

Если $B \subseteq A$, но $B \neq A$, то пишут $B \subset A$ и B называют **строгим подмножеством**
(или *собственным подмножеством*) множества A, а символ $\subset$ — **символом строгого включения**.

.

Для всякого множества A может быть образовано множество всех подмножеств множества A.
Его обозначают 2^A:

$$2^A = \{X\colon X \subseteq A\}.$$

Это множество часто называют **булеаном множества** A
Булеан множества $\{a, b\}$ состоит из четырех множеств $\varnothing$, $\{a\}$, $\{b\}$, $\{a, b\}$, т.е. $2^{\{a, b\}} = \{\varnothing, \{a\}, \{b\}, \{a, b\}\}$.
Булеан $2^{\mathbb{N}}$ состоит из всех возможных, конечных или бесконечных, подмножеств множества $\mathbb{N}$.
Так, $\varnothing \in 2^{\mathbb{N}}$, $\{5\} \in 2^{\mathbb{N}}$, вообще для любого n множество $\{n\} \in 2^{\mathbb{N}}$, множество $\{1, \ldots, n\} \in 2^{\mathbb{N}}$, $\{n\colon n = 2k,\ k = 1, 2, \ldots\} \in 2^{\mathbb{N}}$ и т.п.

Теоретико-множественные тождества.

1. $A \cup B = B \cup A$;
2. $A \cap B = B \cap A$;
3. $A \cup (B \cup C) = (A \cup B) \cup C$;
4. $A \cap (B \cap C) = (A \cap B) \cap C$;
5. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$;
6. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$;
7. $\overline{A \cup B} = \overline{A} \cap \overline{B}$;
8. $\overline{A \cap B} = \overline{A} \cup \overline{B}$;

9. $A \cup \varnothing = A$;
10. $A \cap \varnothing = \varnothing$;
11. $A \cap U = A$;
12. $A \cup U = U$;
13. $A \cup \overline{A} = U$;
14. $A \cap \overline{A} = \varnothing$;
15. $A \cup A = A$;
16. $A \cap A = A$;
17. $\overline{\overline{A}} = A$;

18. $A \setminus B = A \cap \overline{B}$;

19. $A \triangle B = (A \cup B) \setminus (A \cap B)$;

20. $(A \triangle B) \triangle C = A \triangle (B \triangle C)$;

21. $A \triangle B = B \triangle A$;

22. $A \cap (B \triangle C) = (A \cap B) \triangle (A \cap C)$.

Докажем тождество $A \cap (B \cap C) = (A \cap B) \cap C$;
Покажем первое включение

$$
\begin{aligned}
x \in (A \cap (B \cap C)) \Rightarrow \\
&\Rightarrow (x \in A) \wedge (x \in (B \cap C)) \Rightarrow \\
&\Rightarrow (x \in A) \wedge ((x \in B) \wedge (x \in C)) \Rightarrow \\
&\Rightarrow ((x \in A) \wedge (x \in B)) \wedge (x \in C) \Rightarrow \\
&\Rightarrow (x \in (A \cap B)) \wedge (x \in C) \Rightarrow \\
&\Rightarrow x \in ((A \cap B) \cap C)
\end{aligned}
$$

Первое включение установлено.

Покажем второе включение.

$$
\begin{aligned}
& x \in ((A \cap B) \cap C) \\
& \qquad \Rightarrow (x \in (A \cap B)) \wedge (x \in C) \Rightarrow \\
& \qquad \Rightarrow ((x \in A) \wedge (x \in B)) \wedge (x \in C) \Rightarrow \\
& \qquad \Rightarrow (x \in A) \wedge ((x \in B) \wedge (x \in C)) \Rightarrow \\
& \qquad \Rightarrow (x \in A) \wedge (x \in (B \cap C)) \Rightarrow \\
& \qquad \Rightarrow x \in (A \cap (B \cap C)) \Rightarrow
\end{aligned}
$$

Оба включения имеют место, тождество доказано.

Для доказательства теоретико-множественных тождеств могут быть использованы:
метод характеристических функций (см. семинар 1);
метод эквивалентных преобразований.

При доказательстве теоретико-множественных тождеств **методом эквивалентных преобразований** используются ранее доказанные тождества для преобразования левой части к правой или наоборот.

Докажем методом эквивалентных преобразований тождество

$$A \setminus (B \cup C) = (A \setminus B) \setminus C,$$

пользуясь тождествами 1–19. Преобразуем левую часть к правой:

$$\begin{aligned} A \setminus (B \cup C) = \\ &= A \cap \overline{(B \cup C)} = \\ &= A \cap (\overline{B} \cap \overline{C}) = \\ &= (A \cap \overline{B}) \cap \overline{C} = \\ &= (A \setminus B) \cap \overline{C} = \\ &= (A \setminus B) \setminus C \end{aligned}$$

Тождество доказано.

ДИСКРЕТНАЯ МАТЕМАТИКА

ИУ5 - 4 семестр
2015

Лекция 2.

2.1. Кортеж. Декартово произведение

Упорядоченная пара $(a,\, b)$ на множествах A и B, определяется не только самими элементами $a \in A$ и $b \in B$, но и порядком, в котором они записаны.
Если $A = B$, то говорят об упорядоченной паре на множестве A.

Определение 2.1. Две упорядоченные пары $(a,\, b)$ и $(a',\, b')$ на множествах A и B называют **равными**, если $a = a'$ и $b = b'$.

Обобщением понятия упорядоченной пары является **упорядоченный n -набор**, или *кортеж*.
Кортеж $(a_1, \ldots, a_n)$ на множествах $A_1, \ldots, A_n$ характеризуется не только входящими в него элементами $a_1 \in A_1, \ldots, a_n \in A_n$, но и порядком, в котором они перечисляются.
Роль порядка в кортеже фиксируется определением равенства кортежей.

Определение 2.2. Два кортежа $(a_1, \ldots, a_n)$ и $(b_1, \ldots, b_n)$ на множествах $A_1, \ldots, A_n$ равны, если $a_i = b_i$, $i = \overline{1, n}$.

Число n называется **длиной кортежа** (или **размерностью кортежа**), а элемент a_i — i -й **проекцией** (компонентой) **кортежа**.
Для двух кортежей одинаковой размерности их компоненты с одинаковыми номерами называют **одноименными компонентами**.

Определение 2.3. Множество всех кортежей длины n на множествах $A_1, \ldots, A_n$ называют **декартовым (прямым) произведением множеств** $A_1, \ldots, A_n$ и обозначают $A_1 \times \ldots \times A_n$.

$$A_1 \times \ldots \times A_n = \{(a_1, \ldots, a_n) \colon a_1 \in A_1, \ldots, a_n \in A_n\}.$$

Если все множества A_i, $i = \overline{1, n}$, равны между собой, то указанное декартово произведение называют n-й **декартовой степенью множества** A и обозначают A^n.
При $n = 2$ получаем **декартов квадрат**, при $n = 3$ — **декартов куб** множества A.

Первая декартова степень любого множества A есть само множество A, т.е. $A^1 = A$.

Свойства декартова произведения :

- $A \times (B \cup C) = (A \times B) \cup (A \times C)$;
- $A \times (B \cap C) = (A \times B) \cap (A \times C)$;
- $A \times \varnothing = \varnothing \times A = \varnothing$.

Докажем первое тождество *методом двух включений*.

$$
\begin{aligned}
(x,\, y) \in A \times (B \cup C) \Rightarrow \\
&\Rightarrow (x \in A \wedge (y \in B \cup C)) \\
&\Rightarrow (x \in A \wedge (y \in B \vee y \in C)) \\
&\Rightarrow ((x \in A \wedge y \in B) \vee (x \in A \wedge y \in C)) \\
&\Rightarrow ((x,\, y) \in A \times B \vee (x,\, y) \in A \times C) \\
&\Rightarrow ((x,\, y) \in (A \times B) \cup (A \times C))
\end{aligned}
$$

Следовательно, $A \times (B \cup C) \subseteq (A \times B) \cup (A \times C)$.
Доказательство обратного включения аналогично.

2.2. Соответствия и бинарные отношения

Отображение f из множества A в множество B $(f\colon A \to B)$ считается заданным, если каждому элементу $x \in A$ сопоставлен **единственный** элемент $y \in B$.

Элемент $y \in B$, который отображением f сопоставляется элементу $x \in A$, называют **образом элемента** x **при отображении** f и обозначают $f(x)$.

Каждое отображение **однозначно** определяет множество **упорядоченных пар** $\{(x, y)\colon x \in A,\ y = f(x)\}$, являющееся подмножеством *декартова произведения* $A \times B$ множества A на множество B и называемое **графиком отображения** f.

Обратно, если в декартовом произведении $A \times B$ фиксировано подмножество упорядоченных пар f, такое, что для любых двух пар (x, y) и (x', y') множества f из $x = x'$ следует равенство $y = y'$, то f единственным образом определяет некоторое отображение из A в B.

Отображение f, элементу $x \in A$ сопоставляет такой элемент $y \in B$, что $(x, y) \in f$. Можем отождествить отображения и их графики.

Отображение есть подмножество декартова произведения.
Отображение f множества A в себя называют **тождественным**, если $f(x) = x$ при всех x из A.
В общем случае для отображения $f\colon A \to B$ может существовать несколько различных элементов множества A, образы которых совпадают.
Множество всех элементов $x \in A$, для которых $f(x) = y_0$, называют **прообразом элемента** $y_0 \in B$ **при отображении** f.
Прообраз числа a, $|a| \leq 1$, при отображении $y = \sin x$ есть множество всех решений уравнения $\sin x = a$, т.е. множество

$$\{x\colon x = \arcsin a + 2\pi n,\ n \in \mathbb{Z}\} \cup \{x\colon x = \pi - \arcsin a + 2\pi n,\ n \in \mathbb{Z}\}.$$

Прообраз элемента $y_0 \in B$ может быть *пустым множеством*. Например, для числа 2 при отображении $y = \sin x$.

Множество A называют **областью определения отображения** f.
Область определения отображения f будем обозначать $D(f)$

Множество всех $y \in B$, таких, что найдется $x \in A$, для которого $y = f(x)$, называют **областью значений отображения** f.

Область значений отображения f будем обозначать $R(f)$.

Отображение $f\colon A \to B$ называют **инъективным** (**инъекцией**), если каждый элемент из области его значений имеет единственный прообраз, т.е. из $f(x_1) = f(x_2)$ следует $x_1 = x_2$.

Отображение $f\colon A \to B$ называют **сюръективным** (**сюръекцией**), если его область значений совпадает со всем множеством B .

Сюръективное отображение из A в B называют также **отображением** множества A **на** множество B .

Отображение $f\colon A \to B$ называют **биективным** (**биекцией**), если оно одновременно инъективно и сюръективно.

Если отображение $f\colon A \to B$ биективно, то каждому элементу множества A отвечает единственный элемент множества B и наоборот. Тогда говорят, что множества A и B находятся между собой во **взаимно однозначном соответствии**.

Пример 2.1.

а. Отображение, заданное равенством $\nu(n) = n + 1$, есть биекция множества натуральных чисел $\mathbb{N}$ на его подмножество $\mathbb{N} \setminus \{1\}$.

б. Отображение $\nu(n) = 2n$ есть биекция множества всех натуральных чисел на множество всех четных натуральных чисел.

в. Любая *показательная функция* $y = a^x$, $a > 0$, есть биекция множества $\mathbb{R}$ всех действительных чисел на множество $\mathbb{R}^+$ всех положительных действительных чисел.

г. Функция $y = \operatorname{arctg} x$ есть биекция множества $\mathbb{R}$ на интервал $(-\pi/2, \pi/2)$.

Пусть задано отображение $f\colon A \to B$ и $C \subseteq A$ — некоторое множество.
Множество $f(C)$ элементов $y \in B$, таких, что $y = f(x)$, $x \in C$, называют **образом множества** C при отображении f.
Например, при отображении $y = \sin x$ отрезок $[0,\, 1]$ является образом множества (отрезка) $[0,\, \pi]$ (и любого объединения отрезков вида $[2\pi k,\, (2k+1)\pi]$ (для произвольного целого k)).
При $k = 0$ это можно записать следующим образом: $\sin([0,\, \pi]) = [0,\, 1]$.

Для любого отображения $f\colon A \to B$ образ $f(A)$ всего множества A есть область значений данного отображения.

Для произвольного множества $D \subseteq B$ множество всех элементов $x \in A$, таких, что $f(x) \in D$, называют **прообразом множества** D при отображении f.

Например, для любого действительного числа $a \in [0, 1)$ множество, которое является объединением всех отрезков вида $[\arcsin a + 2\pi k,\ \pi - \arcsin a + 2\pi k]$, $k \in \mathbb{Z}$, есть прообраз отрезка $[a, 1]$ при отображении $y = \sin x$.

Прообраз области значений произвольного отображения $f\colon A \to B$ совпадает со всем множеством A.

Множество всех отображений из A в B будем обозначать как B^A.

Понятие отображения можно обобщить.

1.**Частичное отображение**.
Пусть образ определен не для каждого элемента множества A, а для некоторых элементов этого множества (отказ от полной определенности отображения).
Мы пришли к понятию **частичного отображения**.
При этом подмножество всех элементов A, для которых определен образ, называют **областью определения** данного **частичного отображения**.

2.**Соответствие**.
Пусть данному $x \in A$ сопоставлен не один, а несколько образов (множество образов) в множестве B (отказ от однозначности отображения).
В этом случае говорят, что задано **соответствие**
Соответствие ρ из A в B будем обозначать $\rho(x)$ (по аналогии с обозначением $f(x)$ для отображений).
$\rho(x)$ есть элемент **подмножества** B .

График соответствия

Графиком соответствия ρ из множества A в множество B называется множество C_ρ упорядоченных пар (x, y), таких, что $x \in A$, $y \in B$ и элементы x, y связаны соответствием ρ, т.е. $y \in \rho(x)$.

Указанное множество C_ρ упорядоченных пар есть подмножество декартова произведения $A \times B$.

Обратно, фиксируя на декартовом произведении $A \times B$ какое-либо подмножество C, мы однозначно определяем некоторое соответствие ρ_C из A в B, а именно $\rho_C(x) = \{y\colon y \in B \wedge (x, y) \in C\}$.

Графиком соответствия ρ_C будет множество C, а соответствием, отвечающим графику C_ρ, будет ρ.

Можно отождествить соответствие с его графиком и считать, что соответствие из множества A в множество B есть некоторое подмножество ρ декартова произведения $A \times B$, т.е. $\rho \subseteq A \times B$.
При $\rho = \varnothing$ получаем **пустое соответствие**.
При ρ, совпадающем со всем указанным декартовым произведением, — **универсальное соответствие**.
$(x,\ y) \in \rho$ –упорядоченные пары, связанных соответствием ρ .

Область определения соответствия $\rho \subseteq A \times B$ из множества A в множество B — это множество всех первых компонент упорядоченных пар из ρ:

$$D(\rho) = \{x\colon (\exists y \in B)(x,\, y) \in \rho\}\,.$$

Область значения соответствия ρ — это множество всех вторых компонент упорядоченных пар из ρ:

$$R(\rho) = \{y\colon (\exists x \in A)(x,\, y) \in \rho\}\,.$$

$D(\rho) \subseteq A\,,\ R(\rho) \subseteq B\,.$

Соответствие из A в B называют **всюду определенным**, если его область определения совпадает с множеством A: $D(\rho) = A\,.$

Сечением соответствия $\rho \subseteq A \times B$ для фиксированного элемента $x \in A$ называют множество $\rho(x) = \{y\colon (x, y) \in \rho\}$.
Сечение соответствия $\rho(x)$ есть множество всех „образов“ элемента x при данном соответствии.

Соответствие $\rho \subseteq A \times A$ из множества A в себя, т.е. подмножество множества A^2, называют **бинарным отношением на множестве** A.

Пример 2.2. Отношение нестрогого неравенства на множестве действительных чисел $\mathbb{R}$. Здесь каждому $x \in \mathbb{R}$ поставлены в соответствие такие $y \in \mathbb{R}$, для которых справедливо $x \leq y$.

Для произвольного бинарного отношения на некотором множестве часто используют запись $x \rho y$ вместо $(x, y) \in \rho$, говоря при этом об **элементах, связанных бинарным отношением** ρ.

Например, $x \leq y$, а не $(x, y) \in \leq$.

Бинарное отношение на множестве A, состоящее из всех пар (x, x), т.е. пар с совпадающими компонентами, называют **диагональю** множества A и обозначают id_A.
Диагональ A есть **тождественное отображение** A на себя.

Для наглядного изображения соответствий из A в B будем использовать два способа.

1. **График соответствия**

Соответствие интерпретируется как подмножество декартова произведения и изображается на плоскости как подмножество декартова квадрата числовых множеств.

2.**Граф соответствия**

Для конечных множеств A и B ,применяется построение **графа соответствия**.

Пример 2.3. Множество точек окружности $x^2 + y^2 = 1$ есть график бинарного отношения на множестве действительных чисел, состоящего из всех таких упорядоченных пар (x, y), что $y = \pm\sqrt{(1 - x^2)}$, (компоненты пары удовлетворяют уравнению $x^2 + y^2 = 1$.)
Область определения бинарного отношения есть отрезок $[-1, 1]$, область значения — также отрезок $[-1, 1]$.

Соответствие $\rho \subseteq A \times B$ называют **функциональным по** второй (первой) **компоненте**, если для любых двух упорядоченных пар $(x, y) \in \rho$ и $(x', y') \in \rho$ из равенства $x = x'$ следует $y = y'$ (и из $y = y'$ следует $x = x'$).

Функциональность соответствия по второй компоненте означает, что, фиксируя в любой упорядоченной паре, принадлежащей данному соответствию, первую компоненту, мы однозначно определяем и вторую компоненту.

Соответствие, **функциональное** по **второй компоненте**, есть **отображение** (возможно, частичное).
Соответствие $f \subseteq A \times B$ является отображением из A в B, если и только если оно всюду определено (т.е. $D(f) = A$) и функционально по второй компоненте.
Отображение из A в B является инъекцией тогда и только тогда, когда оно функционально по первой компоненте.

Определение 2.4. Произвольное подмножество ρ декартова произведения $A_1 \times \ldots \times A_n$ называют (**n -арным** или **n -местным) отношением** на множествах $A_1, \ldots, A_n$.
В случае если все множества $A_1, \ldots, A_n$ совпадают, т.е. $A_1 = \ldots = A_n = A$, говорят об **n -арном отношении на множестве** A.

Если ρ — n-арное отношение на множествах $A_1, \ldots, A_n$ и $(a_1, \ldots, a_n) \in \rho$, то говорят об **элементах** $a_1, \ldots, a_n$, **связанных отношением** ρ.
При $n = 2$ получаем **бинарное отношение** на множествах A_1, A_2.
Это соответствие из A_1 в A_2, где множества A_1 и A_2 различны.
При $A_1 = A_2 = A$ получаем введенное ранее бинарное отношение на множестве, т.е. подмножество декартова квадрата A.
В общем случае (при произвольном $n \geq 2$) следует, сторого говоря, различать термины „n-арное отношение“ и „n-арное отношение на множестве“.

Пусть n-арное отношение $\rho \subseteq A_1 \times \ldots \times A_n$ удовлетворяет условию: для любых двух кортежей $(x_1, \ldots, x_i, \ldots, x_n) \in \rho$ и $(y_1, \ldots, y_i, \ldots, y_n) \in \rho$ из выполнения равенств $x_k = y_k$ для любого $k \neq i$ ($0 \leq k \leq n$) следует, что и $x_i = y_i$.

Тогда отношение ρ называют **функциональным по** i-й **компоненте** $(1 \leq i \leq n)$.

Функциональность n-местного отношения по i-й $(i \leq n)$ компоненте равносильна условию, что, фиксируя все компоненты, кроме i-й, мы однозначно определяем и i-ю компоненту.

Пример 2.4. Рассмотрим на множестве V_3 геометрических векторов в пространстве тернарное (трехместное) отношение ρ, состоящее из всех упорядоченных троек $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ компланарных векторов.

Это отношение не является функциональным ни по одной компоненте, так как любым двум векторам соответствует бесконечно много векторов, образующих с ними компланарную тройку.

2.3. Операции над соответствиями

Поскольку **соответствия** можно считать множествами, то все операции над множествами (**пересечение, объединение, разность, дополнение** и т.д.) можно применить и к соответствиям.

Говоря о дополнении соответствия из A в B, мы имеем в виду дополнение до **универсального соответствия** из A в B, т.е. до **декартова произведения** $A \times B$.

Равенство соответствий можно трактовать как **равенство множеств**.
На соответствия можно распространить операции, определяемые для отображений.

Композиция соответствий

Композицией (произведением) соответствий $\rho \subseteq A \times B$ и $\sigma \subseteq B \times C$ называют соответствие

$$\rho \circ \sigma = \{(x,\, y)\colon (\exists z \in B)((x,\, z) \in \rho) \wedge ((z,\, y) \in \sigma)\}\,. \qquad (2.1)$$

Пример 2.5. Соответствие ρ задано следующим образом:
Есть множество программистов $A = \{И,\ П,\ С\}$ и множество программ $B = \{n_1,\ n_2,\ n_3,\ n_4,\ n_5\}$.
Соответствие ρ из A в B, связывает программистов и разрабатываемые ими программы:

$$\rho = \{(И,\ n_1),\ (И,\ n_3),\ (И,\ n_5),(П,\ n_2),\ (П,\ n_4),\ (С,\ n_2),\ (С,\ n_5)\}.$$

Соответствие σ зададим как соответствие из множества программ $\{n_1,\ n_2,\ n_3,\ n_4,\ n_5\}$ в множество заказчиков ПО $\{З_1,\ З_2,\ З_3,\ З_4\}$.
Пусть

$$\sigma = \{(n_1,\ З_3),\ (n_1,\ З_4),\ (n_2,\ З_1),\ (n_3,\ З_2),\ (n_4,\ З_4),\ (n_5,\ З_3)\}.$$

Построим композицию соответствий ρ и σ.

Имеем
$\rho(И) = \{n_1,\ n_3,\ n_5\}$,
$\sigma(n_1) = \{З_3,\ З_4\}$,
$\sigma(n_3) = \{З_2\}$,
$\sigma(n_5) = \{З_3\}$.

Получаем $\sigma(n_1) \cup \sigma(n_3) \cup \sigma(n_5) = \{З_2,\ З_3,\ З_4\}$ сечение композиции по элементу $И$.
Аналогично, получим
$(\rho \circ \sigma)(П) = \{З_1,\ З_4\}$
и $(\rho \circ \sigma)(С) = \{З_1,\ З_3\}$.

Операция композиции соответствий $\rho \subseteq A \times B$ и $\sigma \subseteq C \times D$ **не коммутативна**.

В общем случае $\rho \circ \sigma \neq \sigma \circ \rho$, поскольку $\rho \circ \sigma \subseteq A \times D$, а $\sigma \circ \rho \subseteq C \times B$.

Бинарное отношение на множестве является частным случаем соответствия. Для двух бинарных отношений ρ и σ, заданных на множестве A, их композиция $\rho \circ \sigma$ (2.1) как соответствий является бинарным отношением на том же множестве A.

В этом случае говорят о **композиции бинарных отношений на множестве** A.

Композицию $\rho \circ \rho$ бинарного отношения ρ на некотором множестве с самим собой называют **квадратом бинарного отношения** ρ и обозначают ρ^2 .

В общем случае для двух бинарных отношений τ и φ также имеет место неравенство

$$\tau \circ \varphi \neq \varphi \circ \tau,$$

хотя обе композиции заданы на одном и том же множестве.

Пример 2.6.
Зададим на множестве $A = \{1,\ 2,\ 3,\ 4\}$ бинарные отношения
$\tau = \{(x,\ y)\colon x + 1 < y\}$,
$\varphi = \{(x, y)\colon |x - y| = 2\}$
Найдем композицию $\tau \circ \varphi$.
Имеем $\tau(1) = \{3,\ 4\}$, $\varphi(3) = \{1\}$ и $\varphi(4) = \{2\}$.
Следовательно, $(\tau \circ \varphi)(1) = \varphi(3) \cup \varphi(4) = \{1,\ 2\}$.
Далее $\tau(2) = \{4\}$, $\varphi(4) = \{2\}$ и $(\tau \circ \varphi)(2) = \{2\}$.
Так как $\tau(3) = \tau(4) = \varnothing$, то в итоге получим

$$\tau \circ \varphi = \{(1,\ 1),\ (1,\ 2),\ (2,\ 2)\}.$$

Свойства композиции соответствий.

1) $\rho \circ (\sigma \circ \tau) = (\rho \circ \sigma) \circ \tau$;

2) для любого соответствия ρ имеет место $\rho \circ \varnothing = \varnothing \circ \rho = \varnothing$;

3) $\rho \circ (\sigma \cup \tau) = (\rho \circ \sigma) \cup (\rho \circ \tau)$;

4) для любого бинарного отношения на множестве A имеет место равенство $\rho \circ \mathrm{id}_A = \mathrm{id}_A \circ \rho = \rho$.

5) $\rho \circ (\sigma \cap \tau) \subseteq \rho \circ \sigma \cap \rho \circ \tau$, обратное включение в общем случае не имеет места.

Роль **пустого соответствия** в операции композиции, определенной на множестве всех бинарных отношений на A, аналогична роли нуля при умножении чисел.

Диагональ множества A играет роль, аналогичную роли единицы.

Первые четыре свойства можно доказать *методом двух включений*.

Обратное соответствие

Соответствие, **обратное** к соответствию $\rho \subseteq A \times B$, есть соответствие из B в A, обозначаемое ρ^{-1} и равное, по определению,

$$\rho^{-1} = \{(y,\, x)\colon (x,\, y) \in \rho\}\,.$$

Пример 2.7.
Соответствие ρ, задано следующим образом:
Соответствие ρ из A в B, связывает множество программистов $A = \{И, П, С\}$ и разрабатываемые ими программы ($B = \{n_1, n_2, n_3, n_4, n_5\}$):

$$\rho = \{(И, n_1), (И, n_3), (И, n_5), (П, n_2), (П, n_4), (С, n_2), (С, n_5)\}.$$

Обратное соответствие

$$\rho^{-1} = \{(n_1, И), (n_2, П), (n_2, С), (n_3, И), (n_4, П), (n_5, И), (n_5, С)\}.$$

Свойства обратного соответствия

1) $(\rho^{-1})^{-1} = \rho$;
2) $(\rho \circ \sigma)^{-1} = \sigma^{-1} \circ \rho^{-1}$.

Для бинарного отношения ρ на множестве A обратное соответствие есть бинарное отношение на том же множестве.

В этом случае говорят о **бинарном отношении** ρ^{-1} **на множестве** A, **обратном** к ρ.

Соответствия $\rho \circ \rho^{-1}$ и $\rho^{-1} \circ \rho$ в общем случае не совпадают.
Для бинарного отношения ρ на множестве A $\rho \circ \rho^{-1} \neq \rho^{-1} \circ \rho$
$\rho \circ \rho^{-1} \neq \mathrm{id}_A$ и $\rho^{-1} \circ \rho \neq \mathrm{id}_A$.

Ограничение отношения.

Любое бинарное отношение заданное на множестве A задает бинарное отношение на любом подмножестве B множества A.

Это отношение называется **сужением отношения** ρ **на подмножество** C и обозначается обозначаемое $\rho|_B$,

2.4. Специальные свойства бинарных отношений

Бинарное отношение ρ на множестве A называют **рефлексивным**, если *диагональ* множества A *содержится в* ρ: $\mathrm{id}_A \subseteq \rho$, т.е. $x \rho x$ для любого элемента x множества A.

Если же $\mathrm{id}_A \cap \rho = \varnothing$, то бинарное отношение ρ на множестве A называют **иррефлексивным**.

Указанные свойства бинарных отношений на множестве A называют **рефлексивностью** и **иррефлексивностью**.

Иррефлексивное отношение нерефлексивно, но не всякое нерефлексивное отношение иррефлексивно.

Иррефлексивному отношению на A не принадлежит ни один элемент диагонали id_A, а нерефлексивное отношение может содержать некоторые (но не все!) элементы диагонали.

Примеры рефлексивных бинарных отношений.
Бинарные отношения равенства и подобия на множестве геометрических фигур,все отношения равенства, нестрогого неравенства на множестве действительных чисел, **отношение** $\subseteq$ **включения** множеств.

Примеры иррефлексивных бинарных отношений.
Бинарное отношение на множестве действительных чисел, задаваемое строгим неравенством $x < y$, отношение $\subset$ *строгого включения* множеств.

Бинарное отношение ρ на множестве A называют:
1) **симметричным**, если для любых $x, y \in A$ из $x \rho y$ следует $y \rho x$;

2) **антисимметричным**, если для любых $x, y \in A$ из $x \rho y$ и $y \rho x$ следует, что $x = y$.

Соответствующие свойства бинарных отношений на множестве A называют **симметричностью** и **антисимметричностью**.

График симметричного бинарного отношения на множестве A симметричен относительно диагонали

Теорема 1. Бинарное отношение ρ на множестве A симметрично, если и только если *бинарное отношение на множестве* A, *обратное* к ρ, совпадает с ρ: $\rho^{-1} = \rho$.

Теорема 2. Бинарное отношение ρ на множестве A антисимметрично тогда и только тогда, когда $\rho \cap \rho^{-1} \subseteq \mathrm{id}_A$.

Для антисимметричного бинарного отношения на множестве A может иметь место равенство $\rho \cap \rho^{-1} = \varnothing$.

Все бинарные отношения в геометрии типа равенства или подобия симметричны.

Если треугольник ABC подобен треугольнику $A'B'C'$, то и второй из этих треугольников подобен первому.

Бинарные отношения включения множеств ($A \subset B$, $A \subseteq B$), как строгие, так и не строгие, антисимметричны.

Бинарное отношение ρ на множестве A называют **транзитивным**, если для любых x, y, $z \in A$ из того, что $x \rho y$ и $y \rho z$, следует $x \rho z$. Соответствующее свойство бинарного отношения называют **транзитивностью**.

Пример 2.8. а. Пусть M — некоторое множество населенных пунктов. Зададим на нем бинарное отношение достижимости: из пункта A достижим пункт B, если есть дорога, по которой можно доехать из A в B. Это отношение транзитивно, поскольку если из пункта A можно доехать до пункта B, а из B есть дорога до C, то из A можно проехать в C.

б. Бинарное отношение неравенства на множестве действительных чисел не транзитивно, так как из того, что $x \neq y$ и $y \neq z$, вовсе не следует, что $x \neq z$.

Свойство транзитивного бинарного отношения.

Теорема 3. Бинарное отношение ρ на множестве A транзитивно тогда и только тогда, когда его *квадрат* содержится в нем, т.е. $\rho \circ \rho \subseteq \rho$ ($\rho \circ \rho = \rho^2$).

Данное свойство целесообразно использовать для проверки транзитивности бинарного отношения ρ на некотором множестве в тех случаях, когда построение квадрата ρ является более легкой задачей по сравнению с исследованием свойства транзитивности ρ на основе определения.

Материал для самостоятельного изучения

Построение композиции .

1. Композиция двух отображений (частный случай соответствий).

Пусть заданы отображения : f из A в B и g из B в C .

Композиция $f \circ g$ определяется как отображение из A в C , задаваемое формулой $y = g(f(x))$.

Тем самым задается **график отображения** $f \circ g$, т.е. множество **упорядоченных пар** $(x,\, y)$, таких, что $y = g(f(x))$.

При этом упорядоченная пара $(x,\, y)$ будет принадлежать графику отображения $f \circ g$,

если и только если найдется элемент $z \in B$, такой, что $z = f(x)$ и $y = g(z)$.

График композиции отображений f и g есть

$$f \circ g = \{(x,\, y)\colon (\exists z)(z = f(x) \text{ и } y = g(z))\} = \\ = \{(x,\, y)\colon y = g(f(x))\} . \quad (2.2)$$

При построении композиции отображений обычно предполагается, что пересечение области значений отображения f и области определения отображения не пусто ($R(f) \cap D(g) \neq \varnothing$), поскольку в противном случае композиция была бы пуста.

Для отображений, не являющихся частичными, $R(f) \subseteq D(g)$, так как $D(g) = B$. Поэтому в данном случае пересечение $R(f) \cap D(g)$ всегда не пусто.

Если f и g — биекции, то и композиция их тоже будет биекцией.

2.Композиция двух соответствий $\rho \circ \sigma$.

Возьмем произвольный элемент $x \in D(\rho)$. (область определения $D(\rho)$ соответствия ρ не пуста)

Пусть **сечение** $\rho(x) \subseteq B$ соответствия ρ не пусто и найдется такой элемент $z \in \rho(x)$, что сечение $\sigma(z) \subseteq C$ также не пусто.

Тогда непустое множество $\{(x,\, t)\colon t \in \sigma(z)\}$ будет подмножеством сечения соответствия $\rho \circ \sigma$ в точке x .

Сечением соответствия $\rho \circ \sigma$ в точке x будет непустое множество всех упорядоченных пар $(x,\, t) \in A \times C$ таких, что $x \in D(\rho)$, а $t \in \sigma(z)$ для некоторого $z \in \rho(x)$.

Нужно перебрать все элементы z из сечения $\rho(x)$.

Различие в построении композиции соответствий и композиции отображений заключается в том, что „промежуточный“ элемент z в общем случае не единственный и каждому такому элементу также ставится в соответствие не единственный элемент $y \in C$.

Обратное соответствие (дополнение).

Если $f\colon A \to B$ — отображение, то оно является соответствием. Обратное к f соответствие из B в A в общем случае не является отображением.

Соответствие f^{-1}, обратное к f, состоит из всех упорядоченных пар вида $(f(x), x)$, $x \in A$.

В общем случае могут найтись такие два различных элемента x и x', что $f(x) = f(x')$, то соответствие f^{-1} случае не будет *функционально по второй компоненте* и поэтому не будет отображением.

Если *отображение* f *инъективно*, то обратное соответствие есть *частичное отображение* из B в A.

Если **отображение** f **биективно**, то обратное соответствие является отображением из B в A, причем имеют место равенства

$$f \circ f^{-1} = \mathrm{id}_A, \quad f^{-1} \circ f = \mathrm{id}_B .$$

Отображение f^{-1} в этом случае называют **отображением**, **обратным** к f.

ТЕОРЕМА 1 . Бинарное отношение ρ на множестве A симметрично, если и только если *бинарное отношение на множестве* A, *обратное* к ρ, совпадает с ρ: $\rho^{-1} = \rho$.

◀ Пусть бинарное отношение ρ на множестве A симметрично
Докажем, что $\rho^{-1} = \rho$.

$$(x,\, y) \in \rho^{-1} \Rightarrow (y,\, x) \in \rho \Rightarrow$$
$$\text{(в силу симметричности } \rho : (x,\, y) \in \rho \Rightarrow (y,\, x) \in \rho)$$
$$\Rightarrow (x,\, y) \in \rho \Rightarrow \rho^{-1} \subseteq \rho$$

Аналогично доказывается включение $\rho \subseteq \rho^{-1}$.

Пусть $\rho = \rho^{-1}$.
Докажем, что бинарное отношение ρ на множестве A симметрично.

$$(x,\, y) \in \rho \Rightarrow (x,\, y) \in \rho^{-1} \Rightarrow$$

(по определению обратного отношения)

$$\Rightarrow (y,\, x) \in \rho.$$

Следовательно, ρ — симметричное бинарное отношение. ▶

ТЕОРЕМА 2. Бинарное отношение ρ на множестве A антисимметрично тогда и только тогда, когда $\rho \cap \rho^{-1} \subseteq \mathrm{id}_A$.

◀ Пусть отношение ρ на множестве A антисимметрично,
т.е. $\forall (x,\, y) \in A : (x,\, y) \in \rho \wedge (y,\, x) \in \rho \Rightarrow x = y$
Докажем, что $\rho \cap \rho^{-1} \subseteq \mathrm{id}_A$.

$$(x,\, y) \in \rho \cap \rho^{-1} \Rightarrow (x,\, y) \in \rho \wedge (x,\, y) \in \rho^{-1} \Rightarrow$$
$$\Rightarrow (x,\, y) \in \rho \wedge (y,\, x) \in \rho \Rightarrow$$
(по определению антисимметричности) $\Rightarrow x = y \Rightarrow (x,\, y) \in \mathrm{id}_A$

Пусть $\rho \cap \rho^{-1} \subseteq \mathrm{id}_A$.
Докажем, что отношение ρ на множестве A антисимметрично.
От противного: пусть отношение ρ на множестве A не антисимметрично,т.е. $\exists (x, y) : ((x, y) \in \rho) \wedge ((y, x) \in \rho) \wedge x \neq y$.

$$((x,\, y) \in \rho \wedge (y,\, x) \in \rho^{-1}) \wedge x \neq y \Rightarrow$$
$$\Rightarrow ((x,\, y) \in (\rho \cap \rho^{-1})) \wedge (\rho \cap \rho^{-1} \subseteq \mathrm{id}_A)(\text{по условию}) \wedge$$
$$\wedge (x,\, y) \notin \mathrm{id}_A(\text{т.к. условию} x \neq y)$$

Получаем противоречие: $(x,\, y) \in \mathrm{id}_A$ и $(x,\, y) \notin \mathrm{id}_A$ одновременно.
►

ТЕОРЕМА 3 . Бинарное отношение ρ на множестве A транзитивно тогда и только тогда, когда его *квадрат* содержится в нем, т.е. $\rho \circ \rho \subseteq \rho$ ($\rho \circ \rho = \rho^2$).

◀ Пусть бинарное отношение ρ на множестве A транзитивно, т.е. $\forall x, y, z \in A : (x,\, y) \in \rho \wedge (y,\, z) \in \rho \Rightarrow (x,\, z) \in \rho$.
Докажем, что: $\rho^2 \subseteq \rho$.

$$(x,\, z) \in \rho^2 \Rightarrow \exists y : (x,\, y) \in \rho \wedge (y,\, z) \in \rho \Rightarrow$$
$$\Rightarrow (\text{по определению транзитивности})(x,\, z) \in \rho \Rightarrow$$
$$\Rightarrow \rho^2 \subseteq \rho$$

Пусть $\rho^2 \subseteq \rho$.
Докажем, что: бинарное отношение ρ транзитивно ($(x, y) \in \rho \wedge (y, z) \in \rho \Rightarrow (x, z) \in \rho$)

$(x, y) \in \rho \wedge (y, z) \in \rho$

(по определению композиции бинарных отношений)$(x, z) \in \rho^2$

$\Rightarrow$ (т.к.$\rho^2 \subseteq \rho$) $(x, z) \in \rho$

►

Бинарное отношение ρ на множестве A называется **плотным**, если для любых $x, y \in A$, отличных друг от друга и таких, что $(x,\ y) \in \rho$, найдется z, отличный и от x и от y, такой, что $(x,\ z) \in \rho$ и $(y,\ z) \in \rho$.
Пусть ρ — плотное бинарное отношение на множестве A.
Тогда

$$\forall x, y \in A : (x,\ y) \in \rho$$
$$\exists z \in A : x \neq z \wedge y \neq z : (x,\ z) \in \rho \wedge (z,\ y) \in \rho \Rightarrow$$
$$\Rightarrow \text{по определению композиции} (x,\ y) \in \rho^2$$

Если ρ плотно, то оно содержится в своем квадрате.

Для транзитивного бинарного отношения $\rho^2 \subseteq \rho$.

Если бинарное отношение ρ плотно и транзитивно одновременно, то $\rho = \rho^2$.

ДИСКРЕТНАЯ МАТЕМАТИКА

ИУ5, 3 семестр, 2015 г.

Лекция 3. ОТНОШЕНИЯ ЭКВИВАЛЕНТНОСТИ И ПОРЯДКА

3.1. Отношения эквивалентности

Пусть A — произвольное множество.
Семейство $(B_i)_{i\in I}$ **непустых** и попарно **не пересекающихся** множеств называют **разбиением** множества A, если
объединение множеств семейства $(B_i)_{i\in I}$ равно A, т.е. $\bigcup\limits_{i\in I} B_i = A$.

Сами множества B_i называют **элементами разбиения** $(B_i)_{i\in I}$.

Пример. Рассмотрим множество точек плоскости. Семейство параллельных прямых образует разбиение плоскости.
Элементом разбиения является множество точек каждой прямой.

Пусть ρ — *эквивалентность* на множестве A и $x \in A$.

Классом эквивалентности по отношению ρ называют множество всех элементов A, *эквивалентных* x, т.е. множество $\{y\colon y \rho x\}$.

Класс эквивалентности обозначают $[x]_\rho$.

Для любого элемента $x \in A$ класс эквивалентности не пуст в силу рефлексивности , так как $x \in [x]_\rho$.

Фактор-множеством множества A по отношению ρ называют множество всех классов эквивалентности по данному отношению эквивалентности ρ на множестве A и обозначают A/ρ.

Утверждение 3.1. Любые два класса эквивалентности по отношению ρ либо не пересекаются, либо совпадают.

◀ Пусть два класса эквивалентности $[x]_\rho$ и $[y]_\rho$ имеют общий элемент $z \in [x]_\rho \cap [y]_\rho$.
Тогда $z \rho x$ и $z \rho y$.
В силу *симметричности* отношения ρ : $x \rho z$, тогда $x \rho z$ и $z \rho y$.
В силу *транзитивности* отношения ρ получим $x \rho y$.
Пусть

$$h \in [x]_\rho \Rightarrow (h \rho x \wedge x \rho y) \Rightarrow h \rho y \Rightarrow h \in [y]_\rho.$$

Это верно для любого элемента $h \in [x]_\rho$.

Обратно,
если

$$h \in [y]_\rho \Rightarrow (h\,\rho\,y) \wedge (x\,\rho\,y) \Rightarrow$$
$$\Rightarrow (\text{ то в силу симметричности}\rho)\ (h\,\rho\,y) \wedge (y\,\rho\,x) \Rightarrow$$
$$\Rightarrow (\text{в силу транзитивности})\ h\,\rho\,x \Rightarrow h \in [x]_\rho \Rightarrow [x]_\rho = [y]_\rho.$$

►

Теорема 1. Для любого *отношения эквивалентности* на множестве A множество классов эквивалентности образует разбиение множества A. Обратно, любое разбиение множества A задает на нем отношение эквивалентности, для которого классы эквивалентности совпадают с элементами разбиения.

◄ Отношение эквивалентности ρ на множестве A определяет некоторое разбиение этого множества.
Каждый элемент множества A принадлежит некоторому классу эквивалентности по отношению ρ т.к. для любого $x \in A$ справедливо $x \in [x]_\rho$ ($x \,\rho\, x$).
Множество всех классов эквивалентности по отношению ρ образует разбиение исходного множества A.

Т.о., любое отношение эквивалентности однозначно определяет некоторое разбиение.

Пусть $(B_i)_{i\in I}$ — некоторое разбиение множества A.
Рассмотрим отношение ρ, такое, что

$$x\,\rho\,y \Leftrightarrow (\exists i \in I)(x \in B_i) \wedge (y \in B_i).$$

Введенное отношение ρ рефлексивно и симметрично.
Если для любых x, y и z имеет место $x\,\rho\,y$ и $y\,\rho\,z$,то x, y и z в силу определения отношения ρ принадлежат одному и тому же элементу B_i разбиения.
Следовательно, $x\,\rho\,z$ и отношение ρ транзитивно.
Таким образом, ρ — эквивалентность на A. ▶

Любая эквивалентность определяет единственное разбиение и наоборот.

Пример 3.1. На множестве **целых чисел** $\mathbb{Z}$ определим отношение $\equiv_{(\mathrm{mod\ k})}$ **отношение равенства по модулю** k , где $k \in \mathbb{N}$:
$x \equiv_{(\mathrm{mod\ k})} y$, если и только если $x - y$ делится на k .

$\equiv_{(\mathrm{mod\ k})}$ — это отношение эквивалентности.

Равенство чисел m и n по модулю k означает, что при делении на k эти числа дают одинаковые остатки.

Различных остатков может быть ровно $k : 0\,,\ 1\,, \ldots,\ k-1\,.$

Получаем ровно k попарно различных классов эквивалентности:
$[0]_{\equiv_{(\mathrm{mod\ k})}}$, $[1]_{\equiv_{(\mathrm{mod\ k})}}$, . . . , $[k-1]_{\equiv_{(\mathrm{mod\ k})}}$,
где класс $[r]_{\equiv_{(\mathrm{mod\ k})}}$ состоит из всех целых чисел, дающих при делении на k остаток r .

3.2. Упорядоченные множества.

Множество вместе с заданным на нем *отношением порядка* называют **упорядоченным множеством**.

Отношение порядка будем обозначать $\leq$ (или значками $\preccurlyeq$, $\sqsubseteq$ и т.п., похожими на $\leq$).
Множество M с заданным на нем отношением порядка $\leq$ будем записывать как пару $(M, \leq)$.

Каждому отношению порядка $\leq$ на множестве M можно сопоставить следующие отношения.

1. Отношение $<$, получается из исходного отношения порядка $\leq$ выбрасыванием всех элементов диагонали id_M.

$$(x < y)\ \forall\, x, y \in M \Leftrightarrow ((x \leq y) \wedge (x \neq y))$$

"Элемент x строго меньше элемента y."

Бинарное отношение $<$ на множестве M —*отношение строгого порядка*. Оно *иррефлексивное*, *антисимметричное* и *транзитивное*.

2. **Двойственный порядок**. Это бинарное отношение на множестве M, обратное к отношению $\leq$.

Обозначение $\geq$.

Тогда для любых x, y условие $x \geq y$ равносильно тому, что $y \leq x$.
Отношение $\geq$ тоже является отношением порядка.
Отношение строгого порядка,ассоциированное с $\geq$, обозначим $>$.

3. Отношение доминирования $x \lhd y$.

Для двух элементов x и y, по определению, $x \lhd y$ тогда и только тогда, когда x строго меньше y и не существует такого элемента z, что $x < z < y$.

Отношение $x \lhd y$ называют **отношением доминирования** (или просто **доминированием**), ассоциированным с отношением порядка $\leq$.

"Элемент y доминирует над элементом x".

Отношение доминирования иррефлексивно, антисимметрично, но не транзитивно.

Пример 3.2. На множестве натуральных чисел $\mathbb{N}$ задано отношение делимости

По отношению делимости 15 доминирует над 3 и 5, но 20 не доминирует над 5, так как существует „промежуточный“ элемент — 10, делитель 20, который делится на 5, но не равен ни 20, ни 5.

3.3. Упорядоченные множества

Рассмотрим упорядоченное множество $(M, \leq)$.

Элементы x и y упорядоченного множества $(M, \leq)$ называют **сравнимыми** по отношению порядка $\leq$, если $x \leq y$ или $y \leq x$.

В противном случае элементы x и y называются **несравнимыми**.

Упорядоченное множество, все элементы которого попарно сравнимы, называют **линейно упорядоченным**, а соответствующее отношение — **отношением линейного порядка** (или просто **линейным порядком**).

Линейно упорядоченное подмножество называют **цепью**.

Любое подмножество попарно не сравнимых элементов данного упорядоченного множества называют **антицепью**.

Пример 3.3. **а.** Отношение естественного числового порядка на множестве $\mathbb{R}$ действительных чисел является отношением линейного порядка, поскольку для любых двух чисел a, b имеет место или неравенство $a \leq b$, или неравенство $b \leq a$.

б. Отношение делимости на множестве $\mathbb{N}$ не является линейным порядком. #

Пусть $(A, \leq)$ — упорядоченное множество.

Элемент $a \in A$ называют **наибольшим элементом** множества A, если для всех $x \in A$ выполняется неравенство $x \leq a$.

Элемент b называют **максимальным элементом** множества A, если для всякого $x \in A$ имеет место одно из двух: или $x \leq b$, или x и b не сравнимы.

Наименьший элемент упорядоченного множества A — это такой его элемент a, что $a \leq x$ для каждого $x \in A$.

Минимальный элемент — это такой элемент $b \in A$, что для любого $x \in A$ элементы b и x не сравнимы или $b \leq x$.

Утверждение 3.2. Наибольший (наименьший) элемент множества, если он существует, является единственным.

◄ Пусть a и a' — наибольшие элементы A по отношению порядка $\leq$.
Для всякого $x \in A$ выполняется $x \leq a$ и $x \leq a'$.
В частности, $a' \leq a$ и $a \leq a'$. Следовательно, $a = a'$ (антисимметричность отношения порядка). ►

Единственность наименьшего элемента доказывается аналогично.
Максимальных (минимальных) элементов может быть сколько угодно.

Пример 3.4.

Отношение порядка на множестве точек плоскости с фиксированной системой координат:

$(a,\ b) \leq (c, d)$, если и только если $a \leq c$ и $b \leq d$.

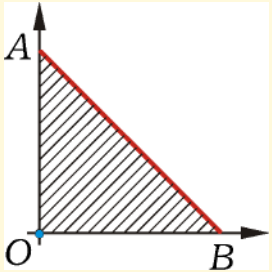

Рис. 1

Рассмотрим множество точек треугольника **OAB** . Точка с координатами $(0,\ 0)$ является наименьшим элементом этого множества.

Максимальными элементами являются все точки, лежащие на стороне AB . Наибольшего элемента нет. #

Пусть $(A, \leq)$ — упорядоченное множество и $B \subseteq A$.
Элемент $a \in A$ называется **верхней** (соответственно **нижней**) **гранью множества** B, если для всех элементов $x \in B$ имеет место $(x \leq a)$ (соответственно $(x \geq a)$).

Точной верхней гранью B называют наименьший элемент множества всех верхних граней множества B и обозначают $\sup B$
Точной нижней гранью B называют наибольший элемент множества всех нижних граней и обозначают($\inf B$).

Элементы $\sup B$ и $\inf B$ могут не принадлежать множеству B.
(наибольший и наименьший элементы множества B всегда принадлежат множеству B)
Точная верхняя (нижняя) грань множества существует не всегда.

Пример 3.5.

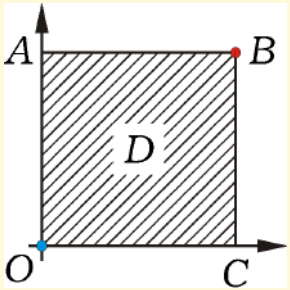

Рис. 2

Рассмотрим множество D точек прямоугольника **OABC** с отношением порядка $(a,\ b) \leq (c, d)$, если и только если $a \leq c$ и $b \leq d$.
Точка O является точной нижней гранью, а точка B — точной верхней гранью этого множества. Обе точки принадлежат множеству.

Пример 3.6.

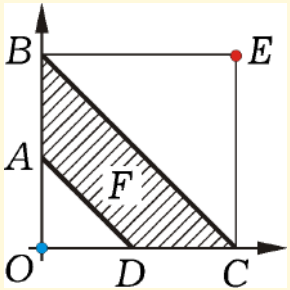

Рис. 3

Рассмотрим множество **F** с тем же отношением порядка.
Точная нижняя грань (точка O) и точная верхняя грань (точка E) множества F существуют, но не принадлежат множеству.

Индуктивное упорядоченное множество

Последовательность $\{x_i\}_{i\in\mathbb{N}}$ элементов упорядоченного **множества** называют **неубывающей**, если для каждого $i \in \mathbb{N}$ справедливо неравенство $x_i \leq x_{i+1}$.

Элемент a упорядоченного множества $(M, \leq)$ называют **точной верхней гранью последовательности** $\{x_i\}_{i\in\mathbb{N}}$, если он есть точная верхняя грань множества всех членов последовательности.

Определение 3.1. Упорядоченное множество $(M, \leq)$ называют **индуктивным**, если:

1) оно содержит наименьший элемент;

2) всякая неубывающая последовательность элементов этого множества имеет точную верхнюю грань.

Пример 3.7. Множество всех подмножеств некоторого множества по отношению включения будет индуктивным.

Наименьший элемент — $\varnothing$.

$$\sup\{A_i\}_{i\in\mathbb{N}} = \cup_{i\in\mathbb{N}} A_i.$$

Определение 3.2. Пусть $(M_1, \leq)$ и $(M_2, \preccurlyeq)$ — индуктивные упорядоченные множества.

Отображение $f\colon M_1 \to M_2$ одного индуктивного упорядоченного множества в другое называют **непрерывным**, если для любой неубывающей последовательности $a_1, \ldots, a_n, \ldots$ элементов множества M_1 образ ее точной верхней грани равен точной верхней грани последовательности образов $f(a_1), \ldots, f(a_n), \ldots$, т.е. справедливо равенство $f(\sup a_n) = \sup f(a_n)$.

Определение 3.3. Отображение $f\colon M_1 \to M_2$ упорядоченных множеств $(M_1, \leq)$ и $(M_2, \preccurlyeq)$ называют **монотонным**, если для любых $a, b \in M_1$ из $a \leq b$ следует $f(a) \preccurlyeq f(b)$.

Функция $f\colon \mathbb{R} \to \mathbb{R}$ будет монотонной в смысле определения 3.3 тогда и только тогда, когда она является неубывающей .

Теорема 2. Всякое непрерывное отображение одного индуктивного упорядоченного множества в другое монотонно.

◄ Пусть f — непрерывное отображение индуктивного упорядоченного множества $(M_1,\, \leq)$ в индуктивное упорядоченное множество $(M_2,\, \preccurlyeq)$.
Пусть $a, b \in M_1$ и $a \leq b$.
Образуем последовательность $\{x_n\}_{n\in\mathbb{N}}$, где $x_1 = a$, а $x_n = b$, $n \geq 2$.
Эта последовательность неубывающая.
Для нее $\sup x_n = \sup\{a,\, b\} = b$.
В силу непрерывности отображения f

$$f(b) = f(\sup x_n) = f(\sup\{a,\, b\}) = \sup\{f(a),\, f(b)\},$$

откуда следует, что $f(a) \preccurlyeq f(b)$. ►

3.4. Теорема о неподвижной точке.

Наиболее важны непрерывные отображения индуктивного упорядоченного множества в себя.

Определение 3.4. Элемент a множества A называют *неподвижной точкой отображения* $f\colon A \to A$, если $f(a) = a$.

Элемент a упорядоченного множества M называют **наименьшей неподвижной точкой отображения** $f\colon M \to M$, если он является наименьшим элементом множества всех неподвижных точек отображения f.

Утверждение 3.3. Если у неубывающей последовательности $\{x_n\}_{n\geq 0}$ отбросить любое конечное число начальных членов, то ее точная верхняя грань не изменится.

Теорема 3 (теорема о неподвижной точке). Любое непрерывное отображение f индуктивного упорядоченного множества $(M, \leq)$ в себя имеет наименьшую неподвижную точку.

◀ $\mathbb{O}$ — наименьший элемент множества M .
Полагаем $f^0(x) = x$.
$f^n(x) = f(f^{n-1}(x))$ для любого $n > 0$,

Рассмотрим последовательность элементов M

$$\{f^n(\mathbb{O})\}_{n\geq 0} = \{\mathbb{O},\, f(\mathbb{O}),\, \ldots,\, f^n(\mathbb{O}),\, \ldots\}. \tag{3.1}$$

Докажем, что последовательность (3.1) неубывающая.
Используем метод математической индукции.
Для наименьшего элемента множества M $\mathbb{O}$ имеем

$$\mathbb{O} = f^0(\mathbb{O}) \leq f(\mathbb{O}).$$

Пусть для некоторого натурального n верно соотношение

$$f^{n-1}(\mathbb{O}) \leq f^n(\mathbb{O}).$$

Отображение f монотонно (теорема 2), поэтому

$$f^n(\mathbb{O}) = f(f^{n-1}(\mathbb{O})) \leq f(f^n(\mathbb{O})) = f^{n+1}(\mathbb{O}),$$

т.е. соотношение верно и для номера $n+1$.

Согласно методу математической индукции,

$$f^n(\mathbb{O}) \leq f^{n+1}(\mathbb{O}) \text{ для любого } n \in \mathbb{N},$$

т.е. последовательность (3.1) неубывающая.

Следовательно, по определению индуктивного упорядоченного множества, она имеет точную верхнюю грань a.

$$a = \sup_{n\geq 0} f^n(\mathbb{O}). \tag{3.2}$$

В силу непрерывности f получаем

$$f(a) = f\left(\sup_{n\geq 0} f^n(\mathbb{O})\right) = \sup_{n\geq 0} f(f^n(\mathbb{O})) = \sup_{n\geq 0} f^{n+1}(\mathbb{O}).$$

Но

$$\sup_{n\geq 0} f^{n+1}(\mathbb{O}) = \sup\{f^1(\mathbb{O}),\ f^2(\mathbb{O}),\ \ldots\} = \sup_{n\geq 1} f^n(\mathbb{O}) = a.$$

a является неподвижной точкой отображения f.

Докажем, что найденная неподвижная точка является наименьшей.

Пусть для некоторого $y \in M$ $f(y) = y$, т.е. y — другая неподвижная точка.
$\mathbb{O} \leq y$ (поскольку $\mathbb{O}$ — *наименьший элемент множества* M.)

Отображение f непрерывно и монотонно.
Следовательно, $f(\mathbb{O}) \leq f(y) = y$, $f(f(\mathbb{O})) \leq f(f(y)) = y$ и т.д.

То есть, для любого $n \geq 0$ $f^n(\mathbb{O}) \leq y$.

Элемент y есть верхняя грань последовательности $\{f^n(\mathbb{O})\}_{n\geq 0}$.

Элемент a — точная верхняя грань (наименьший элемент на множестве всех верхних граней этой последовательности), то $y \geq a$.

Показано, что произвольная неподвижная точка отображения f не меньше элемента a — неподвижной точки отображения f. Следовательно, a — наименьшая неподвижная точка отображения f.

►

Поиск неподвижной точки отображения $f\colon M \to M$ можно рассматривать как задачу поиска наименьшего решения уравнения

$$x = f(x).$$

Доказательство теоремы о неподвижной точке конструктивное: оно дает метод получения неподвижной точки.

Для ее нахождения надо построить последовательность

$$\{\mathbb{O},\ f(\mathbb{O}),\ \ldots,\ f^n(\mathbb{O}),\ \ldots\}$$

и найти ее точную верхнюю грань.

Пример 3.8. Вычисление наименьшей неподвижной точки.

Отрезок $[0,\,1]$ с естественным числовым порядком $\leq$ - это индуктивное упорядоченное множество.

Зададим на этом множестве отображение $f(x) = \frac{1}{2}\,x + \frac{1}{4}$ и рассмотрим уравнение $x = \frac{1}{2}\,x + \frac{1}{4}$.

Для индуктивного упорядоченного множества $([0,\,1], \leq)$ монотонная функция $f\colon [0,\,1] \to [0,\,1]$ — непрерывна.

Для любой неубывающей последовательности $\{x_n\}$ на множестве $[0,\,1]$ справедливо равенство $\sup\{x_n\} = \lim\limits_{n\to\infty} x_n$.

В силу непрерывности функции f *в смысле определения математического анализа* имеем

$$f(\lim_{n\to\infty} x_n) = \lim_{n\to\infty} f(x_n)$$

Так как функция f монотонна, то $\{f(x_n)\}_{n\geq 0}$ — неубывающая последовательность .

$$\lim_{n\to\infty} f(x_n) = \sup f(x_n)$$

В итоге получаем

$$f(\sup x_n) = f(\lim_{n\to\infty} x_n) = \lim_{n\to\infty} f(x_n) = \sup f(x_n)$$

Следовательно, правая часть данного уравнения непрерывна.

Наименьшим элементом рассматриваемого множества является число 0.

Вычисляем:

$$
\begin{aligned}
&f^0(0) = 0,\\
&f^1(0) = 1/4,\\
&f^2(0) = (1/2)\cdot(1/4) + 1/4 = 3/8,\\
&f^3(0) = (1/2)\cdot(3/8) + 1/4 = 7/16,\\
&\ldots\ldots\ldots\ldots\ldots\ldots
\end{aligned}
$$

получая последовательность приближений к наименьшей неподвижной точке.

Можно проверить с помощью метода математической индукции, что

$$f^n(0) = \frac{2^n - 1}{2^{n+1}}, n \in \mathbb{N}.$$

Предел этой последовательности равен $1/2$.

Таким образом, наименьшая неподвижная точка отображения f, определяемого правой частью уравнения, равна $1/2$.

Это единственная в данном случае неподвижная точка отображения f.

Материал для самостоятельного изучения

Отношение эквивалентности

Пример 3.9. На множестве $\mathbb{R}$ действительных чисел зададим отношение $a \equiv_{(\mathrm{mod}\,1)} b$, полагая, что числа a и b равны по модулю 1 тогда и только тогда, когда число $a - b$ является целым.

Из определения следует, что каждое число по модулю 1 равно своей дробной части

Так как отношение $\equiv_{(\mathrm{mod}\,1)}$ определено через равенство, все свойства отношения эквивалентности для него выполняются.

Каждый класс эквивалентности будет содержать числа с равными дробными частями.

Каждый класс эквивалентности по данному отношению однозначно определяет некоторое число из полуинтервала $[0,\,1)$.
Наоборот, каждому числу $\gamma \in [0,\,1)$ однозначно сопоставляется класс эквивалентности, состоящий из всех действительных чисел, дробная часть которых равна γ.
Таким образом, фактор-множество $\mathbb{R}/\equiv_{(\mathrm{mod}\,1)}$ и полуинтервал $[0,\,1)$ на числовой прямой находятся во взаимно однозначном соответствии.

Связь между понятиями эквивалентности и отображения.

Для любого отношения эквивалентности ρ на множестве A можно определить отображение $f_\rho\colon A \to A/\rho$, сопоставив каждому $x \in A$ содержащий его класс эквивалентности.
$f_\rho(x) = [x]_\rho$
Это *отображение сюръективно*, так как каждый элемент множества A принадлежит некоторому классу эквивалентности, т.е. для каждого $[x]_\rho \in A/\rho$ справедливо $[x]_\rho = f_\rho(x)$.

Отображение f_ρ, определенное таким образом, называют **канонической сюръекцией** множества A.

Любое отображение однозначно определяет некоторое отношение эквивалентности.

Теорема 4. Пусть $f\colon A \to B$ — произвольное отображение.На множестве A определим отношение ρ_f : $(x, y) \in \rho_f$, если и только если $f(x) = f(y)$. Это отношение ρ_f является отношением эквивалентности, причем существует биекция фактор-множества A/ρ_f на множество $f(A)$.

◀ Рефлексивность : $f(x) = f(x)$;
Симметричность : $f(x) = f(y)$ и $f(y) = f(x)$;
Транзитивность : $f(x) = f(y) \wedge f(y) = f(z) \Rightarrow f(x) = f(z)$;
т.е. ρ_f — эквивалентность.

$\varphi\colon A/\rho_f \to f(A)\ \varphi([x]_{\rho_f}) = f(x)$.

Каждому классу эквивалентности поставлен в соответствие единственный элемент $y \in f(A)$ (отображение определено корректно).
φ — биекция (инъекция и сюръекция одновременно).
Пусть классы эквивалентности $[x]_{\rho_f}$ и $[y]_{\rho_f}$ не совпадают.
В силу теоремы 1 они не пересекаются, т.е. x не эквивалентно y .
Из определения отношения ρ_f следует, что $f(x) \neq f(y)$.
Таким образом, φ — инъекция.
Если элемент $u \in f(A)$, то найдется такой элемент $x \in A$, что $u = f(x) = \varphi([x]_{\rho_f})$, т.е. φ — сюръекция .
Итак, φ — биекция. ►

Следовательно, в силу доказанных теорем 1 и 4 существует связь между тремя понятиями — отображением множества, отношением эквивалентности на множестве и разбиением множества.

Но **неверно**, что существует взаимно однозначное соответствие между отображениями и отношениями эквивалентности.
Два разных отображения могут определять одно и то же разбиение отображаемого множества, тем самым задавая на нем одно и то же отношение эквивалентности.

Пример 3.10.
a. Любое биективное отображение $f\colon A \to B$ задает на A одно и то же разбиение — тривиальное разбиение на одноэлементные множества.
b. Тождественное отображение множества целых чисел и отображение, сопоставляющее каждому целому n число $n+1$, задают одинаковые разбиения множества целых чисел.

Отношение порядка

Пример 3.11. Рассмотрим множество действительных чисел $\mathbb{R}$ с естественным числовым порядком.

Пусть $a < c$.
Для любых a и c найдется такое b, что $a < b < c$.
Отношение порядка на множестве действительных чисел является плотным.
Поэтому отношение доминирования будет пустым.

Пустым будет и отношение доминирования, ассоциированное с естественным числовым порядком на множестве рациональных чисел.

На множестве целых чисел с естественным числовым порядком отношение доминирования не пусто.
$1 \lhd 2\,,\ -5 \lhd -4\,;$
между 1 и 2 не существует „промежуточный“ элемент.
Записывать $1 \lhd 3$ **неверно**, что, поскольку между единицей и тройкой существует „промежуточный“ элемент — двойка.

ДИСКРЕТНАЯ МАТЕМАТИКА

Ткачев С.Б., каф. ФН-12 МГТУ им. Н.Э. Баумана

ИУ5, 4 семестр, 2015 г.

Лекция 4. ТЕОРЕМА ”О НЕПОДВИЖНОЙ ТОЧКЕ”

4.1. Индуктивное упорядоченное множество

Определение 4.1. Упорядоченное множество $(M, \leq)$ называют **индуктивным**, если:

1) оно содержит наименьший элемент;

2) всякая неубывающая последовательность элементов этого множества имеет точную верхнюю грань.

Пример 4.1. Множество всех подмножеств некоторого множества по отношению включения будет индуктивным.
Наименьший элемент — $\varnothing$.

$$\sup\{A_i\}_{i\in\mathbb{N}} = \cup_{i\in\mathbb{N}} A_i.$$

Определение 4.2. Пусть $(M_1, \leq)$ и $(M_2, \preccurlyeq)$ — индуктивные упорядоченные множества.

Отображение $f\colon M_1 \to M_2$ одного индуктивного упорядоченного множества в другое называют **непрерывным**, если для любой неубывающей последовательности $a_1, \ldots, a_n, \ldots$ элементов множества M_1 образ ее точной верхней грани равен точной верхней грани последовательности образов $f(a_1), \ldots, f(a_n), \ldots$, т.е. справедливо равенство

$$f(\sup a_n) = \sup f(a_n).$$

Определение 4.3. Отображение $f\colon M_1 \to M_2$ упорядоченных множеств $(M_1, \leq)$ и $(M_2, \preccurlyeq)$ называют **монотонным**, если для любых $a, b \in M_1$ из $a \leq b$ следует $f(a) \preccurlyeq f(b)$.

Функция $f\colon \mathbb{R} \to \mathbb{R}$ будет монотонной в смысле определения 4.3 тогда и только тогда, когда она является неубывающей .

Теорема 1. Всякое непрерывное отображение одного индуктивного упорядоченного множества в другое монотонно.

◄ Пусть f — непрерывное отображение индуктивного упорядоченного множества $(M_1, \leq)$ в индуктивное упорядоченное множество $(M_2, \preccurlyeq)$.

Пусть $a, b \in M_1$ и $a \leq b$.
Образуем последовательность $\{x_n\}_{n\in\mathbb{N}}$, где $x_1 = a$, а $x_n = b$, $n \geq 2$.
Эта последовательность неубывающая.
Для нее $\sup x_n = \sup\{a, b\} = b$.
В силу непрерывности отображения f

$$f(b) = f(\sup x_n) = f(\sup\{a, b\}) = \sup\{f(a), f(b)\},$$

откуда следует, что $f(a) \preccurlyeq f(b)$. ►

4.2. Теорема о неподвижной точке.

Наиболее важны непрерывные отображения индуктивного упорядоченного множества в себя.

Определение 4.4. Элемент a множества A называют *неподвижной точкой отображения* $f\colon A \to A$, если $f(a) = a$.

Элемент a упорядоченного множества M называют **наименьшей неподвижной точкой отображения** $f\colon M \to M$, если он является наименьшим элементом множества всех неподвижных точек отображения f.

Утверждение 4.1. Если у неубывающей последовательности $\{x_n\}_{n\geq 0}$ отбросить любое конечное число начальных членов, то ее точная верхняя грань не изменится.

Теорема 2 (теорема о неподвижной точке). Любое непрерывное отображение f индуктивного упорядоченного множества $(M, \leq)$ в себя имеет наименьшую неподвижную точку.

◀ $\mathbb{O}$ — наименьший элемент множества M .

Полагаем

$$f^0(x) = x,$$

$$f^n(x) = f(f^{n-1}(x)) \quad \text{для любого } n > 0\,.$$

Рассмотрим последовательность элементов M

$$\{f^n(\mathbb{O})\}_{n\geq 0} = \{\mathbb{O},\ f(\mathbb{O}),\ f(f(\mathbb{O})),\ \ldots,\ f^n(\mathbb{O}),\ \ldots\}. \qquad (4.1)$$

Докажем, что последовательность (4.1) неубывающая.
Используем метод математической индукции.
Для наименьшего элемента $\mathbb{O}$ множества M имеем

$$\mathbb{O} = f^0(\mathbb{O}) \leq f(\mathbb{O}).$$

Пусть для некоторого натурального n верно соотношение

$$f^{n-1}(\mathbb{O}) \leq f^n(\mathbb{O}).$$

Отображение f монотонно (теорема 1), поэтому

$$f^n(\mathbb{O}) = f(f^{n-1}(\mathbb{O})) \leq f(f^n(\mathbb{O})) = f^{n+1}(\mathbb{O}),$$

т.е. соотношение верно и для номера $n+1$.

Согласно методу математической индукции,

$$f^n(\mathbb{O}) \leq f^{n+1}(\mathbb{O}) \text{ для любого } n \in \mathbb{N},$$

т.е. последовательность (4.1) неубывающая.

Следовательно, по определению индуктивного упорядоченного множества, она имеет точную верхнюю грань

$$a = \sup_{n \geq 0} f^n(\mathbb{O}). \tag{4.2}$$

В силу непрерывности f получаем

$$f(a) = f\left(\sup_{n \geq 0} f^n(\mathbb{O}) \right) = \sup_{n \geq 0} f(f^n(\mathbb{O})) = \sup_{n \geq 0} f^{n+1}(\mathbb{O}).$$

Но

$$\sup_{n \geq 0} f^{n+1}(\mathbb{O}) = \sup \{ f^1(\mathbb{O}),\ f^2(\mathbb{O}),\ \ldots \} = \sup_{n \geq 1} f^n(\mathbb{O}) = a.$$

Следовательно, a является неподвижной точкой отображения f.

Докажем, что найденная неподвижная точка является наименьшей.

Пусть для некоторого $y \in M$ $f(y) = y$, т.е. y — другая неподвижная точка.

$\mathbb{O} \leq y$ (поскольку $\mathbb{O}$ — *наименьший элемент множества* M.)

Отображение f непрерывно и монотонно.
Следовательно, $f(\mathbb{O}) \leq f(y) = y$, $f(f(\mathbb{O})) \leq f(f(y)) = y$ и т.д.

То есть, для любого $n \geq 0$ $f^n(\mathbb{O}) \leq y$.

Элемент y есть верхняя грань последовательности $\{f^n(\mathbb{O})\}_{n\geq 0}$.

Элемент a — точная верхняя грань (наименьший элемент на множестве всех верхних граней этой последовательности), поэтому $y \geq a$.

Показано, что произвольная неподвижная точка отображения f не меньше элемента a — неподвижной точки отображения f . Следовательно, a — наименьшая неподвижная точка отображения f .

►

Поиск неподвижной точки отображения $f\colon M \to M$ можно рассматривать как задачу поиска наименьшего решения уравнения

$$x = f(x).$$

Доказательство теоремы о неподвижной точке конструктивное: оно дает метод получения неподвижной точки.

Для ее нахождения надо построить последовательность

$$\{\mathbb{O},\ f(\mathbb{O}),\ \ldots,\ f^n(\mathbb{O}),\ \ldots\}$$

и найти ее точную верхнюю грань.

Пример 4.2. Вычисление наименьшей неподвижной точки.

Отрезок $[0,\ 1]$ с естественным числовым порядком $\leq$ - это индуктивное упорядоченное множество.

Зададим на этом множестве отображение $f(x) = \frac{1}{2}\,x + \frac{1}{4}$ и рассмотрим уравнение $x = \frac{1}{2}\,x + \frac{1}{4}$.

Для индуктивного упорядоченного множества $([0,\ 1], \leq)$ монотонная функция $f\colon [0,\ 1] \to [0,\ 1]$ — непрерывна.

Для любой неубывающей последовательности $\{x_n\}$ на множестве $[0,\ 1]$ справедливо равенство $\sup\{x_n\} = \lim\limits_{n\to\infty} x_n$.

В силу непрерывности функции f *в смысле определения математического анализа* имеем

$$f(\lim_{n\to\infty} x_n) = \lim_{n\to\infty} f(x_n)$$

Так как функция f монотонна, то $\{f(x_n)\}_{n\geq 0}$ — неубывающая последовательность .

$$\lim_{n\to\infty} f(x_n) = \sup f(x_n)$$

В итоге получаем

$$f(\sup x_n) = f(\lim_{n\to\infty} x_n) = \lim_{n\to\infty} f(x_n) = \sup f(x_n).$$

Следовательно, правая часть данного уравнения непрерывна.

Наименьшим элементом рассматриваемого множества является число 0.

Вычисляем:

$$
\begin{aligned}
&f^0(0) = 0,\\
&f^1(0) = 1/4,\\
&f^2(0) = (1/2)\cdot(1/4) + 1/4 = 3/8,\\
&f^3(0) = (1/2)\cdot(3/8) + 1/4 = 7/16,\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots
\end{aligned}
$$

получая последовательность приближений к наименьшей неподвижной точке.

Можно проверить с помощью метода математической индукции, что

$$f^n(0) = \frac{2^n - 1}{2^{n+1}}, n \in \mathbb{N}.$$

Предел этой последовательности равен $1/2$.

Таким образом, наименьшая неподвижная точка отображения f, определяемого правой частью уравнения, равна $1/2$.

Это единственная в данном случае неподвижная точка отображения f.

Материал для самостоятельного изучения. Упорядоченные множества

Пример. На числовой прямой с „выколотой“ точкой b для полуинтервала $[a,\ b)$ множество верхних граней есть $(b,\ +\infty)$, но точной верхней грани нет.

Упорядоченное множество $(M,\ \leq)$ называют **вполне упорядоченным**, если его любое непустое подмножество имеет наименьший элемент.
Множество натуральных чисел с отношением естественного числового порядка вполне упорядоченное.
Множество целых чисел не вполне упорядоченное, поскольку оно не имеет наименьшего элемента.
Аналогично множества рациональных и действительных чисел не являются вполне упорядоченными.

Для **упорядоченных множеств** справедлив принцип двойственности. Пусть $(M, \leq)$ — произвольное упорядоченное множество.

Тогда любое утверждение, доказанное для порядка $\leq$, останется справедливым для двойственного порядка $\geq$, если в нем:

1) порядок $\leq$ заменить на порядок $\geq$ и наоборот;
2) наименьший (минимальный) элемент заменить наибольшим (максимальным) элементом и наоборот;
3) $\inf$ заменить на $\sup$ и наоборот.

Например, если для некоторого $a \in M$ и для $B \subseteq M$ мы доказали, что $a = \sup B$ при заданном отношении порядка, то для двойственного порядка $a = \inf B$.

Взаимно двойственные определения: если в любом определении, связанном с упорядоченным множеством, произвести взаимные замены согласно принципу двойственности, то получится новое определение, называемое двойственным к исходному.

Определение наибольшего (максимального) элемента множества двойственно к определению наименьшего (минимального) элемента, и наоборот.

Наглядное представление упорядоченных множеств.

Упорядоченное множество можно графически изобразить в виде **диаграммы Хассе**.

Диаграммы Хассе для упорядоченных множеств делителей чисел 2, 6, 30 по отношению делимости.

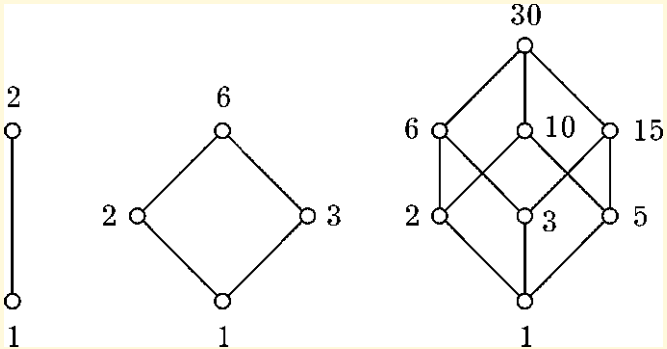

Рис. 1

Утверждение 4.1
Если у неубывающей последовательности $\{x_n\}_{n\geq 0}$ отбросить любое конечное число начальных членов, то ее точная верхняя грань не изменится.

◀ $a = \sup\limits_{n\geq 0} x_n \Rightarrow \forall n\ a \geq x_n\ n \geq 0$.

Зафиксируем произвольное $k > 0$.

$\forall n \geq k\ a \geq x_n$

т.е. a будет верхней гранью подпоследовательности

Докажем, что a является **точной** верхней гранью этой подпоследовательности.

Пусть b — другая верхняя грань , $\forall n \geq k \; b \geq x_n$.
Последовательность $\{x_n\}_{n\geq 0}$ *неубывающая*, следовательно, $x_p \leq x_k$ для каждого $p = \overline{0,\, k-1}$.
$x_k \leq b \Rightarrow x_p \leq b$ (транзитивность отношения порядка)
$\Rightarrow \forall n \geq 0 \; b \geq x_n$ т.е. b есть верхняя грань всей последовательности $\{x_n\}_{n\geq 0}$.
Поскольку $a = \sup\limits_{n\geq 0} x_n$, то $a \leq b$ и $a = \sup\limits_{n\geq k>0} x_n$.
Следовательно, a — **точная** верхняя грань подпоследовательности $\{x_n\}_{n\geq k}$. ▶

ДИСКРЕТНАЯ МАТЕМАТИКА

Ткачев С.Б., ФН-12, МГТУ им. Н.Э. Баумана

ИУ5 - 4 семестр, 2015 г.

Лекция 5. БУЛЕВЫ ФУНКЦИИ

5.1. Понятие булевой функции. Булев куб

Конечной функцией называют **отображение** одного конечного множества в другое. Важный класс таких функций образуют булевы функции.

Булева функция (от n переменных) — это произвольное отображение вида

$$f\colon \{0,\,1\}^n \to \{0,\,1\}, \tag{5.1}$$

т.е. булева функция определена на множестве всех n-элементных (при $n \geq 0$) **последовательностей** (или n-компонентных **кортежей**) нулей и единиц и принимает два возможных значения: 0 и 1.

Можно задать булеву функцию (5.1) в виде

$$y = f(x_1, \ldots, x_n),$$

где каждое булево переменное x_i, $i = \overline{1,\, n}$, и функция f принимают два возможных значения: 0 и 1.
Понятию булевой функции можно придать содержательный смысл, рассматривая элементы множества $\{0,\, 1\}$ так: единицу —как „истину“, нуль — как „ложь“.

Булево переменное часто называют логическим переменным.

Булева функция от n переменных есть отображение n-й декартовой степени множества $\{0,\, 1\}$ в множество $\{0,\, 1\}$.

Обозначим через $\mathcal{P}_{2,n}$ множество всех булевых функций от n переменных (для фиксированного n),

а через $\mathcal{P}_2$ множество всех булевых функций (для всех возможных значений n числа переменных):

$$\mathcal{P}_2 = \bigcup_{n \geq 0} \mathcal{P}_{2,n}.$$

Областью определения любой булевой функции от n переменных является множество $\{0,\,1\}^n$, т.е. **булев куб размерности** n.

Элементы булева куба $\{0,\,1\}^n$ называют n**-мерными булевыми векторами** (или **наборами**).

Число всех элементов булева куба $\{0,\,1\}^n$ составляет 2^n.

Обозначение — $\mathbb{B}^n$.

Элементы булева куба будем также называть его вершинами.

Булев порядок

Для произвольных двух наборов $\tilde{\alpha} = (\alpha_1, \ldots, \alpha_n)$ и $\tilde{\beta} = (\beta_1, \ldots, \beta_n)$ из $\mathbb{B}^n$ имеет место $\tilde{\alpha} \leq \tilde{\beta}$ тогда и только тогда, когда $\alpha_i \leq \beta_i$ для каждого $i = \overline{1, n}$, т.е. $\alpha_i \vee \beta_i = \beta_i$.

Другими словами, $\tilde{\alpha} \leq \tilde{\beta}$ тогда и только тогда, когда $\alpha_i = \beta_i$ или $\alpha_i = 0$, а $\beta_i = 1$ для каждого $i = \overline{1, n}$.

Если существует хотя бы одно i, для которого выполняется $\alpha_i = 0$, $\beta_i = 1$, то имеет место строгое неравенство $\tilde{\alpha} < \tilde{\beta}$.

В частности, если существует ровно одно такое i, то набор $\tilde{\beta}$ **доминирует** над набором $\tilde{\alpha}$, так как этом случае нельзя найти такой набор $\tilde{\gamma}$, что $\tilde{\alpha} < \tilde{\gamma} < \tilde{\beta}$.

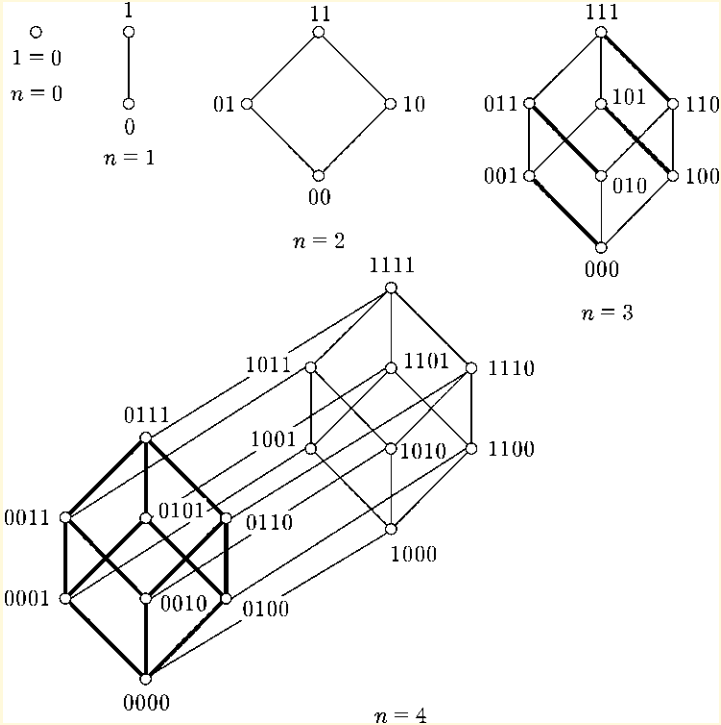

Рис. 1. Диаграммы Хассе для булевых кубов.

Пример 5.1. В булевом кубе $\mathbb{B}^4$

$$(0,\,0,\,1,\,1) < (1,\,0,\,1,\,1) < (1,\,1,\,1,\,1),$$

второй из этих наборов доминирует над первым, а третий — над вторым (но, естественно, третий уже не доминирует над первым, а лишь строго больше его).

Наборы же $(0,\,1,\,0,\,1)$ и $(1,\,0,\,1,\,1)$ — **несравнимые элементы**, так как первая компонента второго набора больше первой компоненты первого набора, но зато вторая компонента первого набора больше второй компоненты второго набора.

Описанное сравнение наборов возможно только для фиксированной размерности и никак нельзя сравнивать наборы разных размерностей. #

5.2. Таблицы булевых функций

Задать булеву функцию $f(x_1, \ldots, x_n)$ от n переменных можно, указав значение функции на каждом из наборов значений переменных.

Поскольку каждое переменное может принимать только два значения — 0 и 1, имеется 2^n попарно различных наборов.

Булева функция от n переменных может быть задана таблицей, состоящей из двух столбцов и 2^n строк.

В первом столбце перечисляют все наборы из $\mathbb{B}^n$, а во втором — значения функции на соответствующих наборах. В таблице применяют стандартное расположение наборов: каждый набор рассматривают как запись натурального числа в двоичном исчислении и располагают наборы в соответствии с естественным числовым порядком.

На каждом наборе булева функция может принимать только два значения — 0 и 1, следовательно число булевых функций от n переменных равно 2^{2^n}.

Форма таблицы произвольной булевой функции:

Таблица 5.1

$x_1 \dots x_n$	$f(x_1, \dots, x_n)$
$0 \dots 0$	$f(0, \dots, 0)$
$\dots$	$\dots$
$(\alpha_{k,1} \dots \alpha_{k,n})$	$f(\alpha_{k,1}, \dots, \alpha_{k,n})$
$\dots$	$\dots$
$1 \dots 1$	$f(1, \dots, 1)$

В $(k+1)$ -й строке таблицы расположен набор

$$\tilde{\alpha}_k = (\alpha_{k,1} \dots \alpha_{k,n}),$$

являющийся двоичным кодом числа k
(при $0 \leq k \leq 2^n - 1$).

Примеры булевых функций, заданных таблицами.
Для функции одного переменного ($n = 1$) можно задать четыре булевых функции

Таблица 5.2

x	$f_1(x)$	$f_2(x)$	$f_3(x)$	$f_4(x)$
0	0	0	1	1
1	1	0	1	0

Функцию f_1 называют **тождественной функцией**.
Функцию f_4 — **отрицанием**.
Функции f_2 и f_3 являются функциями (от одного переменного), принимающими постоянное значение (0 и 1 соответственно). Их также называют константой 0 и константой 1 .

Существуют 16 различных булевых функций от двух переменных. В таб. 5.3 указаны семь (из $2^{2^2} = 16$) наиболее часто употребляемых булевых функций от двух переменных.

Таблица 5.3

x_1	x_2	f_1	f_2	f_3	f_4	f_5	f_6	f_7
0	0	0	0	0	1	1	1	1
0	1	1	0	1	1	0	1	0
1	0	1	0	1	0	0	1	0
1	1	1	1	0	1	1	0	0

Каждая булева функция от двух переменных есть одновременно и бинарная операция на множестве $\{0,\ 1\}$, для таких функций можно использовать запись, принятую для бинарных операций: $x\omega y$ вместо $\omega(x, y)$.

Для функций, записанных в табл. 5.3, принимаются следующие обозначения:

$$f_1(x_1, x_2) = x_1 \vee x_2, \qquad f_2(x_1, x_2) = x_1 \cdot x_2$$
$$(\text{или } f_2(x_1, x_2) = x_1 \wedge x_2),$$
$$f_3(x_1, x_2) = x_1 \oplus x_2, \qquad f_4(x_1, x_2) = x_1 \to x_2,$$
$$f_5(x_1, x_2) = x_1 \sim x_2, \qquad f_6(x_1, x_2) = x_1 \mid x_2,$$
$$f_7(x_1, x_2) = x_1 \downarrow x_2.$$

Функцию f_1 называют **дизъюнкцией**,
f_2 — **конъюнкцией**,
f_3 — **сложением по модулю 2** (mod 2),
f_4 — **импликацией**;
f_5 — **эквивалентностью**,
f_6 — **штрихом Шеффера**,
f_7 — **стрелкой Пирса**.

Штрих Шеффера есть отрицание конъюнкции, а стрелка Пирса — отрицание дизъюнкции:

$$x_1 \mid x_2 = \overline{x_1 \cdot x_2}, \quad x_1 \downarrow x_2 = \overline{x_1 \vee x_2}.$$

В число булевых функций от двух переменных также входят две функции, принимающие постоянные значения 0 и 1.

Пример 5.2. Таблица булевой функции от трех переменных:

Таблица 5.4

Номер набора	x_1	x_2	x_3	$f(x_1, x_2, x_3)$
0	0	0	0	0
1	0	0	1	0
2	0	1	0	0
3	0	1	1	1
4	1	0	0	0
5	1	0	1	1
6	1	1	0	1
7	1	1	1	1

Эта функция называется **мажоритарной функцией** или **функцией голосования**.

Табличный способ задания булевых функций применяют ограниченно, поскольку, например, для задания некоторой функции от десяти переменных потребуется таблица из 1024 строк.

Даже все функции от пяти переменных практически невозможно задать таблицами, поскольку число таких функций равно $2^{2^5} = 4\ 294\ 967\ 296$.

Различных функций от трех переменных — 256.

Для задания функции от n переменных можно применять более экономные способы — достаточно записать вектор значений булевой функции на всех наборах в порядке их следования в таблице.

Например, мажоритарная функция f может быть задана в виде

$$f = (0,\ 0,\ 0,\ 1,\ 0,\ 1,\ 1,\ 1)$$

Можно также перечислить номера тех наборов, на которых функция принимает значение 1:

$$f = \{3,\ 5,\ 6,\ 7\}.$$

5.3. Фиктивные переменные. Равенство булевых функций

Две булевы функции от одного и того же числа переменных n равны, если совпадают таблицы этих функций.
Возможно ли равенство при различном количестве переменных?

Пример 5.3. Рассмотрим булевы функции
$f(x, y) = x \vee y$ и $g(x, y, z) = xz \vee x\overline{z} \vee yz \vee y\overline{z}$.
Используя тождества, запишем $g(x, y, z) = (x \vee y)(z \vee \overline{z})$.
Поскольку $z \vee \overline{z} = 1$, то $g(x, y, z) = (x \vee y) = f(x, y)$.
Функции f и g можно считать равными, хотя они формально зависят от разного числа переменных.
Значение переменного z не влияет на значение функции g.

Определение 5.1. Переменное x_i называют **фиктивным переменным** булевой функции $f(x_1, \ldots, x_n)$, если значение функции не зависит от значения этого переменного, т.е. если для любых значений переменных x_1, $\ldots$, x_{i-1}, x_{i+1}, $\ldots$, x_n

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

Переменное x, не являющееся фиктивным переменным функции f, называют **существенным переменным** данной функции.

Говорят, что функция f существенно зависит от переменного x.

Чтобы определить по таблице булевой функции, что переменное x_i является фиктивным, нужно проверить, что на всех парах наборов, у которых одинаковы все компоненты, кроме i-й, а i-я компонетна в одном наборе равна 0, а в другом — 1, функция принимает равные значения.

Пример 5.4. Для функции g переменное x_2 является фиктивным, остальные — существенны.

Таблица 5.5

	x_1	x_2	x_3	x_4	$f(x_1, x_2, x_3, x_4)$
0	0	0	0	0	0
1	0	0	0	1	0
2	0	0	1	0	0
3	0	0	1	1	1
4	0	1	0	0	0
5	0	1	0	1	0
6	0	1	1	0	0
7	0	1	1	1	1
8	1	0	0	0	0
9	1	0	0	1	1
10	1	0	1	0	1
11	1	0	1	1	1
12	1	1	0	0	0
13	1	1	0	1	1
14	1	1	1	0	1
15	1	1	1	1	1

Определение 5.2. Булевы функции f и g называют **равными**, если их существенные переменные соответственно равны и на каждом наборе значений этих переменных функции f и g принимают равные значения.

Возможно добавление к множеству переменных булевой функции одной или нескольких фиктивных переменных. Например

$$\hat{f}(x_1, \ldots, x_n, y) = f(x_1, \ldots, x_n) \cdot (y \vee \overline{y}).$$

Понятие фиктивного переменного позволяет также произвольные две булевы функции рассматривать как функции от одного и того же числа переменных.

5.4. Формулы

Пусть даны некоторое счетное множество P, элементы которого будем называть булевыми переменными, и некоторое множество булевых функций F.

Элементы множества P будем обозначать символами $x_1, x_2, \ldots, x_n, \ldots$ Элементы множества F будем обозначать строчными латинскими буквами.

Константы 0 и 1 также могут включаться в F.

1. **Базис индукции.** Формулой над множеством F считают любую константу из F (если она там есть) и любое булево переменное из P.

2. **Индуктивный переход.** Если $\Phi_1, \ldots, \Phi_n$ ($n \geq 1$) — формулы над множеством F, а f — функция из F от n переменных, то выражение $f(\Phi_1, \ldots, \Phi_n)$ есть формула над множеством F.

3. **Замыкание.** Никаких других формул над множеством F, кроме определенных выше, не существует.

Промежуточные формулы, используемые при построении некоторой формулы, называют **подформулами** этой формулы.

При построении формул с использованием функций двух переменных будем применять традиционную форму записи функции как бинарной операции.

Множество $F = \{\vee, \cdot, \bar{}\}$, состоящее из функций дизъюнкции, конъюнкции и отрицания, называют **стандартным базисом**.

Формулами над стандартным базисом будут любые переменные, например x_1, x_2, x_3.

Из переменных x_1, x_2 как формул и функции $\vee$ построим новую формулу

$$\Phi_1 = (x_1 \vee x_2).$$

Используя $\cdot$, Φ_1 и x_3, получим

$$\Phi_2 = \Phi_1 \cdot x_3 = (x_1 \vee x_2) \cdot x_3.$$

Каждому набору значений переменных, входящих в заданную формулу, сопоставляется значение этой формулы. Это значение вычисляется путем последовательного вычисления значений всех подформул, входящих в формулу. Процесс вычисления значения формулы повторяет процесс построения формулы из подформул.

Каждая формула, построенная по указанным выше правилам, однозначно определяет некоторую булеву функцию.

Если булева функция $f(x_1, \ldots, x_n)$ представляется формулой $\Phi(x_1, \ldots, x_n)$, то пишут $f = \Phi(x_1, \ldots, x_n)$.

Пример 5.5. Для основных функций от двух переменных мы можем записать следующие формулы над стандартным базисом:

$$x_1 \oplus x_2 = x_1 \cdot \overline{x}_2 \vee \overline{x}_1 \cdot x_2,$$
$$x_1 \to x_2 = \overline{x}_1 \vee x_2,$$
$$x_1 \sim x_2 = (\overline{x}_1 \vee x_2) \cdot (\overline{x}_2 \vee x_1),$$
$$x_1 | x_2 = \overline{x_1 \cdot x_2},$$
$$x_1 \downarrow x_2 = \overline{x_1 \vee x_2}.$$

Соответствие между формулами над фиксированным множеством и представляемыми ими функциями **не является взаимно однозначным**. Возможны различные разложения функции по одному и тому же базису.
Например, формулы

$$\overline{(x \vee y)} \quad \text{и} \quad \overline{x} \cdot \overline{y}$$

над стандартным базисом представляют одну и ту же функцию.

Эквивалентными называются формулы, которые представляют равные функции.

Эквивалентным (или **тождественным**) **преобразованием формулы** Φ называют переход (по определенным правилам) к любой формуле Ψ, эквивалентной формуле Φ.

Тождеством (над множеством $F \subseteq \mathcal{P}_2$) называют выражение

$$\Phi(x_1, \ldots, x_n) = \Psi(x_1, \ldots, x_m), \tag{5.2}$$

где формулы Φ и Ψ — эквивалентные формулы над F.

Правила тождественных преобразований.

Утверждение 5.1.

1. Если в тождестве некоторые переменные заменить произвольными формулами (над множеством F), то тождество сохранится, т.е. полученные в результате такой замены новые формулы останутся эквивалентными.

2. Если в формуле Φ произвольную ее подформулу заменить любой эквивалентной ей, то получится формула, эквивалентная формуле Φ. #

5.5. Дизъюнктивные и конъюнктивные нормальные формы

Любая формула вида x или $\overline{x}$ над стандартным базисом, где x — произвольное переменное, называется **литералом**. Литерал есть обозначение либо самого переменного x, либо его отрицания.

Для $\sigma \in \{0, 1\}$ пишут x^σ, понимая под этим само переменное x, если $\sigma = 1$, и отрицание x, если $\sigma = 0$, т.е.

$$x^\sigma = \begin{cases} x, & \sigma = 1; \\ \overline{x}, & \sigma = 0. \end{cases} \tag{5.3}$$

Подставляя в (5.3) 0 и 1 вместо x, получаем

$$0^\sigma = \begin{cases} 0, & \sigma = 1; \\ 1, & \sigma = 0, \end{cases} \qquad 1^\sigma = \begin{cases} 1, & \sigma = 1; \\ 0, & \sigma = 0. \end{cases}$$

Используют также обозначение $\tilde{x}$, понимая под этим любой из двух литералов — x или $\overline{x}$.

Формула вида $\tilde{x}_1\tilde{x}_2\ldots\tilde{x}_m$, где все фигурирующие в ней переменные попарно различны, называется **элементарной конъюнкцией**.

Определение 5.3. Дизъюнктивная нормальная форма (ДНФ) от переменных $x_1, \ldots, x_n$ — это формула вида

$$K_1 \vee \ldots \vee K_m,$$

где K_i, $i = \overline{1, m}$, — элементарная конъюнкция, содержащая некоторые из литералов $x_1, \ldots, x_n$.

ДНФ называется **совершенной дизъюнктивной нормальной формой (СДНФ)**, если в каждую конъюнкцию K_i для каждого номера $j = \overline{1, n}$ входит в точности один из литералов $\tilde{x_j}$.

Формула вида $\tilde{x}_1 \vee \tilde{x}_2 \vee \ldots \vee \tilde{x}_m$, где все фигурирующие в ней переменные попарно различны, называется **элементарной дизъюнкцией**.

Определение 5.4. Конъюнктивная нормальная форма (ДНФ) от переменных $x_1, \ldots, x_n$ — это формула вида

$$D_1 \wedge \ldots \wedge D_m,$$

где D_i, $i = \overline{1, m}$, — элементарная дизъюнкция, содержащая некоторые из литералов $x_1, \ldots, x_n$.

КНФ называется **совершенной конъюктивной нормальной формой (СКНФ)**, если в каждую дизъюнкцию D_i для каждого номера $j = \overline{1, n}$ входит в точности один из литералов $\tilde{x_j}$.

Теорема 1. Любая булева функция, отличная от константы 0, представима в виде СДНФ.

◄ Для функции $f \in \mathcal{P}_{2,n}$, не равной тождественно 0, рассмотрим множество $C_f^1 = \{\tilde{\alpha}\colon f(\tilde{\alpha}) = 1\}$.
Это множество C_f^1 не пусто.
Каждый набор из C_f^1 называется **конституентой единицы** функции f.
Каждому набору $\tilde{\alpha} \in C_f^1$ поставим в соответствие элементарную конъюнкцию $K_{\tilde{\alpha}} = x_1^{\alpha_1} x_2^{\alpha_2} \dots x_n^{\alpha_n}$.
$K_{\tilde{\alpha}}$ обращается в единицу только на наборе $\tilde{\alpha}$.
Тогда искомой СДНФ для функции f будет

$$f = \bigvee_{\tilde{\alpha} \in C_f^1} K_{\tilde{\alpha}}.$$

►

Теорема 2. Любая булева функция, отличная от константы 1 , представляется в виде СКНФ.

СКНФ функции f будет иметь вид

$$f = \bigwedge_{\tilde{\alpha} \in C_f^0} D_{\tilde{\alpha}},$$

где множество наборов $C_f^0 = \{\tilde{\alpha} \colon f(\tilde{\alpha}) = 0\}$, и каждый набор $\tilde{\alpha}$ из C_f^0 называют **конституентой нуля** функции f .

Таким образом любая булева функция может быть представлена в виде формулы над стандартным базисом (СДНФ или СКНФ).

Пример 5.6. Построим СДНФ и СКНФ для **мажоритатрной функции**

$$f = (0, 0, 0, 1, 0, 1, 1, 1).$$

Конституентами единицы для нее служат наборы

$$\tilde{\alpha}_1 = (0,\ 1,\ 1)\,,\ \tilde{\alpha}_2 = (1,\ 0,\ 1)\,,$$
$$\tilde{\alpha}_3 = (1,\ 1,\ 0)\,,\ \tilde{\alpha}_4 = (1,\ 1,\ 1)\,.$$

Им соответствуют элементарные конъюнкции

$$K_{\tilde{\alpha}_1} = \overline{x}_1 x_2 x_3\,,\ K_{\tilde{\alpha}_2} = x_1 \overline{x}_2 x_3\,,$$
$$K_{\tilde{\alpha}_3} = x_1 x_2 \overline{x}_3\,,\ K_{\tilde{\alpha}_4} = x_1 x_2 x_3\,.$$

Тогда СДНФ, представляющая мажоритарную функцию, имеет вид

$$\overline{x}_1 x_2 x_3 \vee x_1 \overline{x}_2 x_3 \vee x_1 x_2 \overline{x}_3 \vee x_1 x_2 x_3. \qquad (5.4)$$

Для получения СКНФ для той же функции выпишем все конституенты нуля данной функции:

$$(0,\,0,\,0)\,,\ (0,\,0,\,1)\,,\ (0,\,1,\,0)\,,\ (1,\,0,\,0)\,.$$

Сопоставим им элементарные дизъюнкции

$$x_1 \vee x_2 \vee x_3\,,\ x_1 \vee x_2 \vee \overline{x}_3\,,\ x_1 \vee \overline{x}_2 \vee x_3\,,\ \overline{x}_1 \vee x_2 \vee x_3$$

соответственно.
СКНФ для мажоритарной функции:

$$(x_1 \vee x_2 \vee x_3)(x_1 \vee x_2 \vee \overline{x}_3)(x_1 \vee \overline{x}_2 \vee x_3)(\overline{x}_1 \vee x_2 \vee x_3). \quad (5.5)$$

5.6. Минимизация булевых функций

Проблема минимизации

Задача: среди всех ДНФ, представляющих данную функцию, выбрать наиболее простую (минимальную) ДНФ в смысле некоторого критерия.

Критерии простоты: число литералов, число элементарных конъюнкций и число отрицаний, используемых для записи формулы.

Определение 5.5. Число входящих в ДНФ элементарных конъюнкций называют ее **длиной**.

Определение 5.6. ДНФ называют **кратчайшей**, если она имеет наименьшую длину среди всех ДНФ, представляющих функцию f.

ДНФ называют **минимальной**, если она содержит наименьшее число литералов среди всех ДНФ, представляющих функцию f.

Пример 5.7. ДНФ $x_1x_2 \vee \overline{x}_1x_2$ не является минимальной, так как ее можно преобразовать к эквивалентной ДНФ:

$$x_1x_2 \vee \overline{x}_1x_2 = (x_1 \vee \overline{x}_1)x_2 = x_2.$$

Полученная ДНФ не содержит ни одного из литералов $\tilde{x}_1$, т. е. x_1 или $\overline{x}_1$. Вместо четырех литералов в исходной ДНФ получаем ДНФ, состоящую из одного литерала. ►

Возможен ли прямой перебор?

Теоретически можно перебрать все ДНФ от n переменных и среди них выбрать нужные.

В каждую элементарную конъюнкцию каждое переменное может входить, не входить или входить с отрицанием, число различных конъюнкций (включая пустую) равно 3^n .

Для проверки соответствия ДНФ функции придется выполнить не менее 2^{3^n} более мелких операций. При $n = 3$ получим 134 217 728 операций.

Практически метод прямого перебора неприменим.

Для функции f, принимающей хотя бы одно ненулевое значение, всегда строится СДНФ.

Идея: максимально сократить количество анализируемых вариантов, используя СДНФ, а затем выбрать среди найденных вариантов наилучший по заданным критериям.

Минимизация с использованием тождеств склейки и поглощения

Тождества простой склейки и тождество поглощения:

$$xK \vee \overline{x}K = K, \quad xK \vee K = K, \tag{5.6}$$

где
K — некоторая элементарная конъюнкция,
xK и $\overline{x}K$ — некоторые элементарные конъюнкции, присутствующие в СДНФ минимизируемой функции.

Тождество склейки позволяет получить элементарную конъюнкцию K, которая содержит на один литерал меньше, чем исходные конъюнкции.

Тождество поглощения позволяет удалить из формулы элементарные конъюнкции xK и $\overline{x}K$, заменив их на K. Применение тождества поглощения уменьшает длину ДНФ на единицу.

К полученной ДНФ можно снова применить эти тождества. Продолжать следует до тех пор, пока применение тождеств не станет невозможным.

Тождество склейки применяют к различным парам конъюнкций, входящих в ДНФ, результат применения тождества склейки дописывают в ДНФ без удаления „склеиваемых“ конъюнкций. Удаляют конъюнкции с использованием тождества поглощения после того, как все возможные склейки выполнены.

Пример 5.8. Рассмотрим функцию трех переменных, заданную в виде СДНФ:

$$f = \overline{x}_1\overline{x}_2\overline{x}_3 \vee \overline{x}_1\overline{x}_2x_3 \vee x_1\overline{x}_2x_3 \vee x_1x_2\overline{x}_3 \vee x_1x_2x_3. \quad (5.7)$$

1. Применим тождество склейки к первой и второй, второй и третьей, а также к четвертой и пятой конъюнкциям. Результаты выполнения тождества склейки добавим к исходной СДНФ, такое добавление дает ДНФ, эквивалентную исходной. Получим

$$\begin{aligned}\overline{x}_1\overline{x}_2\overline{x}_3 \vee \overline{x}_1\overline{x}_2x_3 \vee \overline{x}_1\overline{x}_2 \vee \\ \vee \overline{x}_1\overline{x}_2x_3 \vee x_1\overline{x}_2x_3 \vee \overline{x}_2x_3 \vee \\ \vee x_1x_2\overline{x}_3 \vee x_1x_2x_3 \vee x_1x_2.\end{aligned}$$

После применения тождества поглощения запишем ДНФ

$$\overline{x}_1\overline{x}_2 \vee \overline{x}_2 x_3 \vee x_1 x_2. \qquad (5.8)$$

Дальнейшее упрощение формулы с использованием тождеств склейки и поглощения невозможно.

2. Вариант 1 преобразования не является единственным. Применим тождество склейки к первой и второй, третьей и пятой, четвертой и пятой конъюнкциям.
$(f = \overline{x}_1\overline{x}_2\overline{x}_3 \vee \overline{x}_1\overline{x}_2x_3 \vee x_1\overline{x}_2x_3 \vee x_1x_2\overline{x}_3 \vee x_1x_2x_3)$
Запишем

$$\begin{aligned}\overline{x}_1\overline{x}_2\overline{x}_3 \vee \overline{x}_1\overline{x}_2x_3 \vee \overline{x}_1\overline{x}_2 \vee \\ \vee x_1\overline{x}_2x_3 \vee x_1x_2x_3 \vee x_1x_3 \vee \\ \vee x_1x_2\overline{x}_3 \vee x_1x_2x_3 \vee x_1x_2.\end{aligned}$$

Получим ДНФ

$$\overline{x}_1\overline{x}_2 \vee x_1x_3 \vee x_1x_2, \tag{5.9}$$

отличную от ДНФ (5.8). Дальнейшее упрощение ДНФ (5.9) невозможно.

3. В исходной СДНФ применим тождество склейки ко всем возможным парам (первой и второй, второй и третьей, третьей и пятой, четвертой и пятой конъюнкциям), а затем применим тождество поглощения, получим ДНФ

$$\overline{x}_1\overline{x}_2 \vee \overline{x}_2x_3 \vee x_1x_2 \vee x_1x_3, \tag{5.10}$$

которая содержит все конъюнкции, входящие в ДНФ (5.8) и (5.9). Других ДНФ, эквивалентных СДНФ (5.7), построить не удается.

Итог:

1. $\overline{x}_1\overline{x}_2 \vee \overline{x}_2x_3 \vee x_1x_2$.
2. $\overline{x}_1\overline{x}_2 \vee x_1x_3 \vee x_1x_2$
3. $\overline{x}_1\overline{x}_2 \vee \overline{x}_2x_3 \vee x_1x_2 \vee x_1x_3$

Две первые ДНФ являются кратчайшими и минимальными, ДНФ 2 содержит на одно отрицание меньше.

Все три ДНФ содержат общую часть $D = \overline{x}_1\overline{x}_2 \vee x_1x_2$, ДНФ 1 и 2 образуются добавлением к D конъюнкций $\overline{x}_2x_3$ и x_1x_3 соответственно, добавляемые конъюнкции содержатся в (5.10).

Определение 5.7. Булеву функцию g называют **импликантой** булевой функции f, если для любых наборов значений переменных из $g = 1$ следует $f = 1$.

Простой импликантой булевой функции f называют такую элементарную конъюнкцию в составе некоторой ДНФ, представляющей функцию f, что удаление из нее любого литерала приводит к тому, что она перестает быть импликантой.

Например, конъюнкция $\overline{x}_1\overline{x}_2\overline{x}_3$ (см. пример 5.8) не является простой импликантой функции f, так как из нее можно удалить литерал $\overline{x}_3$ и получить конъюнкцию $\overline{x}_1\overline{x}_2$. Эта конъюнкция будет простой импликантой.

Сокращенная ДНФ функции f есть ДНФ, представляющая функцию f, в которую входят все простые импликанты булевой функции f.

Алгоритм минимизации

1) последовательно выполняя все возможные склейки, а затем применяя к результату тождество поглощения, получить сокращенную ДНФ, содержащую все конъюнкции, дальнейшее упрощение которых с помощью указанных тождеств невозможно;

2) выделить из сокращенной ДНФ общую часть, входящую в любую представляющую функцию f ДНФ, составленную из простых импликант, затем выписать все возможные ДНФ (тупиковые), состоящие из простых импликант и представляющие функцию f;

3) найти среди выписанных ДНФ кратчайшие (по количеству слагаемых);

4) выбрать из кратчайших минимальные (по количеству литералов).

Геометрическая интерпретация первого шага алгоритма.

Установим смысл простой склейки с точки зрения геометрии булева куба.

Каждому набору $\tilde{\alpha} = (\alpha_1, \ldots, \alpha_n)$, для которого $f(\tilde{\alpha}) = 1$ в СДНФ соответствует элементарная конъюнкция $K_{\tilde{\alpha}} = x_1^{\alpha_1} \cdot \ldots \cdot x_n^{\alpha_n}$, принимающая значение 1 только на наборе $\tilde{\alpha}$.

Простая склейка может быть применена лишь к таким двум элементарным конъюнкциям $K_{\tilde{\alpha}}$ и $K_{\tilde{\beta}}$, которые отличаются только одним литералом.

Соответствующие наборы $\tilde{\alpha}$, $\tilde{\beta}$ различаются значением всего одной компоненты, т. е. они образуют ребро булева куба $\mathbb{B}^n$.

Тождество простой склейки можно применить только к тем элементарным конъюнкциям исходной СДНФ, представляющей функцию f, которые соответствуют элементам какого-либо ребра булева куба, на котором функция f принимает единичное значение.

Применяя простую склейку к исходной СДНФ Φ, получаем новую ДНФ Φ_1. Если возможно, к ней также применяем простую склейку — получаем ДНФ Φ_2.

Геометрия повторения простой склейки состоит в дальнейшем склеивании каждой пары ребер, принадлежащих одной грани размерности 2 и не имеющих общей вершины (противолежащих), на которых значение функции равно 1, в грани размерности 2.

Два ребра, принадлежащие одной грани размерности 2 и имеющие общую вершину, не склеиваются.

Продолжаем выполнять эту операцию до тех пор, пока не окажется, что для некоторого k в ДНФ Φ_k уже нельзя склеить никакие две элементарные конъюнкции. В силу конечности булева куба такое k всегда найдется.

Пример 5.9. Функция f от трех переменных, задана СДНФ:

$$\overline{x}_1\overline{x}_2\overline{x}_3 \vee \overline{x}_1\overline{x}_2x_3 \vee x_1\overline{x}_2\overline{x}_3 \vee x_1\overline{x}_2x_3. \qquad (5.11)$$

Применим тождество простой склейки к первой и третьей, ко второй и четвертой элементарным конъюнкциям в (5.11):
$\overline{x}_1\overline{x}_2\overline{x}_3 \vee x_1\overline{x}_2\overline{x}_3 = \overline{x}_2\overline{x}_3, \quad \overline{x}_1\overline{x}_2x_3 \vee x_1\overline{x}_2x_3 = \overline{x}_2x_3.$
В результате получим

$$f = \overline{x}_2\overline{x}_3 \vee \overline{x}_2x_3. \qquad (5.12)$$

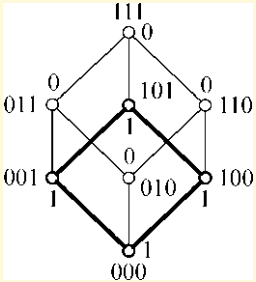

Рис. 2

С геометрической точки зрения склейка первой и третьей конъюнкций в формуле (5.11) означает, что функция f принимает единичное значение на ребре $[000, 100]$, а склейка второй и четвертой конъюнкций — на ребре $[001, 101]$.

Эти ребра являются соседними. Функция f принимает единичное значение и на другой паре соседних ребер: $[000, 001]$ и $[100, 101]$.

Если функция принимает единичное значение на двух соседних противолежащих ребрах булева куба, то она равна 1 в любой точке образуемой ими грани размерности 2. Применяя простую склейку к (5.12) (по переменному x_3), получаем $f(x_1, x_2, x_3) = \overline{x}_2$.

Карты Карно

Для булевых функций от трех и четырех переменных процедура склейки наглядно и просто выполняется на **картах Карно**, представляющих собой прямоугольные таблицы.

В карте Карно для функции от трех переменных строки отмечены наборами значений переменного x_1, а столбцы — x_2, x_3. В клетках таблицы пишут 1 в том случае, если на соответствующем наборе исследуемая функция принимает значение 1. Карту Карно можно рассматривать как специальную таблицу, задающую булеву функцию.

Порядок следования наборов переменных в строках и столбцах определен так, чтобы двум соседним по горизонтали или вертикали клеткам соответствовали наборы, соединенные в булевом кубе ребром.

Пример 5.10. Рассмотрим функцию от трех переменных $f = \{0, 1, 2, 4, 5\}$.

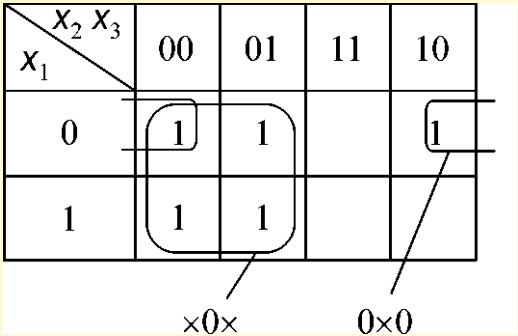

Рис. 3

Карта Карно функции и максимальные склейки показаны на рис. 3.

Отметим, что одна из склеек выполняется „через край“, а другая — сразу на четыре единицы. Сокращенная ДНФ имеет вид $\overline{x}_2 \vee \overline{x}_1\overline{x}_3$. ►

Карта Карно для функции от четырех переменных есть табличное представление булева куба $\mathbb{B}^4$. Строки карты помечены наборами значений переменных x_1, x_2, а столбцы — x_3, x_4. На рис. 4 показана карта Карно для функции $f = \{0, 1, 2, 4, 5, 7, 8, 10\}$. На карте обозначены максимальные склейки, соответствующие простым импликантам функции.

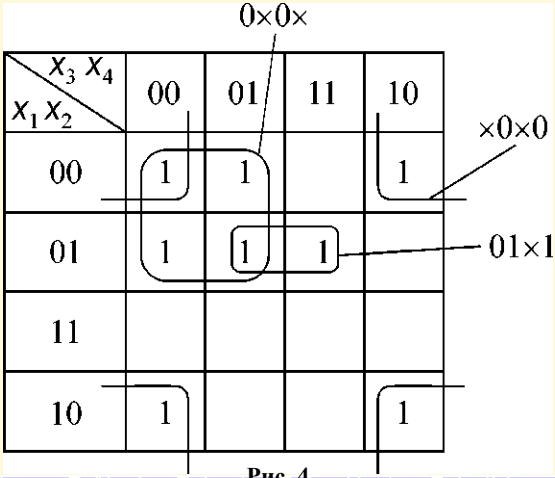

Рис. 4

Для карты, показанной на рис. 4, получим сокращенную ДНФ в виде $\overline{x}_2\overline{x}_4 \vee \overline{x}_1\overline{x}_3 \vee \overline{x}_1 x_2 x_4$.

Для функции от четырех переменных возможно получение склейки на восемь единиц, такая склейка отвечает грани куба размерности 3. В принципе, для функции четырех переменных возможна склейка шестнадцати единиц, однако это — тривиальный случай.

Определение ядра. ДНФ Квайна

С помощью карты Карно выписали все простые импликанты и получили сокращенную ДНФ.

Выделим общую часть, входящей во все представляющие заданную функцию ДНФ, которые можно получить из сокращенной ДНФ вычеркиванием некоторых конъюнкций.

Элементарная конъюнкция K покрывает элементарную конъюнкцию L ($K \succ L$), если любой литерал, входящий в K , входит в L .

$K \succ L \Rightarrow K \vee L = K$(согласно тождеству поглощения)

Например, $x_1x_2 \succ x_1x_2x_3$, $\quad x_1x_3 \succ x_1\overline{x}_2x_3$, но $x_1x_3 \not\succ x_1x_2\overline{x}_3$, т.к. вторая конъюнкция содержит литерал $\overline{x}_3$, отсутствующий в первой конъюнкции.

Каждая входящая в сокращенную ДНФ простая импликанта покрывает некоторую элементарную конъюнкцию исходной СДНФ. На карте Карно этому отвечает прямоугольник, закрывающий соответствующую единицу.

Простую импликанту называют **ядровой**, если она покрывает некоторую элементарную конъюнкцию исходной СДНФ, не покрываемую никакой другой простой импликантой.

На карте Карно прямоугольник, соответствующий ядровой импликанте — это такой прямоугольник, удалив который, получим единицу, не закрытую никаким другим прямоугольником.

Ни одна ядровая импликанта не может быть удалена из искомой минимальной ДНФ исходной функции.

Множество всех ядровых импликант (склеек) сокращенной ДНФ называют **ядром**.

Пример 5.11.

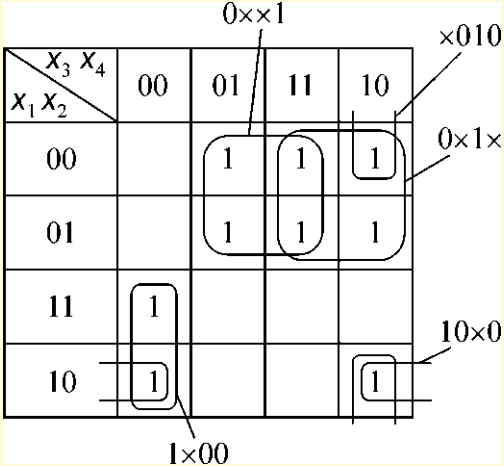

Рис. 5

На карте Карно (рис. 5) склейки $0\times\times1$, $0\times1\times$, 1×00, являются ядровыми. Две склейки 10×0 и $\times010$ не ядровые, поскольку удаление любого из изображающих их прямоугольников не приведет к появлению открытых единиц. ►

На карте Карно могут быть склейки, не являющиеся ядровыми, но покрывающие только единицы, уже покрытые ядровыми склейками.

Такие склейки оказываются лишними, т. е. удаление соответствующих им простых импликант из сокращенной ДНФ не приводит к нарушению эквивалентности этой ДНФ с исходной СДНФ.

ДНФ, получающаяся из сокращенной ДНФ после удаления всех простых импликант, соответствующих максимальным склейкам, целиком покрываемых ядром, называют **ДНФ Квайна**.

Для любой булевой функции, отличной от константы 0, существует единственная ДНФ Квайна.

Перечисление тупиковых ДНФ

Выделить из сокращенной ДНФ ядро, входящее в любую ДНФ, которую можно получить из сокращенной, и перейти от сокращенной ДНФ к ДНФ Квайна.

Дальнейшая минимизация основана на переборе всех возможных ДНФ, представляющих исследуемую функцию, которые можно получить из ДНФ Квайна путем удаления некоторых конъюнкций.

Простую импликанту называют **избыточной** (относительно некоторой ДНФ, содержащей только простые импликанты и эквивалентной исходной СДНФ), если ее можно удалить из этой ДНФ без потери эквивалентности ее исходной СДНФ.

Можно пошагово удалить избыточные импликанты, начиная с сокращенной ДНФ. В результате получим ДНФ, не содержащую ни одной избыточной склейки.

Любую ДНФ, эквивалентную исходной СДНФ, содержащую все ядровые импликанты и не содержащую ни одной избыточной импликанты, называют **тупиковой**.

Для перечисления всех тупиковых ДНФ может быть использован алгоритм, основанный на построении вспомогательной КНФ, которую называют **функцией Патрика**.

Обозначим склейки на карте Карно (т. е. простые импликанты сокращенной ДНФ) через K_1, K_2, ..., K_m.

Для каждой единицы карты Карно, не покрываемой ядром, запишем элементарную дизъюнкцию вида $K_i \vee K_j \ldots \vee K_l$, в которую включим только имена склеек, покрывающих данную единицу. Из полученных дизъюнкций составляем КНФ (**функцию Патрика**).

Пример 5.12.

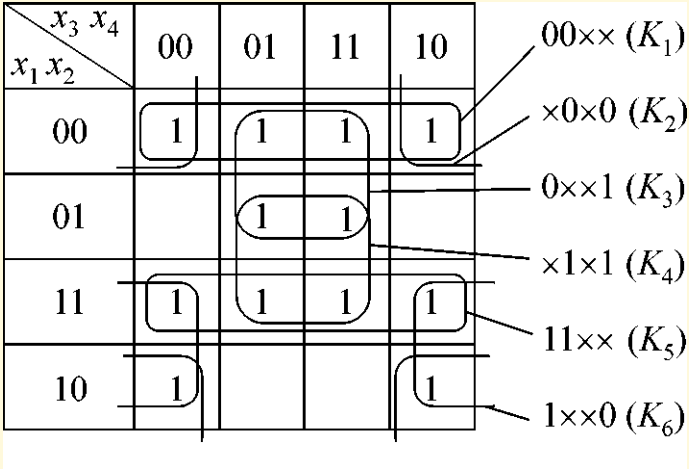

Рис. 6

Функция $f(x_1, x_2, x_3, x_4)$ представлена на карте Карно (рис. 6).

Ни одна склейка не является ядровой.

$$K_1 = \overline{x}_1\overline{x}_2\,(00\times\times); \qquad K_2 = \overline{x}_2\overline{x}_4\,(\times0\times0);$$
$$K_3 = \overline{x}_1 x_4\,(0\times\times1); \qquad K_4 = x_2 x_4\,(\times1\times1);$$
$$K_5 = x_1 x_2\,(11\times\times); \qquad K_6 = x_1\overline{x}_4\,(1\times\times0).$$

Запишем функцию Патрика, просматривая единицы слева направо по строкам карты:

$$(K_1 \vee K_2) \wedge (K_1 \vee K_3) \wedge (K_1 \vee K_3) \wedge (K_1 \vee K_2) \wedge$$
$$\wedge (K_3 \vee K_4) \wedge (K_3 \vee K_4) \wedge (K_5 \vee K_6) \wedge (K_4 \vee K_5) \wedge$$
$$\wedge (K_4 \vee K_5) \wedge (K_5 \vee K_6) \wedge (K_2 \vee K_6) \wedge (K_2 \vee K_6).$$

Применим тождество $K \wedge K = K$. Функцию можно упростить, удалив повторяющиеся дизъюнкции.

$$(K_1 \vee K_2) \wedge (K_1 \vee K_3) \wedge (K_3 \vee K_4) \wedge$$
$$\wedge (K_5 \vee K_6) \wedge (K_4 \vee K_5) \wedge (K_2 \vee K_6).$$

Скобки удобно раскрывать по парам, выбирая пары так, чтобы в них содержались одинаковые конъюнкции. Затем применить тождества поглощения $K_i \vee K_i K_j = K_i$

$$(K_1 \vee K_2) \wedge (K_1 \vee K_3) =$$
$$= K_1 \vee K_1 K_3 \vee K_1 K_2 \vee K_2 K_3 = K_1 \vee K_2 K_3.$$

Аналогично

$$(K_3 \vee K_4) \wedge (K_4 \vee K_5) = K_4 \vee K_3K_5,$$
$$(K_5 \vee K_6) \wedge (K_2 \vee K_6) = K_6 \vee K_2K_5.$$

$$
\begin{aligned}
&(K_1 \vee K_2K_3)(K_4 \vee K_3K_5)(K_6 \vee K_2K_5) = \\
&= (K_1K_4 \vee K_1K_3K_5 \vee K_2K_3K_4 \vee K_2K_3K_5)(K_6 \vee K_2K_5) \\
&= K_1K_4K_6 \vee K_1K_2K_4K_5 \vee K_1K_3K_5K_6 \vee K_1K_2K_3K_5 \vee \\
&\vee K_2K_3K_4K_6 \vee K_2K_3K_4K_5 \vee K_2K_3K_5K_6 \vee K_2K_3K_5.
\end{aligned}
$$

Упростим ДНФ, используя тождество поглощения.

$K_1K_2K_3K_5 \vee K_2K_3K_5 = K_2K_3K_5,$

$K_2K_3K_4K_5 \vee K_2K_3K_5 = K_2K_3K_5,$

$K_2K_3K_5K_6 \vee K_2K_3K_5 = K_2K_3K_5.$

В результате получим ДНФ

$$K_1K_4K_6 \vee K_1K_2K_4K_5 \vee K_1K_3K_5K_6 \vee \\ \vee K_2K_3K_4K_6 \vee K_2K_3K_5,$$

Каждая элементарная конъюнкция соответствует некоторой тупиковой ДНФ, каждой тупиковой ДНФ может быть сопоставлена одна из этих конъюнкций.

Тупиковые ДНФ можно получить на основе функции Патрика, заменив в каждой конъюнкции знаки $\wedge$ на знаки $\vee$.

1) $K_1 \vee K_4 \vee K_6 = \overline{x}_1\overline{x}_2 \vee x_2x_4 \vee x_1\overline{x}_4$;
2) $K_1 \vee K_2 \vee K_4 \vee K_5 = \overline{x}_1\overline{x}_2 \vee \overline{x}_2\overline{x}_4 \vee x_2x_4 \vee x_1x_2$;
3) $K_1 \vee K_3 \vee K_5 \vee K_6 = \overline{x}_1\overline{x}_2 \vee \overline{x}_1x_4 \vee x_1x_2 \vee x_1\overline{x}_4$;
4) $K_2 \vee K_3 \vee K_4 \vee K_6 = \overline{x}_2\overline{x}_4 \vee \overline{x}_1x_4 \vee x_2x_4 \vee x_1\overline{x}_4$;
5) $K_2 \vee K_3 \vee K_5 = \overline{x}_2\overline{x}_4 \vee \overline{x}_1x_4 \vee x_1x_2$.

Перечисление тупиковых ДНФ — самый трудоемкий этап всего алгоритма минимизации. ►

Отыскание кратчайших и минимальных ДНФ

Среди найденных тупиковых ДНФ находят кратчайшие и минимальные. Минимальная ДНФ всегда является кратчайшей, обратное неверно.

Для функции, рассмотренной в примере 5.12, первая и пятая ДНФ являются кратчайшими.

Количество литералов в обеих ДНФ совпадает, поэтому обе ДНФ минимальные. Количество отрицаний в этих ДНФ также одинаково, и по этому критерию они неразличимы.

На рис. 7 а) изображено покрытие всех единиц склейками, соответствующей первой ДНФ, а на рис. 7 б) — покрытие, соответствующее пятой ДНФ.

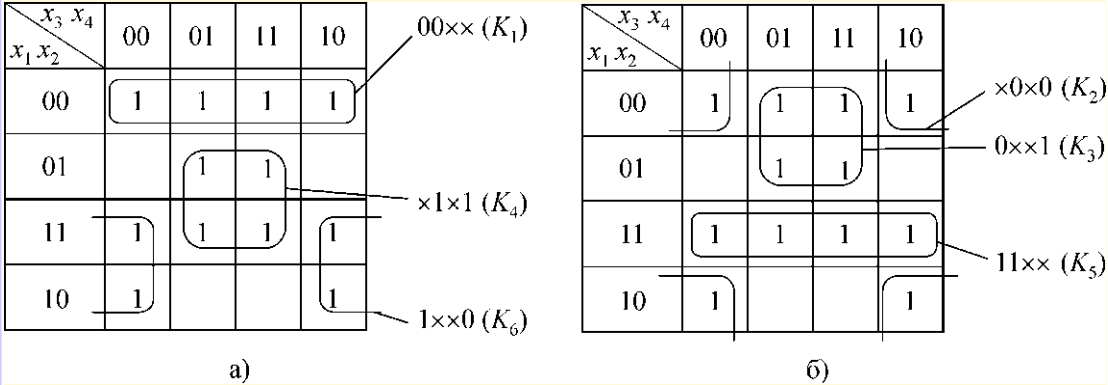

Рис. 7

Ткачев С.Б.

каф. Математического моделирования
МГТУ им. Н.Э. Баумана

ДИСКРЕТНАЯ МАТЕМАТИКА

ИУ5 — 4 семестр, 2015 г.

Лекция 6. ТЕОРЕМА ПОСТА

6.1. Полные множества булевых функций

Определение 6.1. Множество булевых функций F называют **полным**, если любая булева функция может быть представлена некоторой формулой над F.

Стандартный базис $\{\vee, \wedge, \bar{}\}$ является **полным множеством**, в силу теоремы о **представлении** любой **булевой функции дизъюнктивной** или **конъюнктивной нормальной формой**.

Теорема 1. Пусть F и G — некоторые множества булевых функций, причем F — полное множество. Тогда, если каждая функция из F может быть представлена некоторой **формулой над множеством** G, то G — полное множество. (без доказательства)#

Базис $\{\wedge, \overline{\ }\}$ является полным множеством. Согласно **законам де Моргана**, **дизъюнкцию** можно выразить через **конъюнкцию** и **отрицание**.

Базис $\{\vee, \overline{\ }\}$ является полным множеством. Согласно **закону де Моргана**, **конъюнкцию** можно выразить через **дизъюнкцию** и **отрицание**

$$x \wedge y = \overline{\overline{x} \vee \overline{y}}.$$

Пример 6.1. Рассмотрим множество, состоящее из единственной функции — **штриха Шеффера**: $\{|\}$
($x|y = \overline{x \cdot y}$).
Это множество полно, т.к. любая функция стандартного базиса может быть представленя формулой над $\{|\}$:

$$\overline{x} = (x|x),$$
$$x \cdot y = \overline{(\overline{x \cdot y})} = \overline{(x|y)} = (x|y)|(x|y),$$
$$x \vee y = \overline{\overline{x} \cdot \overline{y}} = \overline{(x|x) \cdot (y|y)} = (x|x)|(y|y).$$

6.2. Базис Жегалкина

Рассмотрим **базис Жегалкина** $\{\oplus, \cdot, 1\}$.
Чтобы доказать полноту этого множества, представим каждый элемент стандартного базиса формулой над базисом Жегалкина

$$x \vee y = x \cdot y \oplus x \oplus y, \quad \overline{x} = x \oplus 1.$$

В силу полноты стандартного базиса и теоремы 1 базис Жегалкина является полным.

Полином Жегалкина.
Общий вид полинома Жегалкина от трех переменных:

$$a_{123}x_1x_2x_3 \oplus a_{12}x_1x_2 \oplus a_{13}x_1x_3 \oplus \\ \oplus a_{23}x_2x_3 \oplus a_1x_1 \oplus a_2x_2 \oplus a_3x_3 \oplus a_0 \quad (6.1)$$

Полином Жегалкина от n переменных можно записать в виде

$$P(x_1, \ldots, x_n) = \sum_{\{i_1,\, i_2,\, \ldots,\, i_m\} \subseteq I} (\mathrm{mod}\ 2)\, a_{i_1 i_2 \ldots i_m} x_{i_1} x_{i_2} \ldots x_{i_m},$$

где $I = \{1, 2, \ldots, n\}$, а коэффициенты полинома $a_{i_1 i_2 \ldots i_m} \in \{0, 1\}$ индексированы всеми возможными подмножествами множества $\{1, 2, \ldots, n\}$ (коэффициент a_0 соответствует пустому множеству).

Утверждение 6.1. Полином Жегалкина для любой булевой функции определен однозначно.

Доказательство.
Количество коэффициентов в полиноме Жегалкина от n переменных равно числу подмножеств множества $\{1, 2, \ldots, n\}$, т.е. 2^n. Каждый коэффициент может принимать два значения — 0 и 1. Следовательно, различных полиномов Жегалкина столько же, сколько булевых функций от n переменных — $2^{(2^n)}$.

Для функций от небольшого числа переменных (не превышающего 4) можно использовать **метод неопределенных коэффициентов**, позволяющий получить полином Жегалкина данной функции. Проиллюстрируем этот метод на примере.

Пример 6.2. Пусть $f = (1,1,0,0,1,0,1,1)$. Найдем полином Жегалкина, представляющий f.

Функция f представляется некоторым полиномом Жегалкина третьей степени, общий вид которого дает формула

$$a_{123}x_1x_2x_3 \oplus a_{12}x_1x_2 \oplus a_{13}x_1x_3 \oplus \\ \oplus a_{23}x_2x_3 \oplus a_1x_1 \oplus a_2x_2 \oplus a_3x_3 \oplus a_0 \quad (6.2)$$

Значение функции f на наборе 000 равно коэффициенту a_0 :

$$f(0,0,0) = a_0 = 1.$$

Чтобы найти коэффициенты a_3 , a_2 и a_1 , нужно рассмотреть значения функции на наборах 001 , 010 и 100 соответственно.

$$f(0,0,1) = a_3 \oplus a_0 = a_3 \oplus 1 = 1,$$

Решая это уравнение относительно a_3 в поле $\mathbb{Z}_2$, получим $a_3 = 0$;

$$f(0,1,0) = a_2 \oplus 1 = 0 \ \Rightarrow \ a_2 = 1;$$

$$f(1,0,0) = a_1 \oplus 1 = 1 \ \Rightarrow \ a_1 = 0;$$

Чтобы найти коэффициенты a_{12}, a_{13} и a_{23}, нужно рассмотреть значения функции на наборах 110, 101 и 011 соответственно.

Для первого набора получим

$$f(1,1,0) = a_{12}x_1x_2 \oplus a_1x_1 \oplus a_2x_2 \oplus a_0 = \\ = a_{12} \oplus a_2 \oplus a_0 = a_{12} \oplus 1 \oplus 1 = a_{12}$$

(сумма по модулю 2 любого четного числа равных слагаемых равна 0).
Поскольку $f(1,1,0) = 1$, то $a_{12} = 1$.
Аналогично находим a_{13}. $f(1,0,1) = 0$

$$f(1,0,1) = a_{13} \oplus a_0 = a_{13} \oplus 1 = 0$$

откуда $a_{13} = 1$;

$$f(0,1,1) = a_{23} \oplus a_2 \oplus a_0 = a_{23} \oplus 1 \oplus 1 = 0,$$

$a_{23} = 0$.

$$f(1,1,1) = a_{123} \oplus a_{12} \oplus a_{13} \oplus a_2 \oplus a_0 = a_{123} = 1.$$

Полином Жегалкина, представляющий f есть:

$$f = x_1x_2x_3 \oplus x_1x_2 \oplus x_1x_3 \oplus x_2 \oplus 1.$$

6.3. Классы Поста

Рассмотрим некоторые специальные множества функций.

Определение 6.2. Функцию f называют **функцией, сохраняющей константу 0**, если $f(\tilde{0}) = 0$, где $\tilde{0}$ — нулевой набор значений переменных функции f.

Определение 6.3. Функцию f называют **функцией, сохраняющей константу константу 1**, если $f(\tilde{1}) = 1$, где $\tilde{1}$ — единичный набор значений переменных функции f.

Например, функция $f = (00111101)$ является функцией, сохраняющей и константу 0, и константу 1.
Отрицание не сохраняет ни 0, ни 1.

Множество всех функций, сохраняющих константу 0 обозначается через T_0 .

Множество всех функций, сохраняющих константу 1 обозначается через T_1 .

Наборы $\tilde{\alpha}$ и $\overline{\tilde{\alpha}}$ из булева куба $\mathbb{B}^n = \{0, 1\}^n$ (для произвольного фиксированного n) будем называть **взаимно противоположными**.

Говорят, что набор $\overline{\tilde{\alpha}}$ есть **отрицание набора** $\tilde{\alpha}$.

Определение 6.4. Функцию $g \in \mathcal{P}_{2,n}$ называют **двойственной к функции** $f \in \mathcal{P}_{2,n}$, если для всякого $\tilde{\alpha} \in \{0, 1\}^n$ ($n > 0$) имеет место

$$g(\tilde{\alpha}) = \overline{f}(\overline{\tilde{\alpha}}).$$

Константа 0 является двойственной к константе 1 и наоборот.

Пример 6.3.

а. Стрелка Пирса есть функция, двойственная к штриху Шеффера, так как

$$x \downarrow y = \overline{x \vee y} = \overline{\overline{x} \cdot \overline{y}} = \overline{x|y}.$$

б. Сумма по модулю 2 двойственна к эквивалентности, так как

$$x \sim y = \overline{x \oplus y} = \overline{\overline{x} \oplus \overline{y}}.$$

В общем случае в силу свойства единственности отрицания функция h, двойственная к функции g, которая двойственна к f, равна f.

Определение 6.5. Функцию $f \in \mathcal{P}_{2,n}$ называют **самодвойственной**, если она двойственна к себе самой, т.е.

$$(\forall \tilde{\alpha} \in \{0,\ 1\}^n)(f(\tilde{\alpha}) = \overline{f}(\overline{\tilde{\alpha}})),$$

или

$$(\forall \tilde{\alpha} \in \{0,\ 1\}^n)(f(\overline{\tilde{\alpha}}) = \overline{f}(\tilde{\alpha})).$$

Функция самодвойственна тогда и только тогда, когда на взаимно противоположных наборах она принимает взаимно противоположные значения.

Для того чтобы убедиться в несамодвойственности заданной функции f, достаточно найти хотя бы одну пару взаимно противоположных наборов $\tilde{\alpha}$ и $\overline{\tilde{\alpha}}$, таких, что значения функции на них совпадают, т.е. $f(\tilde{\alpha}) = f(\overline{\tilde{\alpha}})$.

Так, функция $f_1 = (0101)$ является самодвойственной, поскольку

$$f_1(0,0) = 0 = \overline{f}_1(\overline{0},\overline{0}) = \overline{f}_1(1,1) = \overline{1} = 0,$$

$$f_1(0,1) = 1 = \overline{f}_1(\overline{0},\overline{1}) = \overline{f}_1(1,0) = \overline{0} = 1.$$

Функция $f_2 = (1001)$ (эквивалентность) не является самодвойственной, поскольку при $\tilde{\alpha} = (0,\,0)$ $0 \sim 0 = 1$ и $1 \sim 1 = 1$.

Множество всех самодвойственных функций (при всех $n \geq 1$) обозначим S.

Определение 6.6. Функцию $f \in \mathcal{P}_{2,n}$ называют **монотонной**, если для любых наборов $\tilde{\alpha}$, $\tilde{\beta} \in \mathbb{B}^n$, таких, что $\tilde{\alpha} \leq \tilde{\beta}$, имеет место $f(\tilde{\alpha}) \leq f(\tilde{\beta})$.

Функция $f = (0011)$ монотонна. Штрих Шеффера — немонотонная функция, так как $00 < 11$, но $0|0 = 1$, а $1|1 = 0$. Множество всех монотонных функций принято обозначать через M.

Формула вида

$$\sum_{i=1}^{n} (\mathrm{mod}\ 2)\, a_i x_i \oplus a_0 \tag{6.3}$$

называется **полиномом Жегалкина первой степени** от переменных. В таком полиноме отсутствуют „нелинейные“ слагаемые.

Определение 6.7. Функцию $f \in \mathcal{P}_{2,n}$ называют **линейной**, если она может быть представлена полиномом Жегалкина первой степени от n переменных.

Например, $1 \oplus x_1 \oplus x_2 \oplus x_3$ — полином первой степени от 3 переменных.
Множество всех линейных функций обозначают через L.

Определение 6.8. Множества функций T_0, T_1, S, M, L называются **классами Поста**.

Пример 6.4. Штрих Шеффера не принадлежит ни одному из классов Поста.
Все свойства, кроме нелинейности, следуют из таблицы этой функции. Нелинейность же доказывается выводом полинома Жегалкина для штриха Шеффера:

$$x|y = \overline{x \cdot y} = x \cdot y \oplus 1.$$

Полученный полином Жегалкина имеет степень выше первой.

Определение 6.9. Множество булевых функций F называют **замкнутым**, если любая формула над F представляет некоторую функцию из F.

Фундаментальным свойством каждого класса Поста является его замкнутость (в смысле определения 6.9).

Определение 6.10. Функция $f(g_1, \ldots, g_n)$ называется **суперпозицией функций** f, g_1, $\ldots$, g_n. Для любого $\tilde{\alpha} \in \mathbb{B}^m$ имеет место равенство

$$f(g_1, \ldots, g_n)(\tilde{\alpha}) = f(g_1(\tilde{\alpha}), \ldots, g_n(\tilde{\alpha})).$$

Замкнутость классов Поста: для любого из классов Поста C всякая **суперпозиция над** C снова есть элемент C.

Теорема 2. Каждый класс Поста замкнут.

◄ Для каждого класса Поста $C \in \{T_0, T_1, S, M, L\}$ нужно доказать, что замыкание $[C]$ множества булевых функций C совпадает с C, т.е. любая функция, представляемая формулой построенной над классом C, принадлежит этому классу.

Пусть $f(g_1, \ldots, g_n)$ — какая-то суперпозиция над C. Обозначим ее через φ.

Без ограничения общности можно считать, что все функции f, g_1, $\ldots$, g_n зависят от n переменных (для некоторого n).

1. Рассмотрим $C = T_0$.

Пусть f, $g_1, \ldots, g_n \in T_0$, т.е. $f(\tilde{0}) = 0$ и $g_i(\tilde{0}) = 0$.

Тогда $\varphi(\tilde{0}) = f(g_1(\tilde{0}), \ldots, g_n(\tilde{0})) = f(0, \ldots, 0) = 0$.

Следовательно, $\varphi \in T_0$.

2. При $C = T_1$ рассуждаем точно так же.

3. Пусть $C = S$, т.е. f, $g_1, \ldots, g_n \in S$. Докажем, что $\varphi = f(g_1, \ldots, g_m) \in S$.

Фиксируем произвольный набор $\tilde{\alpha} \in \{0, 1\}^n$ и покажем, что $\varphi(\overline{\tilde{\alpha}}) = \overline{\varphi}(\tilde{\alpha})$, используя самодвойственность всех функций:

$$\begin{aligned}\varphi(\overline{\tilde{\alpha}}) &= f(g_1(\overline{\tilde{\alpha}}), \ldots, g_n(\overline{\tilde{\alpha}})) = \\ &= f(\overline{g_1}(\tilde{\alpha}), \ldots, \overline{g_n}(\tilde{\alpha})) = \\ &= \overline{f}(g_1(\tilde{\alpha}), \ldots, g_n(\tilde{\alpha})) = \overline{\varphi}(\tilde{\alpha}).\end{aligned}$$

Следовательно, $\varphi \in S$.

4. $C = M$, т.е. f, $g_1, \ldots, g_n \in S$.
Берем произвольно наборы $\tilde{\alpha}$ и $\tilde{\beta}$ так, что $\tilde{\alpha} \leq \tilde{\beta}$.
Докажем, что $\varphi = f(g_1, \ldots, g_m) \in M$. Имеем

$$\varphi(\tilde{\alpha}) = f(g_1(\tilde{\alpha}), \quad \ldots, \quad g_n(\tilde{\alpha})) \leq f(g_1(\tilde{\beta}), \quad \ldots, \quad g_n(\tilde{\beta}))$$

так как все функции g_i, $i = \overline{1, n}$, монотонны, вектор $(g_1(\tilde{\alpha}), \ldots, g_n(\tilde{\alpha}))$ не больше вектора $(g_1(\tilde{\beta}), \ldots, g_n(\tilde{\beta}))$, функция f также монотонна. Следовательно, $\varphi \in M$.

5. Если же $C = L$, то очевидно, что при подстановке в линейную функцию (полином Жегалкина первой степени) вместо ее переменных произвольных линейных функций получится снова линейная функция.
Доказана замкнутость каждого класса Поста. ►

Приведем теорему, характеризующую важное свойство немонотонных функций.

Теорема 3. Если функция f не является монотонной, т.е. $f \notin M$, то найдутся два таких набора $\tilde{\alpha}$, $\tilde{\beta}$, что

$$\tilde{\alpha} = (\alpha_1, \ldots, \alpha_{i-1}, 0, \alpha_{i+1}, \ldots, \alpha_n),$$
$$\tilde{\beta} = (\alpha_1, \ldots, \alpha_{i-1}, 1, \alpha_{i+1}, \ldots, \alpha_n),$$

и $f(\tilde{\alpha}) = 1$, $f(\tilde{\beta}) = 0$, т.е. эти два набора различаются значениями в точности одной компоненты, а значение функции равно 0 на большем наборе и равно 1 на меньшем.

6.4. Реализация функций формулами

Реализация констант 0 и 1.

Рассмотрим два случая реализации **констант** 0 и 1 .

1. Пусть имеется функция f_0, не сохраняющая константу 0 ($f_0 \notin T_0$), и сохраняющая константу 1 ($f_0 \in T_1$), т.е. $f_0(0,\ldots,0) = 1$ и $f_0(1,\ldots,1) = 1$.

Константа 1 представляется формулой

$$1 = f_0(x,\ldots,x).$$

Чтобы выразить константу 0, используем любую функцию $g \in F$, не сохраняющую константу 1 ($g \notin T_1$):

$$0 = g(1,\ldots,1) = g(f_0(x,\ldots,x),\ldots,f_0(x,\ldots,x)).$$

Пусть имеется функция f_1, не сохраняющая константу 1 ($f_0 \notin T_1$), но сохраняющая константу 0 ($f_0 \in T_0$), т.е. $f_0(0,\ldots,0) = 0$ и $f_0(1,\ldots,1) = 0$.
Константа 0 представляется формулой

$$0 = f_0(x,\ldots,x).$$

Если существует функция $g \notin T_0$, то

$$1 = g(0,\ldots,0) = g(f_0(x,\ldots,x),\ldots,f_0(x,\ldots,x)).$$

2. Реализация **констант** 0 и 1 из **несамодвойственной** функции с использованием отрицания.

Утверждение. Если функция несамодвойственная, то с использованием отрицания из нее можно реализовать константу.

Пусть функция f_S — несамодвойственная. Тогда найдется такой набор $\tilde{\alpha} = (\alpha_1, \ldots, \alpha_n)$, что

$$f_S(\overline{\tilde{\alpha}}) = f_S(\tilde{\alpha}).$$

Введем функцию от одного переменного

$$h(x) = f_S(x^{\alpha_1}, \ldots, x^{\alpha_n}).$$

Заметим, что

$$0^\sigma = \begin{cases} 1, & \sigma = 0; \\ 0, & \sigma = 1. \end{cases}$$

Поэтому $0^{\alpha_i} = \overline{\alpha_i}$. Аналогично $1^{\alpha_i} = \alpha_i$.
Получим

$$h(0) = f_S(\overline{\tilde{\alpha}}) = f_S(\tilde{\alpha}) = h(1),$$

Таким образом, значение $h(x)$ есть константа.

Реализация отрицания.

Утверждение. Если функция f_M **немонотонная**, то с использованием констант 0 и 1 из нее можно реализовать отрицание.

Для немонотонной функции f_M согласно теореме 3 найдутся два таких набора $\tilde{\alpha}$ и $\tilde{\beta}$, что

$$\tilde{\alpha} = (\alpha_1, \ldots, \alpha_{i-1}, 0, \alpha_{i+1}, \ldots, \alpha_n),$$
$$\tilde{\beta} = (\alpha_1, \ldots, \alpha_{i-1}, 1, \alpha_{i+1}, \ldots, \alpha_n),$$

$f_M(\tilde{\alpha}) = 1$, а $f_M(\tilde{\beta}) = 0$.

Тогда

$$\overline{x} = f_M(\alpha_1, \ldots, \alpha_{i-1}, x, \alpha_{i+1}, \ldots, \alpha_n),$$

где $\alpha_1, \ldots, \alpha_{i-1}, \alpha_{i+1}, \ldots, \alpha_n \in \{0, 1\}$.

Частный случай. Если для немонотонной функции имеет место $f_M \notin T_0$ и $f_M \notin T_1$, то

$$f_M(0, \ldots, 0) = 1,$$

$$f_M(1, \ldots, 1) = 0.$$

Тогда

$$\overline{x} = f_2(x, \ldots, x).$$

Реализация конъюнкции с использованием констант и отрицания.

Утверждение. Если функция f_L нелинейная, то с использованием констант и отрицания из нее можно реализовать конъюнкцию.

В полиноме Жегалкина этой функции выбираем произвольное нелинейное слагаемое, содержащее наименьшее число переменных.
Пусть это будет слагаемое $x_{i_1}, \ldots, x_{i_k}$ при $2 \leq k \leq n$.
Вместо каждого переменного x_m функции f_L, где $m \notin \{i_1, \ldots, i_k\}$, подставим константу 0.

Получим новую функцию

$$f'_L(x_{i_1}, \ldots, x_{i_k}) = \\ = x_{i_1} \ldots x_{i_k} \oplus a_{i_1} x_{i_1} \oplus \ldots \oplus a_{i_k} x_{i_k} \oplus a_0 = \\ = f_L(0, \ldots, 0, x_{i_1}, 0, \ldots, 0, x_{i_k}, 0, \ldots, 0).$$

Разобьем множество переменных $\{x_{i_1}, \ldots, x_{i_k}\}$ на две части: $\{x_{i_1}, \ldots, x_{i_m}\}$ и $\{x_{i_{m+1}}, \ldots, x_{i_k}\}$, где $1 \leq m \leq k-1$ так, чтобы после замены всех переменных первой части переменным x, а переменных второй части — переменным y, получить функцию от двух переменных

$$\chi(x, y) = xy \oplus ax \oplus by \oplus c,$$

где $a = a_{i_1} \oplus \ldots \oplus a_{i_m}$, $b = a_{i_{m+1}} \oplus \ldots \oplus a_{i_k}$, $c = a_0$.

Функция χ может быть представлена такой формулой над F:

$$\chi(x,y) = f_L(0,\ldots,0,\underbrace{x}_{i_1},0,\ldots,0,\underbrace{x}_{i_m},0,\ldots,0,$$
$$\underbrace{y}_{i_{m+1}},0,\ldots,0,\underbrace{y}_{i_k},0,\ldots,0),$$

Определим функцию

$$\psi(x,y) = \chi(x \oplus b, y \oplus a) \oplus ab \oplus c.$$

Выразив функцию $\psi(x, y)$ из полинома Жегалкина для χ, получим

$$\psi(x, y) = \chi(x \oplus b, y \oplus a) \oplus ab \oplus c =$$
$$= (x \oplus b)(y \oplus a) \oplus a(x \oplus b) \oplus b(y \oplus a) \oplus c \oplus ab \oplus c =$$
$$= xy \oplus ax \oplus by \oplus ab \oplus ax \oplus ab \oplus by \oplus ab \oplus c \oplus ab \oplus c =$$
$$= xy,$$

т.к. сумма по модулю 2 любого четного числа равных слагаемых равна 0.

Функция ψ есть конъюнкция.

Прибавление к любой функции константы по модулю 2 есть либо сама исходная функция, либо ее отрицание. Поскольку отрицание доступно, то конъюнкция реализована формулой.

Теорема 4 (критерий Поста). Множество F булевых функций полно тогда и только тогда, когда оно не содержится целиком ни в одном из классов Поста.

◄ **Необходимость.** Пусть множество F булевых функций полно. Предположим, что оно содержится целиком в одном из классов Поста, то есть для некоторого класса Поста C выполняется $F \subseteq C$.

Всякая суперпозиция над F, согласно теореме 2, снова лежала бы в C.

Существуют функции, не содержащиеся ни в одном из классов Поста, например штрих Шеффера.

Таким образом, нашлась функция, которую нельзя представить в виде суперпозиции над F, что противоречит предположению о полноте F.

Достаточность. Для доказательства полноты множества F, удовлетворяющего условию теоремы, построим формулы над F для **отрицания** и **конъюнкции**, поскольку множество, образованное этими функциями, полно. Тогда в силу теоремы 1 будет полным и множество F.

По условию теоремы в F найдется хотя бы одна функция $f_1 \notin T_0$. Если $f_1 \in T_1$, то можно реализовать константу 1. Если $f_1 \notin T_1$, то можно реализовать отрицание.

По условию теоремы в F найдется хотя бы одна функция $f_2 \notin T_1$. Если $f_2 \in T_0$, то можно реализовать константу 0. Если $f_2 \notin T_1$, то можно реализовать отрицание.

Таким образом, могут быть реализованы либо две константы 0 и 1, либо только отрицание, либо константы и отрицание.

По условию теоремы в F найдется хотя бы одна **немонотонная** функция, хотя бы одна **несамодвойственная** функция и хотя бы одна **нелинейная** функция.

В случае, если из первых двух классов (T_0 , T_1) построены только формулы для констант, с использованием немонотонной функции $f_M \notin M$ можно реализовать отрицание.

Если реализовано только отрицание, с использованием несамодвойственной функции $f_S \notin S$ можно реализовать константы.

Имея константы и отрицание, из нелинейной функции $f_L \notin L$ можно реализовать коньюнкцию.

Таким образом, отрицание и конъюнкция реализованы формулами над F . Множество полно. ►

Чтобы исследовать полноту конкретного множества функций

$$F = \{f_1,\ f_2,\ \ldots,\ f_n\},$$

используют **критериальную таблицу**

Таблица 6.1

	T_0	T_1	S	M	L
f_1					
f_2					
$\vdots$					
f_n					

Строки таблицы соответствуют функциям исследуемого множества, а столбцы — классам Поста.

Пример 6.5. Пусть $F = \{\sim, \vee, 0\}$.
Заполненная критериальная таблица:

Таблица 6.2

	T_0	T_1	S	M	L
$\sim$	$-$	$+$	$-$	$-$	$+$
$\vee$	$+$	$+$	$-$	$+$	$-$
0	$+$	$-$	$-$	$+$	$+$

В множестве F есть функции, не принадлежащие каждому из пяти классов Поста. Согласно теореме Поста, множество F — полное.

Реализация констант, отрицания и конъюнкции над F

Константа 0 принадлежит самому множеству F.

Функция $\sim$ (эквивалентность) не сохраняет константу 0, но сохраняет константу 1, поэтому $1 = x \sim x$.

Поскольку $0 \sim 0 = 1$, $1 \sim 0 = 0$, то $\overline{x} = x \sim 0$.

Конъюнкцию можно представить формулой над F, следуя доказательству теоремы Поста.
Берем единственную нелинейную функцию данного множества, дизъюнкцию, и записываем для нее полином Жегалкина:

$$x_1 \vee x_2 = x_1 \cdot x_2 \oplus x_1 \oplus x_2.$$

Этот полином есть функция

$$\chi(x_1, x_2) = xy \oplus ax \oplus by \oplus c$$

при $a = b = 1$ и $c = 0$.
Следовательно,

$$x_1 \cdot x_2 = \chi(x_1 \oplus 1, x_2 \oplus 1) \oplus 1.$$

Так как $x \oplus 1 = \overline{x} = x \sim 0$, то

$$x_1 \cdot x_2 = (\underbrace{(x_1 \sim 0)}_{x_1 \oplus 1} \vee \underbrace{(x_2 \sim 0)}_{x_2 \oplus 1}) \sim 0.$$

Этот же результат в данном конкретном случае можно получить и гораздо проще:

$$x_1 \cdot x_2 = \overline{\overline{x}_1 \vee \overline{x}_2} = ((x_1 \sim 0) \vee (x_2 \sim 0)) \sim 0.$$

Замечание. Это полное множество $F = \{\sim, \vee, 0\}$ двойственно к базису Жегалкина в том смысле, что каждая из его функций двойственна к соответствующей функции базиса Жегалкина: эквивалентность двойственна к сумме по модулю 2, дизъюнкция — к конъюнкции, константа 0 — к константе 1.

Никакое собственное подмножество заданного множества не будет полным.

Ткачев С.Б.
каф. Математического моделирования
МГТУ им. Н.Э. Баумана

ДИСКРЕТНАЯ МАТЕМАТИКА

ИУ5 — 4 семестр, 2015 г.

Лекция 8. ЭЛЕМЕНТЫ ОБЩЕЙ АЛГЕБРЫ. ПОЛУГРУППЫ И ГРУППЫ

Предметом рассмотрения в абстрактной алгебре являются произвольные множества с заданными на них операциями.

Природа множеств и операций может существенно отличаться от привычных числовых множеств и известных операций над числами.

8.1. Операции. Понятие алгебраической структуры

Определение 8.1. Пусть A — произвольное непустое множество и n — натуральное число. Любое отображение

$$\omega\colon A^n \to A$$

называют **n -арной** (или **n -местной**) **операцией** на множестве A.

n -арная операция ω каждому *кортежу* $(a_1, \ldots, a_n) \in A^n$ однозначно сопоставляет элемент $b \in A$.

Компоненты кортежа называют **аргументами операции** ω, а b — **результатом применения операции** ω к аргументам $a_1, \ldots, a_n$.

Для n -арной операции используют обозначение

$$b = \omega(a_1, \ldots, a_n)$$

или

$$b = a_1 \ldots a_n \omega.$$

Обычно, если $n = 2$, пишут $a_1 \, \omega \, a_2$.

При $n = 1$ говорят об **унарной операции**.

При $n = 2$ — о **бинарной операции**.

Пример унарной операции — *дополнение* заданного *множества* до *универсального множества*.

Примеры бинарных операций:
— сложение и умножение чисел,

— сложение и умножение матриц квадратных матриц типа $n \times n$,

— сложение векторов линейного пространства.

Специально вводят понятие **нульарной операции** (т.е. при $n = 0$).

Под нульарной операцией на множестве A понимают произвольный фиксированный элемент множества A .

Нульарные операции позволяют фиксировать элементы множества A , обладающие некоторыми специальными свойствами.

Пример нульарной операции — фиксирование нуля в множестве целых чисел с операцией сложения.

Рассмотрим бинарную операцию на множестве A, обозначив ее звездочкой $(*)$.
Эту операцию называют:

1) **ассоциативной**, если $(x*y)*z = x*(y*z)$ для любых x, y, $z \in A$;

2) **коммутативной**, если $x*y = y*x$ для любых x, $y \in A$;

3) **идемпотентной**, если $x*x = x$ для любого $x \in A$.

Ассоциативность операции $*$ позволяет для любых элементов a_1, a_2, $\ldots$, $a_n \in A$ однозначно трактовать результат выражения $a_1 * a_2 * \ldots * a_n$, так как

$$a_1 * a_2 * \ldots * a_n = a_1 * (a_2 * \ldots * a_n) =$$
$$= (a_1 * a_2) * (a_3 * \ldots * a_n) = (a_1 * a_2 * \ldots * a_{n-1}) * a_n.$$

Операция сложения, заданная на множестве натуральных чисел, является ассоциативной и коммутативной.

Операция умножения матриц ассоциативна, но не коммутативна.

Идемпотентными являются операции объединения и пересечения множеств.

Определение 8.2. Элемент $\mathbf{1}$ множества A называют **левым нейтральным элементом** относительно операции $*$, если $\mathbf{1} * x = x$ для любого элемента $x \in A$.

Определение 8.3. Элемент $\mathbf{1}$ множества A называют **правым нейтральным элементом** относительно операции $*$, если $x * \mathbf{1} = x$ для любого элемента $x \in A$.

Если существуют левый ($\mathbf{1}'$) и правый ($\mathbf{1}''$) нейтральные элементы, то они совпадают. $\mathbf{1}' = \mathbf{1}' * \mathbf{1}'' = \mathbf{1}''$.

В этом случае элемент $\mathbf{1}$ единственный, и его называют просто **нейтральным элементом**.

Нейтральным элементом относительно операции умножения на множестве натуральных чисел является число 1.

На множестве целых чисел нейтральным элементом относительно операции сложения будет число 0.

Пример 8.1.

На множестве квадратных матриц вида $\begin{pmatrix} a & 0 \\ b & 0 \end{pmatrix}$, где элементы a и b — действительные числа, любая матрица вида $\begin{pmatrix} 1 & 0 \\ d & 0 \end{pmatrix}$ будет правым нейтральным элементом по операции умножения.

$$\begin{pmatrix} a & 0 \\ b & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ d & 0 \end{pmatrix} = \begin{pmatrix} a & 0 \\ b & 0 \end{pmatrix}.$$

Правых нейтральных элементов бесконечно много.

Левого нейтрального элемента по этой операции нет (иначе существовал бы единственный нейтральный элемент). #

Определение 8.4. Алгебра (универсальная алгебра, Ω-алгебра) считается заданной, если заданы некоторое множество A, называемое **носителем** данной **алгебры**, и некоторое множество операций Ω на A, называемое **сигнатурой** данной **алгебры**.

Алгебра — упорядоченная пара множеств $\mathcal{A} = (A, \Omega)$, первая компонента этой пары есть носитель, а вторая — сигнатура.

Результат применения любой операции обязательно должен принадлежать тому же множеству, что и ее аргументы.

Рассмотрим множество V_3 всех свободных векторов в пространстве и операцию скалярного умножения векторов. Это **не алгебра**, т.к. скалярное произведение двух векторов не есть вектор.

Множество V_3 всех свободных векторов в пространстве и операция векторного умножения векторов **является** алгеброй.

Пара $(\mathbb{R} \setminus \{0\},\ \{:\})$ есть алгебра.

Пример 8.2.
Для любого множества M можно определить алгебру

$$\mathcal{A}_2 = \left(2^{M\times M},\ \{\cup,\ \circ,\ ^{-1}\}\right),$$

носителем которой является множество всех подмножеств множества упорядоченных пар на M, т.е. множество всех *бинарных отношений на множестве* M.
Сигнатура состоит из операций объединения, композиции бинарных отношений и взятия обратного отношения. #

8.2. Группоиды, полугруппы, группы

Рассмотрим алгебры, сигнатуры которых состоят из одной бинарной операции. Эту операцию будем обозначать точкой $(\cdot)$ и условно называть в этом случае умножением.

Группоидом называют любую алгебру $\mathcal{G} = (G, \cdot)$, сигнатура которой состоит из одной бинарной операции. В группоиде на бинарную операцию нет никаких ограничений.

Группоид $(G, \cdot)$ называют **полугруппой**, если его операция **ассоциативна**, т.е. для любых элементов a, b, c *носителя* G выполняется равенство $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.

Группоид $\mathcal{G} = (G, \cdot)$ называют **моноидом**, если его операция **ассоциативна** и относительно операции существует **нейтральный элемент**.
Его называют **нейтральным элементом моноида** $\mathcal{G}$ или **единицей моноида** и обозначают $\mathbf{1}$.

Моноид $\mathcal{G} = (G, \cdot)$ есть полугруппа, в которой для любого a имеют место равенства $a \cdot \mathbf{1} = \mathbf{1} \cdot a = a$, где $\mathbf{1}$ — нейтральный элемент (единица) моноида.

При задании моноида можно в сигнатуре указать только бинарную операцию, описав ее свойства дополнительно, а можно включить в сигнатуру нульарную операцию — нейтральный элемент моноида. На практике используют оба способа.

Пример 8.3. а. $(2^{A\times A},\ \circ,\ \mathrm{id}_A)$ Множество всех бинарных отношений на произвольном множестве A с операцией композиции отношений будет моноидом.
Для любых бинарных отношений ρ, τ и σ на множестве A имеют место равенства $\rho \circ (\tau \circ \sigma) = (\rho \circ \tau) \circ \sigma$ — операция ассоциативна.

Нейтральным элементом будет диагональ множества $A \times A$, поскольку $\mathrm{id}_A \circ \rho = \rho \circ \mathrm{id}_A = \rho$.

Группоид $\mathcal{G} = (G, \cdot)$ называют **группой**, если

1) операция $\cdot$ ассоциативна,
2) существует нейтральный элемент (единица) $\mathbf{1}$ относительно умножения,
3) для каждого $a \in G$ существует такой элемент $a' \in G$, называемый **обратным** к a, что $a \cdot a' = a' \cdot a = \mathbf{1}$.

Группа — это моноид, в котором для **каждого** элемента существует обратный элемент.

Теорема 1. В любой группе $\mathcal{G} = (G, \cdot)$ для каждого $a \in G$ элемент, обратный к a, единственный.

◄ Пусть в группе $(G, \cdot)$ с единицей $\mathbf{1}$ для некоторого a существуют два элемента a' и a'', обратных к a.
Тогда $a' = a' \cdot \mathbf{1}$ в силу свойства единицы.
Так как $\mathbf{1} = a \cdot a''$, то $a' = a' \cdot (a \cdot a'')$.
Используя ассоциативность и учитывая, что a' — элемент, обратный к a, получим

$$a' \cdot (a \cdot a'') = (a' \cdot a) \cdot a'' = \mathbf{1} \cdot a'' = a''. \qquad \blacktriangleright$$

Полугруппа, операция которой коммутативна, называется **коммутативной полугруппой**.

Моноид, операция которого коммутативна, называется **коммутативный моноид.**

Среди групп также выделяют те, бинарная операция в которых коммутативна, — **коммутативные (абелевы) группы**.

В коммутативных полугруппах и группах бинарную операцию часто обозначают знаком $+$ и называют **сложением**. Нейтральный элемент (если он существует) обозначают знаком $\mathbf{0}$ и называют нулем.

Свойства операции вычисления обратного элемента.

Теорема 2. Пусть $\mathcal{G} = (G, \cdot)$ — группа. Для любых элементов $a, b \in G$ верны тожества

$$(a \cdot b)^{-1} = b^{-1} \cdot a^{-1}; \quad (1)$$
$$(a^{-1})^{-1} = a. \quad (2)$$

◀ Покажем

$$(a \cdot b)^{-1} = b^{-1} \cdot a^{-1}.$$

В силу ассоциативности умножения группы имеем

$$(a \cdot b) \cdot (b^{-1} \cdot a^{-1}) = ((a \cdot b) \cdot b^{-1}) \cdot a^{-1}.$$

Используя еще раз ассоциативность, определение элемента, обратного к данному, и свойства единицы, получим

$$((a \cdot b) \cdot b^{-1}) \cdot a^{-1} = a \cdot (b \cdot b^{-1}) \cdot a^{-1} = a \cdot a^{-1} = \mathbf{1}.$$

Итак, $(a \cdot b) \cdot (b^{-1} \cdot a^{-1}) = \mathbf{1}$ — для $a \cdot b$ найден правый обратный.

Точно так же доказывается, что $(b^{-1} \cdot a^{-1})(a \cdot b) = \mathbf{1}$, т.е. найден левый обратный.

Поэтому элемент $b^{-1} \cdot a^{-1}$ является обратным к элементу $a \cdot b$.

Обратный элемент единственный в силу теоремы о единственности обратного элемента в группе, и поэтому

$$(a \cdot b)^{-1} = b^{-1} \cdot a^{-1}.$$

Равенство $(a^{-1})^{-1} = a$ следует непосредственно из определения элемента, обратного к данному.

Действительно, определение элемента a^{-1}, обратного к a, равенством $a^{-1} \cdot a = a \cdot a^{-1} = \mathbf{1}$ можно рассматривать как определение $(a^{-1})^{-1}$ — обратного элемента к a^{-1}, которым является, согласно этим равенствам, элемент a. Обратный элемент единственный в силу теоремы о единственности обратного элемента в группе, т.е. $a = (a^{-1})^{-1}$. ▶

Теорема 3. В любой группе $\mathcal{G} = (G, \cdot, \mathbf{1})$ справедливы **левый** и **правый законы сокращения**:
если $a \cdot x = a \cdot y$, то $x = y$,
если $x \cdot a = y \cdot a$, то $x = y$.

◀ Пусть $a \cdot x = a \cdot y$.
Умножим обе части этого равенства слева на элемент a^{-1}.
Получим

$a^{-1} \cdot (a \cdot x) = a^{-1} \cdot (a \cdot y)$
в силу ассоциативности $(a^{-1} \cdot a) \cdot x = (a^{-1} \cdot a) \cdot y.$
поскольку $a^{-1} \cdot a = \mathbf{1} \;\Rightarrow\; \mathbf{1} \cdot x = \mathbf{1} \cdot y \;\Rightarrow\; x = y$

Доказан левый закон сокращения. Аналогично доказывается и правый закон. ▶

8.3. Решение уравнений в группе

Пусть $\mathcal{G} = (G, \cdot, \mathbf{1})$ — группа, a, b — фиксированные элементы G.
Рассмотрим задачу решения уравнений

$$a \cdot x = b, \tag{8.1}$$

$$x \cdot a = b \tag{8.2}$$

в группе $\mathcal{G}$.
т.е. поиска всех таких элементов $x \in G$, для которых уравнение (8.1) (или (8.2)) превращается в тождество.

Теорема 4. В любой группе $\mathcal{G}$ уравнения вида $a \cdot x = b$ (8.1)
и $x \cdot a = b$ (8.2)
имеют решения, и притом единственные.

◀ Покажем, что $x = a^{-1} \cdot b$ есть решение (8.1).

$$a \cdot (a^{-1} \cdot b) = (a \cdot a^{-1} \cdot b) = b.$$

Докажем единственность решения.
Пусть для фиксированных a и b и некоторого x выполнено равенство $a \cdot x = b$.
В группе для любого a существует и однозначно определен элемент a^{-1}, обратный к a.

Умножим на a^{-1} обе части равенства и преобразуем, используя ассоциативность операции в группе.

$$\begin{gathered} a^{-1} \cdot (a \cdot x) = a^{-1} \cdot b \ \Rightarrow \\ \Rightarrow \ (a^{-1} \cdot a) \cdot x = a^{-1} \cdot b \ \Rightarrow \\ \Rightarrow \ \mathbf{1} \cdot x = a^{-1} \cdot b \ \Rightarrow \\ \Rightarrow x = a^{-1} \cdot b. \end{gathered}$$

Это решение единственное в силу единственности обратного элемента.

Аналогично из $x \cdot a = b$ получаем $x = b \cdot a^{-1}$, и это решение также единственное. ▶

ДОПОЛНИТЕЛЬНЫЙ МАТЕРИАЛ.

Нуль относительно операции

Элемент $\mathbf{0}$ множества A называют **левым** (**правым**) **нулем** относительно данной операции $*$, если $\mathbf{0} * x = \mathbf{0}$ ($x * \mathbf{0} = \mathbf{0}$) для любого $x \in A$.

Если $\mathbf{0}'$ — левый нуль, а $\mathbf{0}''$ — правый нуль, то они совпадают.

Если $\mathbf{0}'$ и $\mathbf{0}''$ существуют, то они совпадают, так как $\mathbf{0}' = \mathbf{0}' * \mathbf{0}'' = \mathbf{0}''$, и в этом случае говорят просто о **нуле относительно операции**.

Нуль единственный и для него одновременно выполнены оба равенства $\mathbf{0} * x = \mathbf{0}$ и $x * \mathbf{0} = \mathbf{0}$.

Пример 8.4. **а.** На множестве целых чисел нулем относительно операции умножения будет число 0.

б. На множестве квадратных матриц вида $\begin{pmatrix} a & 0 \\ b & 1 \end{pmatrix}$, где элементы a и b — действительные числа, любая матрица вида $\begin{pmatrix} 0 & 0 \\ d & 1 \end{pmatrix}$ будет правым нулем относительно операции умножения.

$$\begin{pmatrix} a & 0 \\ b & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ d & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ d & 1 \end{pmatrix}.$$

Левого нуля в этом множестве нет .

Правых нулей имеется больше одного. #

Примеры алгебр.

Пример 8.5. Алгебра $(\mathbb{Z}^+, +)$, где носитель — множество $\mathbb{Z}$ неотрицательных целых чисел, а сигнатура состоит из одной операции сложения, есть коммутативный моноид, в котором нейтральный элемент — это число 0.
Сумма двух неотрицательных целых чисел есть целое неотрицательное число, операция сложения ассоциативна, коммутативна и для любого целого числа n имеет место равенство $n + 0 = n$.

Пример 8.6. Алгебра $(\mathbb{Z}, \cdot)$, у которой носителем является множество целых чисел, а сигнатура состоит из одной операции умножения, есть коммутативный моноид. Нейтральным элементом этого моноида является число 1.

Пример 8.7. Пусть A — конечное множество, а A^n — множество кортежей длины n. На множестве всех кортежей $A^+ = \bigcup\limits_{n\geq 1} A^n$ определим операцию **соединения (конкатенации) кортежей** следующим образом:

$$(a_1, \ldots, a_m) \cdot (b_1, \ldots, b_k) = (a_1, \ldots, a_m, b_1, \ldots, b_k).$$

Введенная операция ассоциативна, но не имеет нейтрального элемента. Таким образом, построена полугруппа, но не моноид.

Чтобы превратить эту полугруппу в моноид, расширим носитель полугруппы, введя понятие **нулевой декартовой степени** A^0 произвольного множества A.
Под A^0 понимают одноэлементное множество $\{\lambda\}$, единственный элемент которого называют **пустым кортежем**
Обозначив $A^* = A^0 \cup A^+$, по определению для любого $x \in A^*$ полагаем $x \cdot \lambda = \lambda \cdot x = x$.

В результате получим алгебру $(A^*, \cdot)$.
Это моноид, с нейтральным элементом λ.
Этот моноид называют **свободным моноидом**, порожденным множеством A.

Алгебры однотипные

Для алгебры $\mathcal{A} = (A, \Omega)$ обозначим через $\Omega^{(n)}$ подмножество сигнатур Ω, состоящее из всех n-арных операций. Тогда $\Omega = \bigcup\limits_{n \geq 0} \Omega^{(n)}$.

Рассмотрим алгебру

$$\mathcal{A}_1 = \left(2^M, \left\{\cup, \cap, \setminus, \Delta, \overline{\ }, \varnothing, M\right\}\right).$$

Носителем является множество всех подмножеств произвольно фиксированного множества M. Сигнатура состоит из следующих операций над множествами: объединения, пересечения, разности, симметрической разности, дополнения, пустого множества и множества M. Пустое множество и множество M определяют нульарные операции.

Имеем:

$$\Omega^{(0)} = \{\varnothing,\ M\},$$
$$\Omega^{(1)} = \left\{\overline{\ }\right\},$$
$$\Omega^{(2)} = \{\cup,\ \cap,\ \setminus,\ \Delta\},$$
$$\Omega^{(n)} = \varnothing \text{ при } n > 2.$$

Определение 8.5. Две алгебры $\mathcal{A}_1 = (A_1,\ \Omega_1)$ и $\mathcal{A}_2 = (A_2,\ \Omega_2)$ называют **однотипными**, если существует такая *биекция* Ω_1 на Ω_2, при которой n-арная операция из Ω_1 для любого n переходит в n-арную из Ω_2.

Нередко сигнатуры однотипных алгебр и элементы этих сигнатур — операции — обозначают одинаково. Так, мы пишем $(\mathbb{R},\ +,\ \cdot,\ 0,\ 1)$ и $(\mathbb{Q},\ +,\ \cdot,\ 0,\ 1)$, хотя первая алгебра задана на множестве всех действительных чисел, а вторая — на множестве рациональных чисел, и, например, сложение в первой алгебре, строго говоря, не есть та же самая операция, что сложение во второй алгебре. В общем случае мы часто будем говорить о различных (но однотипных) Ω-алгебрах, заданных на разных носителях, понимая, что Ω есть общее для всех этих алгебр обозначение их сигнатур.

Пример 8.8. Алгебра $(2^M,\ \cup,\ \cap,\ \varnothing,\ M)$, заданная на множестве всех подмножеств множества M, и алгебра $\mathcal{A}_3 = (\mathbb{R},\ +,\ \cdot,\ 0,\ 1)$, заданная на множестве действительных чисел, однотипны.

Биекцию (*взаимно однозначное соответствие*) между их сигнатурами, которая сохраняла бы арность операций, можно определить и так:

$\cup \mapsto +$, $\cap \mapsto \cdot$, $\varnothing \mapsto 0$, $M \mapsto 1$.

Указанный способ задания биекции не единственный. Например, ее можно определить так:

$\cup \mapsto \cdot$, $\cap \mapsto +$, $\varnothing \mapsto 1$, $M \mapsto 0$.

Не являются однотипными и алгебры $\left(2^M, \overline{\ }\right)$ и $(\mathbb{N}, +)$, ибо в первой алгебре единственная операция ее сигнатуры является унарной, а во второй — бинарной. #

Ткачев С.Б.

каф. Математического моделирования
МГТУ им. Н.Э. Баумана

ДИСКРЕТНАЯ МАТЕМАТИКА

ИУ5 — 4 семестр, 2015 г.

Лекция 9. ГРУППЫ, КОЛЬЦА, ПОЛЯ

Существует две формы записи бинарной операции группы.
В **аддитивной записи** операцию обозначают знаком $+$, нейтральный элемент — знаком $\mathbf{0}$, элемент, обратный к a относительно операции $+$, записывают в виде $-a$ и называют **противоположным** к a.
Бинарную операцию группы в этом случае называют **сложением**

В **мультипликативной записи** операцию обозначают знаком $\cdot$, нейтральный элемент — знаком $\mathbf{1}$, элемент, обратный к a, записывают в виде a^{-1}.
В этом случае бинарную операцию группы называют **умножением** (также **умножением группы** или **групповым умножением**), элемент $a \cdot b$ —**произведением** элементов a и b, и записывают в виде ab.

Пример 9.1. Алгебра $(\mathbb{Z}, +)$ — коммутативная группа.
На множестве целых чисел операция сложения ассоциативна и коммутативна.
Число 0 есть нейтральный элемент.
Для каждого целого числа n существует обратный по сложению элемент, число $-n$, противоположное n.
Рассматриваемую группу называют **аддитивной группой целых чисел**.

Пример 9.2. Множество всех *биекций* некоторого множества A на себя с операцией композиции отображений есть группа.
Композиция двух биекций есть биекция.
Операция композиции ассоциативна.
Нейтральный элемент — тождественное отображение id_A — есть биекция.
Для всякой биекции $f\colon A \to A$ отображение f^{-1}, обратное биекции f, определено, является биекцией и выполнены равенства $f \circ f^{-1} = f^{-1} \circ f = \mathrm{id}_A$.

Эту группу называют **симметрической группой множества** A.

Если множество A конечно, — **группой подстановок** множества A.

Пример 9.3.
Алгебры $(\mathbb{Q} \setminus \{0\}, \cdot)$ и $(\mathbb{R} \setminus \{0\}, \cdot)$ есть коммутативные группы.
Их называют
мультипликативной группой рациональных чисел
и **мультипликативной группой действительных чисел** соответственно.

В каждой из них число 1 есть нейтральный элемент (единица) группы.
Обратный к числу x по операции умножения элемент x^{-1} есть число $x^{-1} = 1/x$.

Пример 9.4.
Рассмотрим алгебру $\mathbb{Z}_k^+ = (\{0, 1, \ldots, k-1\}, \oplus_k)$. Операция $\oplus_k$ (**сложения по модулю k**) определяется следующим образом:
для любых двух m и n число $m \oplus_k n$, называемое **суммой** чисел m и n **по модулю** k, равно остатку от деления арифметической суммы $m + n$ на k.

Эта алгебра является коммутативной группой. Ее называют **аддитивной группой вычетов по модулю** k.

Нейтральным элементом служит число 0.

Обратным к числу n будет $k - n$, т.к. $n \oplus_k (k - n) = 0$.

Пример 9.5. Множество всех невырожденных (т.е. имеющих ненулевой определитель) числовых квадратных матриц порядка n с операцией умножения матриц является группой.

Произведение двух невырожденных матриц снова есть невырожденная матрица.

Единичная матрица порядка n невырожденная.

Матрица, обратная к невырожденной, также является невырожденной.

При использовании аддитивной записи операции для коммутативной группы $\mathcal{G} = (G, +, \mathbf{0})$ уравнения $a + x = b$, $x + a = b$ сводятся к одному:

$$a + x = b,$$

Решение уравнения есть $x = b + (-a)$.

Правую часть этого равенства в коммутативной группе называют **разностью** элементов b и a и обозначают $b - a$.

Операцию, сопоставляющую упорядоченной паре (a, b) разность $b - a$, называют операцией **вычитания**.

С учетом введенных обозначений решение уравнения в коммутативной группе можно записать так: $x = b - a$.

9.1. Группа подстановок

Рассмотрим взаимнооднозначное отображение n -элементного множества $\{1, 2, \ldots, n\}$ в себя (биекцию). Такую биекцию называют *подстановкой* этого множества.

$$\sigma = \begin{pmatrix} 1 & 2 & \ldots & n \\ \alpha_1 & \alpha_2 & \ldots & \alpha_n \end{pmatrix}.$$

Образ 1 (при отображении σ) есть α_1 , образ 2 есть α_2 , $\ldots$, образ n есть α_n .

Циклом длины k называют подстановку, которая отображает β_1 в β_2, β_2 в β_3, $\ldots$, β_{k-1} в β_k, а β_k в β_1, где $\beta_i \in \{1, \ldots, n\}$, $i = 1, \ldots, k$ и все β_i попарно различны, а все элементы, отличные от β_1, $\ldots$, β_k, отображаются сами в себя.
Цикл записывают ее в виде $(\beta_1\ \beta_2\ \ldots\ \beta_k)$.

Например, подстановку из группы S_4

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 4 & 1 \end{pmatrix}$$

можно записать в виде $(1\ 3\ 4)$.

Цикл длины 2 называют *транспозицией*.

Обратная подстановка

Подстановка, обратная подстановке

$$\begin{pmatrix} 1 & 2 & \dots & n \\ \alpha_1 & \alpha_2 & \dots & \alpha_n \end{pmatrix},$$

есть подстановка, которая отображает α_1 в 1, α_2 в 2, $\dots$ α_n в n. Отметим, что при записи обратной подстановки элементы первой строки тем не менее записываются в обычном порядке: $1, \dots, n$.

Решение уравнений в группе подстановок

В группе S_3 решим следующее уравнение:

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} \circ X = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}.$$

Умножив обе части уравнения слева на

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix},$$

получим

$$X = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}.$$

Окончательно получим

$$X = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}.$$

9.2. Кольца, тела, поля

Определение 9.1. Кольцом называют **алгебру**

$$\mathcal{R} = (R,\, +,\, \cdot,\, \mathbf{0},\, \mathbf{1}),$$

сигнатура которой состоит из двух бинарных и двух нульарных операций, причем для любых $a,\, b,\, c \in R$ выполняются равенства:

1) $a + (b + c) = (a + b) + c$;
2) $a + b = b + a$;
3) $a + \mathbf{0} = a$;
4) для каждого $a \in R$ существует элемент a', такой, что $a + a' = \mathbf{0}$;
5) $a \cdot (b \cdot c) = (a \cdot b) \cdot c$;
6) $a \cdot \mathbf{1} = \mathbf{1} \cdot a = a$;
7) $a \cdot (b + c) = a \cdot b + a \cdot c$, $(b + c) \cdot a = b \cdot a + c \cdot a$.

Операцию $+$ называют **сложением кольца**.
Операцию $\cdot$ — **умножением кольца**.
Элемент $\mathbf{0}$ — **нулем кольца**.
элемент $\mathbf{1}$ — **единицей кольца**.

Равенства 1–7, указанные в определении, называют **аксиомами кольца**.

Аксиомы кольца 1–4 означают, что алгебра $(R, +, \mathbf{0})$, сигнатура которой состоит только из операций сложения кольца $+$ и нуля кольца $\mathbf{0}$, является **абелевой группой**.
Эту группу называют **аддитивной группой кольца** $\mathcal{R}$

Аксиомы кольца 5 и 6 показывают, что алгебра $(R, \cdot, \mathbf{1})$, сигнатура которой включает только умножение кольца $\cdot$ и единицу кольца $\mathbf{1}$, есть моноид.
Этот моноид называют **мультипликативным моноидом кольца** $\mathcal{R}$

Аксиома 7 устанавливает связь между сложением кольца и умножением кольца. Операция умножения дистрибутивна относительно операции сложения.

Кольцо — это алгебра с двумя бинарными и двумя нульарными операциями $\mathcal{R} = (R,\ +,\ \cdot,\ \mathbf{0},\ \mathbf{1})$, такая, что:
1) алгебра $(R,\ +,\ \mathbf{0})$ — коммутативная группа;
2) алгебра $(R,\ \cdot,\ \mathbf{1})$ — моноид;
3) операция $\cdot$ (умножения кольца) дистрибутивна относительно операции $+$ (сложения кольца).

Определение 9.2. Кольцо называют **коммутативным**, если его операция умножения коммутативна.

Пример 9.6.

а. Алгебра $(\mathbb{Z},\, +,\, \cdot,\, 0,\, 1)$ есть коммутативное кольцо. Отметим, что алгебра $(\mathbb{N},\, +,\, \cdot)$ кольцом не будет, поскольку $(\mathbb{N},\, +)$ — коммутативная полугруппа, но не группа.

б. Рассмотрим алгебру

$$\mathbb{Z}_k = (\{0,\, 1,\, \ldots,\, k-1\},\, \oplus_k,\, \odot_k,\, 0, 1)$$

($k \geq 1$) с операцией $\oplus_k$ сложения по модулю k и $\odot_k$ (умножения по модулю k).

Операция умножения по модулю k аналогична операции сложения по модулю k : $m \odot_k n$ равно остатку от деления на k числа $m \cdot n$.

Эта алгебра есть коммутативное кольцо, которое называют **кольцом вычетов по модулю** k .

Пример 9.7. **а.** Алгебра $(2^A, \triangle, \cap, \varnothing, A)$ — коммутативное кольцо. Это следует из свойств *пересечения* и *симметрической разности множеств*.

б. Множество всех квадратных матриц фиксированного порядка с операциями сложения и умножения матриц — некоммутативное кольцо.
Единицей этого кольца является единичная матрица, а нулем — нулевая.

Теорема 1. В любом кольце выполняются следующие тождества:

1 $\mathbf{0} \cdot a = a \cdot \mathbf{0} = \mathbf{0}$;

2 $(-a) \cdot b = -(a \cdot b) = a \cdot (-b)$;

3 $(a - b) \cdot c = a \cdot c - b \cdot c$, $\quad c \cdot (a - b) = c \cdot a - c \cdot b$.

◀ Докажем тождество $\mathbf{0} \cdot a = \mathbf{0} \quad (1)$.

$$\forall\, a\, (a + \mathbf{0} \cdot a = \mathbf{1} \cdot a + \mathbf{0} \cdot a = (\mathbf{1} + \mathbf{0}) \cdot a = \mathbf{1} \cdot a = a).$$

В аддитивной группе кольца получили уравнение

$$a + \mathbf{0} \cdot a = a$$

относительно неизвестного элемента $\mathbf{0} \cdot a$.
В аддитивной группе любое уравнение вида $a + x = b$ имеет единственное решение $x = b - a$.
$\mathbf{0} \cdot a = a - a = \mathbf{0}$.

Тождество $a \cdot \mathbf{0} = \mathbf{0}$ доказывается аналогично.

Докажем тождество $-(a \cdot b) = a \cdot (-b) \quad (2)$. Имеем

$$a \cdot (-b) + a \cdot b = a \cdot ((-b) + b) = a \cdot \mathbf{0} = \mathbf{0} \Rightarrow$$
$$\Rightarrow \; a \cdot (-b) = -(a \cdot b)$$

$(-a) \cdot b = -(a \cdot b)$ можно доказать точно так же.

Докажем тождества (3).
Рассмотрим $(a - b) \cdot c = a \cdot c - b \cdot c$.
С учетом доказанного выше имеем

$$a \cdot (b - c) = a \cdot (b + (-c)) = a \cdot b + a \cdot (-c) = a \cdot b - a \cdot c,$$

т.е. тождество справедливо.
Тождество $c \cdot (a - b) = c \cdot a - c \cdot b$ доказывается аналогично. ►

Следствие 9.1. В любом кольце справедливо тождество

$$(-\mathbf{1}) \cdot x = x \cdot (-\mathbf{1}) = -x.$$

◄ Указанное следствие вытекает из второго тождества теоремы 1 при $a = \mathbf{1}$ и $b = x$. ►

Первые два тождества в теореме выражают свойство, называемое **аннулирующим свойством нуля** в кольце.

Тождества (3) теоремы 1 выражает свойство дистрибутивности операции умножения кольца относительно операции вычитания.

В любом кольце производя вычисления , можно раскрывать скобки и менять знаки так же, как и при сложении, вычитании и умножении действительных чисел.

Определение 9.3. Ненулевые элементы a и b кольца $\mathcal{R}$ называют **делителями нуля**, если $a \cdot b = \mathbf{0}$ или $b \cdot a = \mathbf{0}$.

Пример 9.8. Кольцо вычетов по модулю k, если k — составное число.
В этом случае произведение по модулю k любых m и n, дающих при обычном перемножении число, кратное k, будет равно нулю.
В кольце вычетов по модулю 6 элементы 2 и 3 являются делителями нуля, поскольку $2 \odot_6 3 = 0$.

Пример 9.9. Кольцо квадратных матриц фиксированного порядка (не меньшего двух).
Например, для матриц второго порядка имеем

$$\begin{pmatrix} 0 & a \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

При отличных от нуля a и b приведенные матрицы являются делителями нуля.

Определение 9.4. Кольцо, в котором множество всех ненулевых элементов по умножению образует группу, называют **телом**.

Определение 9.5. Коммутативное тело называют **полем**, а группу ненулевых элементов тела (поля) по умножению —**мультипликативной группой** этого **тела**(**поля**).

Поле есть частный случай кольца, в котором операции обладают дополнительными свойствами.

Аксиомы поля

Поле есть алгебра $\mathcal{F} = (F, +, \cdot, \mathbf{0}, \mathbf{1})$, сигнатура которой состоит из двух бинарных и двух нульарных операций, причем справедливы тождества:

1) $a + (b + c) = (a + b) + c$;
2) $a + b = b + a$;
3) $a + \mathbf{0} = a$;
4) для каждого $a \in F$ существует элемент $-a$, такой, что $a + (-a) = \mathbf{0}$;
5) $a \cdot (b \cdot c) = (a \cdot b) \cdot c$;
6) $a \cdot b = b \cdot a$;
7) $a \cdot \mathbf{1} = \mathbf{1} \cdot a = a$;
8) для каждого $a \in F$, отличного от $\mathbf{0}$, существует элемент a^{-1}, такой, что $a \cdot a^{-1} = \mathbf{1}$;
9) $a \cdot (b + c) = a \cdot b + a \cdot c$.

Пример 9.10.
а. Алгебра $(\mathbb{Q},\ ,+,\ \cdot,\ 0,\ 1)$ есть поле, называемое **полем рациональных чисел**.

б. Алгебры $(\mathbb{R},\ +,\ \cdot,\ 0,\ 1)$ и $(\mathbb{C},\ +,\ \cdot,\ 0,\ 1)$ есть поля, называемые **полями действительных** и **комплексных чисел** соответственно.

9.3. Области целостности

Областью целостности называют **коммутативное кольцо** без **делителей нуля**.

Так, кольцо целых чисел есть область целостности.

Утверждение 9.1. Если A — конечное множество и $f : A \to A$ — инъекция, то она является сюръекцией и следовательно биекцией

Теорема 2. Конечная область целостности является **полем**.

◄ Поле — это кольцо, умножение которого **коммутативно**, каждый ненулевой элемент a имеет **обратный элемент** относительно **умножения**.

Область целостности является **коммутативным** кольцом без делителей нуля.

Докажем, что для конечной области целостности любой ненулевой элемент обратим, т.е.

$$\forall\ (a \neq \mathbf{0}) \exists x\ (\text{единственный}) \mid a \cdot x = \mathbf{1}.$$

Фиксируем произвольный элемент $a \neq \mathbf{0}$.
Определим отображение f_a множества всех ненулевых элементов в себя по формуле $f_a(x) = a \cdot x$
($a \cdot x \neq \mathbf{0}$ в области целостности при $a \neq \mathbf{0}$ и $x \neq \mathbf{0}$).
Докажем, что отображение f_a — инъекция (каждый элемент из области значений имеет единственный прообраз).

$$a \cdot x = a \cdot y \;\Rightarrow\; a \cdot (x - y) = \mathbf{0} \Rightarrow$$
$$\Rightarrow x - y = \mathbf{0} \text{ (т.к. делители нуля отсутствуют)} \Rightarrow\; x = y$$

Множество носитель по условию теоремы конечено, следовательно, f_a — биекция (утверждение 9.1).
Поэтому $\forall(y)\ \exists(\text{ единственный } x) |\ y = a \cdot x$.
В частности, при $y = \mathbf{1}$ равенство $a \cdot x = \mathbf{1}$ выполнено для некоторого однозначно определенного x, т.е. $x = a^{-1}$. ▶

Следствия теоремы 2.

Следствие 9.2. Кольцо $\mathbb{Z}_p$ вычетов по модулю p является полем тогда и только тогда, когда p — простое число.

◄ Пусть $\mathbb{Z}_p$ является полем. Покажем, что p — простое число.
Предположим — p составное.
Тогда найдутся такие k и l, $0 < k \leq p-1; 0 < l \leq p-1$, что $p = k \cdot l \Rightarrow$
$k \cdot l = 0 \,(\mathrm{mod\ p}) \Rightarrow$ k и l — делители нуля в кольце $\mathbb{Z}_p$.
Следовательно, $\mathbb{Z}_p$ — не поле.
Число p не может быть составным.

Пусть p — простое число.

Предположим, что $m \cdot n = 0 \,(\mathrm{mod\ p})$, т.е. элементы m и n кольца $\mathbb{Z}_p$ будут делителями нуля (кольцо не область целостности).

p — простое число и $(m \cdot n = 0 \,(\mathrm{mod\ p})) \Rightarrow$
$(\ (m = 0 \,(\mathrm{mod\ p})) \vee (n = 0 \,(\mathrm{mod\ p}))\ )$

Т.к. $((0 \leq m \leq p-1) \wedge (0 \leq n \leq p-1)) \ \Rightarrow (m = 0) \vee (n = 0)$. Следовательно, при простом p делителей нуля нет.

Кольцо $\mathbb{Z}_p$ является конечной областью целостности и по теореме 2 — полем. ►

Материал для самостоятельного изучения

9.4. Циклическая полугруппа

В свободном моноиде, порожденном некоторым конечным множеством, оба закона сокращения справедливы, но никаких обратных элементов не существует.

В полугруппе можно умножать любой элемент a сам на себя, причем в силу ассоциативности операции полугруппы элемент $\underbrace{a \cdot a \cdot \ldots \cdot a}_{n \text{ раз}}$ определен однозначно.

Этот элемент называют n-й **степенью элемента** a и обозначают a^n.

При этом $a^1 = a$, $a^n = a \cdot a^{n-1}$, $n = 2, 3, \ldots$

В моноиде вводят также нулевую степень элемента, полагая $a^0 = \mathbf{1}$.

Если $(A, \cdot, \mathbf{1})$ — группа, то можно ввести и отрицательные степени элемента согласно равенству $a^{-n} = (a^{-1})^n$, $n = 1, 2, \ldots$

Сформулируем утверждения о свойствах степеней (без доказательства).

Утверждение 9.2. Для любой полугруппы $a^m \cdot a^n = a^{m+n}$, $(a^m)^n = a^{mn}$ (m, $n \in \mathbb{N}$).

Утверждение 9.3. Для любой группы $a^{-n} = (a^n)^{-1}$ $(n \in \mathbb{N})$, $a^m \cdot a^n = a^{m+n}$, $(a^m)^n = a^{mn}$ (m, $n \in \mathbb{Z}$).

Определение 9.6. Полугруппу (в частности, группу) $(A, \cdot)$ называют **циклической**, если существует такой элемент a, что любой элемент x полугруппы является некоторой (целой) степенью элемента a.
Элемент a называют **образующим элементом полугруппы (группы)**.

Пример 9.11. а. Полугруппа $(\mathbb{N}, +)$ циклическая, с образующим элементом 1. При аддитивной записи бинарной операции возведение элемента a в положительную степень n есть сумма n этих элементов, и это записывают $n \cdot a$ (или na, без знака умножения).

б. Группа $(\mathbb{Z}, +, 0)$ также циклическая. Для нее образующими элементами могут быть 1 и -1.
Рассмотрим элемент 1. Тогда $0 \cdot 1 = 0$, $n \cdot 1 = \underbrace{1 + \ldots + 1}_{n \text{ раз}} = n$ ($n > 0$) и $(-1) \cdot 1 = -1$, $(-n) \cdot 1 = n \cdot (-1) = \underbrace{(-1) + \ldots + (-1)}_{n \text{ раз}} = -n$ ($n > 0$).
Если в качестве образующего взять элемент -1, то $0 \cdot (-1) = 0$, отрицательные целые числа получаются как положительные „степени“ -1, а положительные — как отрицательные „степени“ -1. Например, $(-2) \cdot (-1) = 2$, $4 \cdot (-1) = -4$.

в. Группа $(\mathbb{Z}_3,\, \oplus_3,\, 0)$ вычетов по модулю 3 циклическая, причем любой ее ненулевой элемент является образующим. Действительно, для 1 имеем $1 \oplus_3 1 = 2$, $1 \oplus_3 1 \oplus_3 1 = 0$, а для 2 получим $2^2 = 2 \oplus_3 2 = 1$, $2 \oplus_3 2 \oplus_3 2 = 0$. #

Рассмотрим подробнее строение конечных циклических групп, используя мультипликативную запись бинарной операции.

Вспомним, *конечная алгебра* (**конечная группа**, в частности) — это алгебра, носитель которой — конечное множество.

Порядком конечной группы называют количество элементов в этой группе.

Например, аддитивная группа вычетов по модулю k имеет порядок k.

Симметрическая группа степени n, т.е. группа подстановок S_n, имеет порядок $n!$.

Мультипликативная группа вычетов по модулю p, где p — простое число, имеет порядок $p-1$.

Порядок элемента a циклической группы — это наименьшее положительное n, такое, что $a^n = \mathbf{1}$.

Теорема 3. Порядок образующего элемента конечной циклической группы равен порядку самой группы.

◀ Пусть $\mathcal{G} = (G, \cdot, \mathbf{1})$ — конечная циклическая группа с образующим элементом a и $n > 0$ — порядок этого элемента.
Тогда все степени $a^0 = \mathbf{1}$, $a^1 = a$, $\ldots$, a^{n-1} попарно различны.
Действительно, если $a^k = a^l$, $0 < l < k < n$, то $a^{k-l} = a^{k+(-l)} = a^k a^{-l} = a^l a^{-l} = a^{l-l} = \mathbf{1}$.
Получено противоречие с выбором n как порядка элемента a, поскольку $k - l < n$ (найдена степень, меньшая n, при возведении в которую элемента a получится единица).

Докажем, что любая степень элемента a принадлежит множеству $\{\mathbf{1},\, a,\, \ldots,\, a^{n-1}\}$.

$\forall(m \in \mathbb{Z})\ \exists(n, k \in \mathbb{Z})|\ (m = kn+q)$, где $(q \in \mathbb{Z} \wedge 0 \leq q < n$

Тогда

$a^m = a^{kn+q} = a^{kn} \cdot a^q = (a^n)^k \cdot a^q = \mathbf{1} \cdot a^q = a^q \in \{\mathbf{1},\, a,\, \ldots,\, a^n$

Поскольку каждый элемент группы $\mathcal{G}$ есть некоторая степень элемента a, то $G = \{\mathbf{1},\, a,\, \ldots,\, a^{n-1}\}$ и порядок группы равен n. ►

Из доказанной теоремы следует, что в бесконечной циклической группе не существует такого $n > 0$, что для образующего элемента a группы выполняется равенство $a^n = \mathbf{1}$.

9.5. Подгруппы.

Пусть $\mathcal{G} = (G, *)$ — произвольный группоид и $H \subseteq G$ — некоторое подмножество множества G.
Рассмотрим свойства бинарной операции $*$ группоида $\mathcal{G}$ на подмножестве H.

Определение 9.7. Множество $H \subseteq G$ замкнуто относительно операции $*$, если $x * y \in H$ для любых x, $y \in H$.

В этом случае подмножество H с операцией $*$ будет группоидом $\mathcal{H} = (H, *)$. Его называют **подгруппоидом** группоида $\mathcal{G}$.

Если подмножество H замкнуто относительно бинарной операции $*$ и эта *бинарная операция ассоциативна* на множестве G, то операция останется ассоциативной и при ее ограничении на подмножество H.

Если группоид $\mathcal{G}$ является полугруппой, то и всякий его подгруппоид будет полугруппой, называемой подполугруппой полугруппы $\mathcal{G}$.

Однако в случае, когда группоид является *моноидом* (*группой*), уже нельзя утверждать, что любой подгруппоид является также моноидом (группой).

Пример 9.12. Рассмотрим группоид —**аддитивную группу целых чисел** $(\mathbb{Z}, +)$.
Выделим в множестве целых чисел подмножество $\mathbb{N}$ натуральных чисел. Это подмножество замкнуто относительно операциии сложения $+$, группоид $(\mathbb{N}, +)$ будет подгруппоидом группоида $(\mathbb{Z}, +)$.
Так как операция сложения чисел ассоциативна, $(\mathbb{N}, +)$ будет подполугруппой. Однако в множестве $\mathbb{N}$ отсутствует **нейтральный элемент** 0 относительно операции сложения. Следовательно, $(\mathbb{N}, +)$ не является группой (не является даже моноидом).

Пусть $\mathcal{M} = (M, \cdot, \mathbf{1})$ — моноид.
Если P есть подмножество M, замкнутое относительно бинарной операции $\cdot$ моноида $\mathcal{M}$ и содержащее нейтральный элемент (**единицу**) $\mathbf{1}$ этого моноида, то $\mathcal{P} = (P, \cdot, \mathbf{1})$ также есть моноид.
Его называют **подмоноидом** моноида $\mathcal{M}$.
Замкнутость подмножества $B \subseteq A$ относительно нульарной операции a на A равносильна соотношению $a \in B$.

Определение 9.8. Моноид $\mathcal{P} = (P, \cdot, \mathbf{1})$ есть подмоноид моноида $\mathcal{M} = (M, \cdot, \mathbf{1})$ тогда и только тогда, когда множество P замкнуто относительно бинарной операции $\cdot$ моноида $\mathcal{M}$, а также относительно его нульарной операции $\mathbf{1}$.

Определение 9.9. Пусть $\mathcal{G} = (G, \cdot, {}^{-1}, \mathbf{1})$ — группа, а H есть подмножество G, замкнутое относительно операции $\cdot$ группы $\mathcal{G}$, содержащее нейтральный элемент (**единицу**) $\mathbf{1}$ этой группы и вместе с каждым элементом $x \in H$ содержащее **элемент** x^{-1}, **обратный** к x, т.е. замкнутое относительно унарной операции ${}^{-1}$ взятия обратного, которая здесь включена в сигнатуру группы.
Тогда $\mathcal{H} = (H, \cdot, {}^{-1}, \mathbf{1})$ также есть группа, которую называют **подгруппой** группы $\mathcal{G}$.

Пусть ω — унарная операция на множестве G моноида $\mathcal{G}$, $\mathcal{H}$ — некоторый его подмоноид.

Подмоноид $\mathcal{H}$ моноида $\mathcal{G}$ называется замкнутым относительно унарной операции ω, если для каждого $x \in H$ имеет место $\omega(x) \in H$.

Группа $\mathcal{H} = (H, \cdot, {}^{-1}, \mathbf{1})$ есть подгруппа группы $\mathcal{G} = (G, \cdot, {}^{-1}, \mathbf{1})$ в том и только в том случае, когда множество H замкнуто относительно всех операций $\cdot$, ${}^{-1}$, $\mathbf{1}$ сигнатуры группы $\mathcal{G}$.

Единица моноида (группы) служит одновременно единицей любого его подмоноида (любой подгруппы).

Пример 9.13.

а. Подмножество всех натуральных четных чисел есть подполугруппа полугруппы $(\mathbb{N}, +)$ (подмножество всех натуральных четных чисел замкнуто относительно сложения, операция сложения ассоциативна).

б. Аддитивная полугруппа натуральных чисел с нулем $(\mathbb{N} \cup \{0\}, +)$ — моноид с нейтральным элементом 0.

Подмножество всех положительных (> 0) четных чисел с операцией сложения не будет подмоноидом моноида $(\mathbb{N} \cup \{0\}, +, 0)$, так как ее носитель не содержит нуля — единицы моноида.

Подмножество всех натуральных чисел вместе с нулем, делящихся на заданное число $k > 1$, замкнуто относительно операции сложения; на нем может быть определен подмоноид моноида $(\mathbb{N} \cup \{0\}, +, 0)$.

в. Группа рациональных чисел $\mathbb{Q}$ с операцией умножения, является подгруппой группы действительных чисел с операцией умножения $(\mathbb{R} \setminus \{0\},\ \cdot,\ 1)$.
г. Алгебра $(\mathbb{Z} \setminus \{0\},\ \cdot,\ 1)$ не является подгруппой группы $(\mathbb{R} \setminus \{0\},\ \cdot,\ 1)$, т.к. не содержит вместе с каждым целым числом m обратного к нему числа $\frac{1}{m}$.
Данное множество будет моноидом т.к. оно замкнуто относительно операции умножения и содержит единицу.

9.6. Циклические подгруппы

Пусть $\mathcal{G} = (G, \cdot, {}^{-1}, \mathbf{1})$ — группа.
Произведение любых **степеней элемента** a есть снова некоторая степень элемента a, нулевая степень дает единицу группы, а обратным к элементу a^k является элемент a^{-k}. Таким образом, множество всех степеней фиксированного элемента a группы $\mathcal{G}$ является подгруппой группы $\mathcal{G}$.

Определение 9.10. Подгруппу группы $\mathcal{G}$, заданную на множестве всех степеней фиксированного элемента a, называют **циклической подгруппой** группы $\mathcal{G}$, **порожденной элементом** a.

Пример 9.14. В группе $\mathbb{Z}_{13}^*$ (мультипликативной группе вычетов по модулю 13) построим циклическую подгруппу, порожденную элементом 5.

Имеем: $5^0 = 1$, $5^1 = 5$, $5^2 = 5 \odot_{13} 5 = 12$, $5^3 = 5 \odot_{13} 12 = 8$, $5^4 = 5 \odot_{13} 8 = 1$.

Порядок этой циклической подгруппы равен 4.

Она состоит из элементов: 1, 5, 8 и 12.

9.7. Теорема Лагранжа

Пусть $\mathcal{G} = (G, \cdot, \mathbf{1})$ — группа, а $\mathcal{H} = (H, \cdot, \mathbf{1})$ — ее подгруппа.

Левым смежным классом подгруппы $\mathcal{H}$ по элементу $a \in G$ называют множество

$$aH = \{y\colon y = a \cdot h,\ h \in H\}\,.$$

Соответственно **правый смежный класс подгруппы $\mathcal{H}$ по элементу** $a \in G$ — это множество

$$Ha = \{y\colon y = h \cdot a,\ h \in H\}\,.$$

Очевидно, что в коммутативной группе $aH = Ha$.

Утверждение 9.4.

$$a \in H \Rightarrow aH = H$$

◀ Рассмотрим левые смежные классы.
Покажем методом двух включений, что если $a \in H$, то $aH = H$.
С одной стороны

$$(x \in aH) \wedge (a \in H) \Rightarrow \exists h \in H \quad x = ah.$$

Поскольку множество H замкнуто относительно умножения группы $\mathcal{G} \Rightarrow x \in H$.

Обратно,

$$x \in H \Rightarrow\ x = \mathbf{1} \cdot x = (aa^{-1})x = a(a^{-1}x) = ah$$

где $h = a^{-1}x \in H \Rightarrow\ x \in aH.$
Окончательно получим $aH = H$. ▶

Покажем, что с использованием смежных классов можно построить отношение эквивалентности.

Введем **бинарное отношение** $\sim_H$ на множестве G следующим образом: элементы a и b связаны отношением $\sim_H$ ($a \sim_H b$), если и только если левые смежные классы подгруппы $\mathcal{H}$ по элементам a и b совпадают ($aH = bH$).

Теорема 4. Бинарное отношение $\sim_H$ есть эквивалентность на G, причем класс эквивалентности произвольного элемента $a \in G$ совпадает с левым смежным классом aH.

◀ Докажем, что $\sim_H$ является эквивалентностью на G.

$\forall a \in G\ (aH = aH) \Rightarrow a \sim_H a \Rightarrow$ **рефлексивно**;

$a \sim_H b \Rightarrow (aH = bH) \Rightarrow (bH = aH) \Rightarrow (b \sim_H a) \Rightarrow$

$\Rightarrow$ **симметричность**;

$a \sim_H b \wedge b \sim_H c \Rightarrow (aH = bH) \wedge (bH = cH) \Rightarrow a \sim_H c \Rightarrow$

$\Rightarrow$ **транзитивность**

$\sim_H$ есть эквивалентность

Методом двух включений докажем , что класс эквивалентности произвольного элемента a равен aH $[a]_{\sim_H} = aH$. Пусть

$$x \in [a]_{\sim_H} \Rightarrow x \sim_H a \Rightarrow xH = aH$$

Из равенства множеств $xH = \{xh_1 | h_1 \in H\}$ и $aH = \{ah | h \in H\}$ следует, что любой элемент вида $ah \in aH$, $h \in H$, может быть представлен в виде $xh_1 \in xH$, где $h_1 \in H$, т.е. $ah = xh_1$.
Отсюда $x = ahh_1^{-1} = ah_2$, где $h_2 = hh_1^{-1}$. $h_2 \in H$ в силу замкнутости подгруппы $\mathcal{H}$ относительно групповой операции, и $ah_2 \in aH$.
Следовательно, $[a]_{\sim_H} \subseteq aH$.

Докажем, что $aH \subseteq [a]_{\sim_H}$ (второе включение).
Пусть

$$x \in aH, \text{ тогда } \exists h \in H \mid x = ah \Rightarrow xH = ahH.$$

$$ahH = \{(ah)h_3 | h_3 \in H\} = \{a(hh_3) | h_3 \in H\} =$$
$$= \{ah_4 | h_4 \in H\} = aH$$

$$\Rightarrow xH = aH \Rightarrow (x \sim_H a) \Rightarrow x \in [a]_{\sim_H} \Rightarrow aH \subseteq [a]_{\sim_H}$$

$\blacktriangleright$

Определение 9.11. Множества A и B называются **равномощными** $(|A| = |B|)$, если существует взаимнооднозначное отображение (биекция) f множества A на множество B.

Теорема 5. Всякий левый смежный класс подгруппы H **равномощен** H.

◄ Для произвольного фиксированного $a \in G$ зададим отображение φ_a: $H \to aH$ следующим образом:

$$\varphi_a(h) = ah.$$

1. Отображение φ_a есть сюръекция, так как если $y \in aH$, то $y = ah$ для некоторого $h \in H$, откуда $y = \varphi_a(h)$.
2. φ_a — инъекция, поскольку из равенства $ah_1 = ah_2$ в силу законов сокращения в группе $\mathcal{G}$ следует $h_1 = h_2$.
Следовательно, φ_a — биекция и $|aH| = |H|$. ►

Определение 9.12. Порядком конечной группы называется количество элементов этой группы.

Теорема 6 (теорема Лагранжа). Порядок конечной группы делится на порядок любой ее подгруппы.

◄ Во введенном выше отношении эквивалентности $\sim_H$ классом эквивалентности элемента a является множество aH (левый смежный класс подгруппы H по элементу a). Согласно теореме 4, все левые смежные классы образуют разбиение множества G на подмножества, равномощные в силу теоремы 5 подгруппе H.
Так как группа $\mathcal{G}$ конечна, то число элементов разбиения конечно. Обозначив это число через k, заключаем, что $|G| = k|H|$. Следовательно, порядок группы $|G|$ делится на порядок группы $|H|$. ►

Следствия теоремы Лагранжа.

Следствие 9.3. Любая группа простого порядка является циклической.

◄ Возьмем в группе, порядок которой есть простое число, какую-то ее циклическую подгруппу, образующий элемент которой отличен от единицы (нейтрального элемента) группы.
Тогда эта подгруппа содержит не менее двух элементов и ее порядок, согласно теореме Лагранжа, должен быть делителем порядка группы.
Поскольку порядок всей группы — простое число, а порядок подгруппы не меньше 2, то он совпадет с порядком всей группы. ►

Рассмотрим моноид (группу) $(M, \cdot)$.
Подмоноид $(P, \cdot)$ (подгруппу) называют **тривиальным подмоноидом (тривиальной подгруппой)**, если **носитель** содержит только единицу исходного моноида ($P = \{\mathbf{1}\}$) или совпадает с носителем исходного моноида (группы) ($P = M$).

Группу называют **неразложимой**, если она не имеет **нетривиальных подгрупп**.

Следствие 9.4. Конечная группа неразложима тогда и только тогда, когда она является циклической группой, порядок которой есть простое число.

◄ Пусть группа циклическая и ее порядок — простое число. Согласно теореме Лагранжа, каждая ее подгруппа имеет порядок, равный либо единице, либо порядку всей группы, следовательно, группа неразложима.

Обратно. Пусть конечная группа $\mathcal{G} = (G, \cdot, \mathbf{1})$ неразложима.
Покажем, что $|G|$ — простое число.
Выберем элемент $a \neq \mathbf{1}$.
Тогда циклическая подгруппа с образующим элементом a совпадает с $\mathcal{G}$.
Допустим, что $|G|$ — составное число, т.е.

$$\exists(k, l \in \mathbb{N}, k \neq 1, l \neq 1, k \neq |G|, l \neq |G|) \mid |G| = kl$$

Тогда циклическая подгруппа с образующим элементом $b = a^k$ не совпадает с $\mathcal{G}$, так как $b^l = a^{kl} = a^{|G|} = \mathbf{1}$ и в этой подгруппе не более l элементов, что противоречит неразложимости группы $\mathcal{G}$.
Следовательно, порядок группы $\mathcal{G}$ есть простое число. ▶

Следствие 9.5. В конечной группе $\mathcal{G}$ для любого элемента $b \in G$ имеет место равенство $b^{|G|} = 1$.

◄ Если группа $\mathcal{G}$ циклическая и элемент b — ее образующий элемент, утверждение очевидно.
Если же элемент b является образующим элементом некоторой циклической подгруппы группы $\mathcal{G}$ порядка $k < |G|$, то в силу теоремы Лагранжа $|G| = kl$ для некоторого натурального l.
Отсюда получаем $b^{|G|} = b^{kl} = \left(b^k\right)^l = \mathbf{1}^l = \mathbf{1}$. ►

Подмоноид, **носитель** которого содержит только единицу исходного моноида ($P = \{\mathbf{1}\}$), а также подмоноид, носитель которого совпадает с носителем исходного моноида ($P = M$), называют **тривиальным подмоноидом** (в частности, **тривиальной подгруппой**).

Подмоноид, не являющийся тривиальным, называют **нетривиальным подмоноидом** (в частности, **нетривиальной подгруппой**).

Подгруппоид (подполугруппу, подмоноид, подгруппу) $(G, *)$ называют **собственным подгруппоидом (подполугруппой**, **подмоноидом**, **подгруппой**) группоида (полугруппы, моноида, группы) $(K, *)$, если его носитель G есть *собственное подмножество* множества K .

С помощью теоремы Лагранжа (точнее, следствия 9.5) можно доказать, что если целое число n не делится на простое число p, то $n^{p-1}-1$ делится на p. В теории чисел это утверждение известно как **малая теорема Ферма**.

Действительно, пусть $n = rp + k$, где r — целое, а $0 < k < p$ (остаток от деления n на p). Тогда ясно, что $n^{p-1} = k^{p-1} \pmod{\mathrm{p}}$ (достаточно разложить $(rp + k)^{p-1}$ по формуле *бинома Ньютона*). Рассмотрим группу $\mathbb{Z}_p^*$ (*мультипликативную группу вычетов по модулю* p) и в этой группе элемент k. Порядок группы $\mathbb{Z}_p^* = p - 1$. Если $k = 1$, то

$$n^{p-1} - 1 = (1^{p-1} - 1) \pmod{\mathrm{p}} = 0 \pmod{\mathrm{p}}$$

и утверждение очевидно. Согласно следствию 9.5, в группе $\mathbb{Z}_p^*$ справедливо равенство $k^{p-1} = 1$, т.е. $k^{p-1} = 1 \pmod{\mathrm{p}}$, и, следовательно, $k^{p-1} - 1 = 0 \pmod{\mathrm{p}}$, т.е. число k^{p-1} равно 1 по модулю p. Поэтому $n^{p-1} = k^{p-1} = 1 \pmod{\mathrm{p}}$.

Малая теорема Ферма дает возможность доказывать утверждения о делимости очень больших чисел. Например, из нее следует, что при $p = 97$ число 97 является делителем $n^{96} - 1$ для любого n, не делящегося на 97. Подобного рода заключения важны при разработке алгоритмов защиты информации.

Кроме того, используя малую теорему Ферма, можно вычислять в *полях вычетов* по *модулю* p (p — простое число) элементы, обратные к заданным относительно умножения. Действительно, если $a \in \mathbb{Z}_p$, то, так как $a^{p-1} = 1$, умножая последнее равенство на a^{-1}, получим $a^{p-2} = a^{-1}$. Таким образом, для того чтобы вычислить элемент, обратный к a по умножению, достаточно возвести его в степень $p - 2$ или, что равносильно, в степень, равную остатку от деления числа $p - 2$ на порядок циклической подгруппы группы $\mathbb{Z}_p^*$, порожденной элементом a.

Пример 9.15. Рассмотрим, как вычислить элемент, обратный к a по умножению в поле $\mathbb{Z}_{17}$. Согласно полученному выше результату, для вычисления обратного к a элемента нужно найти $a^{17-2} = a^{15}$. Однако объем вычислений можно сократить, если порядок циклической подгруппы, порожденной элементом a, меньше порядка группы.
Порядок группы $\mathbb{Z}_{17}^*$ равен 16, следовательно, порядок циклической подгруппы, порожденной элементом a, может составлять, согласно теореме Лагранжа, 2, 4, 8, 16 (т.е. быть каким-то из делителей числа 16). Поэтому при поиске обратного элемента достаточно проверить следующие степени a (кроме 15-й): 1 (остаток от деления 15 на 2), 3 (остаток от деления 15 на 4) и 7 (остаток от деления 15 на 8).

Найдем элемент, обратный к 2. Очевидно, что $2^{-1} \neq 2$, так как $2 \odot_{17} 2 = 4 \neq 1$. Далее получим $2^3 = 4 \odot_{17} 2 = 8$. Поскольку $2 \odot_{17} 8 = 16 \neq 1$, то $2^3 = 8$ также не является обратным к 2. Вычислим $2^7 = 2^3 \odot_{17} 2^3 \odot_{17} 2 = 8 \odot_{17} 8 \odot_{17} 2 = 9$. Поскольку $9 \odot_{17} 2 = 1$, в итоге получаем $2^{-1} = 9$.

Найдем элемент, обратный к 14. Так как $14 \odot_{17} 14 = 9$, то $14^{-1} \neq 14$. Вычисляем $14^3 = 14 \odot_{17} 9 = 7$, но $14 \odot_{17} 7 = 13$, т.е. $14^3 \neq 14^{-1}$. Далее,

$$14^7 = 14^3 \odot_{17} 14^4 = 7 \odot_{17} 13 = 6,$$
$$14 \odot_{17} 6 = 16 = -1.$$

Мы видим, что и $14^7 \neq 14^{-1}$. Следовательно, остается вычислить $14^{-1} = 14^{15}$. Однако в этом случае вычисления можно сократить, заметив, что $14 \odot_{17} 14^7 = 14 \odot_{17} 6 = -1$. Из последнего равенства получим

$$1 = 14 \odot_{17} (-6) = 14 \odot_{17} 11,$$

откуда $14^{-1} = 11$.

Отметим, что $14^{16} = 1$, т.е. порядок циклической подгруппы, порожденной элементом 14, совпадает с порядком всей группы $\mathbb{Z}_{17}^*$, и, следовательно, эта группа является циклической, порожденной элементом 14 (хотя и не только им).

Ткачев С.Б.

каф. Математического моделирования
МГТУ им. Н.Э. Баумана

ДИСКРЕТНАЯ МАТЕМАТИКА

ИУ5 — 4 семестр, 2015 г.

Лекция 10. АЛГЕБРЫ: ПОЛУКОЛЬЦА

Определение 10.1. Полукольцо — это алгебра с двумя **бинарными** и двумя **нульарными операциями**

$$\mathcal{S} = (S, +, \cdot, \mathbf{0}, \mathbf{1}),$$

такая, что для произвольных элементов a, b, c множества S выполняются следующие равенства, называемые **аксиомами полукольца**:

1) $a + (b + c) = (a + b) + c$;
2) $a + b = b + a$;
3) $a + \mathbf{0} = a$;
4) $a \cdot (b \cdot c) = (a \cdot b) \cdot c$;
5) $a \cdot \mathbf{1} = \mathbf{1} \cdot a = a$;
6) $a \cdot (b + c) = a \cdot b + a \cdot c$;
7) $(b + c) \cdot a = b \cdot a + c \cdot a$;
8) $a \cdot \mathbf{0} = \mathbf{0} \cdot a = \mathbf{0}$.

Первую операцию $+$ называют **сложением полукольца**, вторую операцию $\cdot$ — **умножением полукольца** $\mathcal{S}$; элементы $\mathbf{0}$ и $\mathbf{1}$ называют соответственно **нулем** и **единицей полукольца** $\mathcal{S}$.

Аксиомы полукольца называют также **основными тождествами полукольца**.

Аксиому 8 полукольца называют **аннулирующим свойством нуля** в полукольце.

Полукольцо $\mathcal{S} = (S, +, \cdot, \mathbf{0}, \mathbf{1})$ — это алгебра с двумя бинарными и двумя нульарными операциями, такая, что:

1) алгебра $(S, +, \mathbf{0})$ является коммутативным моноидом; **аддитивный моноид полукольца**;

2) алгебра $(S, \cdot, \mathbf{1})$ является моноидом ; **мультипликативный моноид полукольца**;

3) имеют место свойства (двусторонней) дистрибутивности операции сложения относительно операции умножения;

4) выполняется аннулирующее свойство нуля.

Кольцо есть частный случай полукольца: если кольцо по сложению является абелевой группой, то полукольцо — лишь коммутативный моноид.

Выделим два вида полуколец:
1) **коммутативное полукольцо** с коммутативной операцией умножения;

2) **идемпотентное полукольцо** с идемпотентной операцией сложения.

Пример 10.1. Рассмотрим алгебру

$$\mathcal{R}^+ = (\mathbb{R}^+ \cup \{+\infty\},\ \min,\ +,\ +\infty,\ 0),$$

где $\mathbb{R}^+$ — множество неотрицательных действительных чисел, $\min$ — операция взятия наименьшего из двух данных чисел, $+$ — операция сложения действительных чисел, $+\infty$ — „плюс бесконечность“, 0 — число „нуль“.

Эта алгебра — полукольцо.
Операция взятия наименьшего из двух чисел $\min$ является операцией сложения, сложение действительных чисел $\mathcal{R}^+$ —операцией умножения полукольца.
Элемент $+\infty$ — нуль полукольца.
Элемент 0 — единица полукольца.

Проверим аксиомы полукольца.

First Prev Next Last Go Back Full Screen Close Quit

Операция **сложения полукольца**.

Операция взятия $\min$ ассоциативна и коммутативна (доказать самостоятельно).

Элемент $+\infty$ есть нейтральный элемент относительно операции $\min$ (операции сложения в полукольце):

$$\forall(x \in \mathbb{R}^+ \cup \{+\infty\}); \min(x, +\infty) = x.$$

Выполняются аксиомы 1,2,3 полукольца.

Алгебра $(\mathbb{R}^+ \cup \{+\infty\}, \min, +\infty)$ — коммутативный моноид.

Операция **умножения полукольца**.

Операция сложения действительных чисел $+$ ассоциативна и коммутативна.

Элемент 0 есть нейтральный элемент относительно операции $+$ (операции умножения в полукольце):

$$\forall(x \in \mathbb{R}^+ \cup \{+\infty\})\ (x + 0) = x.$$

Выполняются аксиомы 4,5 полукольца.

Алгебра $(\mathbb{R}^+ \cup \{+\infty\},\ +,\ 0)$ — коммутативный моноид.

Проверим **свойства дистрибутивности** (аксиомы 6 и7):

$$a + \min(b,\ c) = \min(a + b,\ a + c).$$

Имеем

$$a + \min(b,\ c) = \begin{cases} a + b, & b \le c; \\ a + c, & b > c. \end{cases}$$

В то же время

$$\min(a + b,\ a + c) = \begin{cases} a + b, & b \le c; \\ a + c, & b > c. \end{cases}$$

Таким образом,

$$a + \min(b,\ c) = \min(a + b,\ a + c).$$

Элемент $+\infty$ также обладает **аннулирующим свойством** относительно операции сложения чисел (операции умножения в полукольце): $x + (+\infty) = +\infty$. Выполняется аксиома 8 полукольца.

В рассматриваемом полукольце умножение $+$ коммутативно, а сложение $\min$ идемпотентно.

Следовательно, $\mathcal{R}^+$ — идемпотентное коммутативное полукольцо.

Пример 10.2. Рассмотрим алгебру $\mathcal{B} = (\{0, 1\}, +, \cdot, 0, 1)$, в которой операции $+$ и $\cdot$ заданы **таблицами Кэли**

Таблица 10.1

$+$	0	1
0	0	1
1	1	1

Таблица 10.2

$\cdot$	0	1
0	0	0
1	0	1

Проверка аксиом полукольца основана на этих таблицах.

Два элемента 0 и 1 одновременно являются соответственно нулем и единицей данного полукольца.

Полукольцо $\mathcal{B}$ коммутативное и идемпотентное.

Операции полукольца $\mathcal{B}$ можно трактовать как логические связки „или“ и „и“, а элементы 0 и 1 — как „ложь“ и „истина“ соответственно.

Алгебры, являющиеся полукольцами.

Пример 10.3. а. Алгебра $\mathcal{N} = (\mathbb{N} \cup 0, +, \cdot, 0, 1)$ с носителем — множеством неотрицательных целых чисел — и операциями сложения и умножения чисел есть коммутативное полукольцо. Оно не является идемпотентным.

б. Алгебра $\mathcal{S}_A = (2^A, \cup, \cap, \varnothing, A)$ с носителем — множеством всех подмножеств некоторого множества A — и операциями объединения и пересечения есть полукольцо. Оно является идемпотентным и коммутативным.

Введем **отношение порядка** на носителе **идемпотентного** полукольца $\mathcal{S} = (S, +, \cdot, \mathbf{0}, \mathbf{1})$:
для произвольных x , $y \in S$ положим $x \leq y$ тогда и только тогда, когда $x + y = y$, т.е.

$$x \leq y \Leftrightarrow x + y = y. \tag{10.1}$$

Покажем, что введенное бинарное отношение **рефлексивно**, **антисимметрично** и **транзитивно**.

Для идемпотентного полукольца

$$(\forall x)\ x + x = x \Rightarrow (x \leq x)\ (\text{согласно (10.1)})$$

Отношение **рефлексивно**.

$$(x \leq y) \wedge (y \leq x) \Rightarrow (x + y = y) \wedge (y + x = x) \Rightarrow (x = y)$$

в силу коммутативности сложения.
Отношение **антисимметрично**.

$$(x \leq y) \wedge (y \leq z) \Rightarrow (x + y = y) \wedge (y + z = z) \Rightarrow$$
$$\Rightarrow x + z = x + (y + z) = (x + y) + z = y + z = z \Rightarrow$$
$$\Rightarrow x \leq z$$

Отношение **транзитивно**.

Отношение $\leq$ на носителе произвольного идемпотентного полукольца есть отношение порядка.
Будем называть его **естественным порядком идемпотентного полукольца** и говорить, что он задан в этом полукольце.

Всякое идемпотентное полукольцо можно рассматривать как **упорядоченное множество**, причем отношение порядка определяется через сложение этого полукольца согласно (10.1).

Нуль идемпотентного полукольца $\mathcal{S} = (S, +, \cdot, \mathbf{0}, \mathbf{1})$ есть **наименьший элемент** относительно естественного порядка этого идемпотентного полукольца .
$\forall(x \in S)\ (\mathbf{0} \leq x\ )$ поскольку $\mathbf{0} + x = x\ \forall(x \in S)$.

First Prev Next Last Go Back Full Screen Close Quit

Пример 10.4. В полукольце $\mathcal{B}$ (пример 10.2) выполняется равенство $0 + 1 = 1$ и, следовательно, $0 \leq 1$.

Пример 10.5. В полукольце $\mathcal{R}^+$ $x \leq y$, если и только если $\min(x, y) = y$.

Обозначим через $\leq_{\mathbb{R}}$ **естественный числовой порядок** на множестве действительных чисел.
Тогда для произвольных элементов x, y полукольца $\mathcal{R}^+$ соотношение $x \leq y$ означает, что $x \geq_{\mathbb{R}} y$, т.е. число x не меньше числа y относительно естественного числового порядка.
Таким образом, порядок в полукольце $\mathcal{R}^+$ — это **двойственный порядок** для отношения $\leq_{\mathbb{R}}$.

First Prev Next Last Go Back Full Screen Close Quit

В полукольце есть **наименьший элемент** относительно введенного порядка — элемент $+\infty$, поскольку для любого элемента x имеем $\min(x, +\infty) = x$.
Существует **наибольший элемент** — единица полукольца, т.е. число 0 .

Не путать число 0 с **нулем** данного полукольца — элементом $+\infty$.

Пример 10.6. В полукольце $\mathcal{S}_A$ (пример 10.3[б]) получаем в качестве отношения естественного порядка полукольца *отношение* включения $\subseteq$.

Для любых двух множеств X , $Y \in 2^A$ из $X \cup Y = Y$ вытекает $X \subseteq Y$ и наоборот.

Наименьшим элементом является нуль полукольца — $\varnothing$ (пустое множество), наибольшим — единица полукольца (множество A).

First Prev Next Last Go Back Full Screen Close Quit

Теорема 1. Если A — конечное подмножество (носителя) идемпотентного полукольца, то $\sup A$ относительно естественного порядка этого полукольца равен сумме всех элементов множества A.

◀ Пусть $A = \{a_1, \ldots, a_n\}$ и $a = a_1 + \ldots + a_n$.
Для произвольного элемента a_i, $i = \overline{1, n}$, в силу коммутативности и идемпотентности сложения имеем

$$\begin{aligned} a_i + a &= a_i + (a_1 + \ldots + a_i + \ldots + a_n) = \\ &= a_1 + \ldots + a_i + a_i + \ldots + a_n = \\ &= a_1 + \ldots + a_i + \ldots + a_n = a, \end{aligned}$$

т.е. $a_i \leq a$, и поэтому a есть **верхняя грань множества** A.
Покажем, что это **точная верхняя грань множества**.

• First • Prev • Next • Last • Go Back • Full Screen • Close • Quit

Пусть b произвольная верхняя грань множества A.
Рассмотрим сумму $b + a$.
Так как для каждого $i = \overline{1, n}$ имеет место $a_i \leq b$, т.е. $a_i + b = b$, то

$$b + a = b + (a_1 + a_2 + \ldots + a_n) =$$
$$= (b+a_1)+(a_2+\ldots+a_n) = (b+a_2)+\ldots+a_n = \ldots = b.$$

Следовательно, $a \leq b$ и a — точная верхняя грань множества A. ▶

First Prev Next Last Go Back Full Screen Close Quit

10.1. Замкнутые полукольца

Определение 10.2. Полукольцо $\mathcal{S} = (S, +, \cdot, \mathbf{0}, \mathbf{1})$ называют **замкнутым**, если:
1) оно идемпотентно;
2) любая последовательность элементов множества S имеет *точную верхнюю грань* относительно *естественного порядка* $\leq$ этого *идемпотентного полукольца*;
3) операция *умножения полукольца* $\mathcal{S}$ сохраняет *точные верхние грани последовательностей*, т.е. для любого $a \in S$ и любой последовательности $X = \{x_n\}_{n\in\mathbb{N}}$ элементов множества S

$$a \sup X = \sup aX, \quad (\sup X)a = \sup(Xa).$$

First Prev Next Last Go Back Full Screen Close Quit

Теорема 2. Любое конечное идемпотентное полукольцо замкнуто.

◀ Поскольку носитель S идемпотентного полукольца

$$\mathcal{S} = (S, +, \cdot, \mathbf{0}, \mathbf{1})$$

есть конечное множество, то множество элементов любой последовательности в этом полукольце конечно.

Для нахождения точной верхней грани такой последовательности нужно найти **точную верхнюю грань множества** $P = \{p_1, \ldots, p_n\}$ ее членов, т.е., согласно теореме 1, вычислить некоторую конечную сумму, которая всегда существует.

В конечном идемпотентном полукольце любая последовательность имеет точную верхнюю грань.

Условия сохранения точных верхних граней имеют вид

$$a(p_1 + \ldots + p_n) = ap_1 + \ldots + ap_n,$$
$$(p_1 + \ldots + p_n)a = p_1a + \ldots + p_na$$

и выполняются в силу **аксиом полукольца**.
Таким образом, полукольцо $\mathcal{S}$ замкнуто. ►

В любом идемпотентном полукольце сумма произвольного конечного множества элементов является точной верхней гранью этого множества.

В замкнутом полукольце точную верхнюю грань последовательности $\{x_n\}_{n\in\mathbb{N}}$ естественно называть **суммой элементов последовательности**, полагая, по определению,

$$\sum_{n=1}^{\infty} x_n = \sup\{x_n \colon n \in \mathbb{N}\}\,. \tag{10.2}$$

$\sum\limits_{n=1}^{\infty} x_n$ всегда есть элемент множества S (условие 2 определения 10.2).

„Пределы суммирования“ будем опускать и писать просто $\sum x_n$, если это не приводит к недоразумению.

Также будем использовать обозначение $\sum_{n\in\mathbb{N}} x_n$.

Если множество элементов x_n бесконечно, сумму, стоящую в левой части (10.2), будем называть **бесконечной суммой**.

First Prev Next Last Go Back Full Screen Close Quit

Утверждение 10.1. Замкнутое полукольцо является **индуктивным упорядоченным множеством**.

◄ Замкнутое полукольцо с отношением естественного порядка является упорядоченным множеством, в котором **наименьшим элементом** служит **нуль полукольца**, поскольку

$$(\forall x)(\mathbf{0} + x = x \Rightarrow \mathbf{0} \leq x).$$

Точной верхней гранью любой *неубывающей последовательности* $\{x_n\}_{n\in\mathbb{N}}$ является бесконечная сумма $\sum x_n$. ►

Заметим, что точная верхняя грань в замкнутом полукольце существует у любой последовательности, а не только у неубывающей.

Операция умножения на произвольный фиксированный элемент a **непрерывна**, поскольку сохраняет точные верхние грани

$$a\sum x_n = \sum ax_n \quad \text{и} \quad \Big(\sum x_n\Big)a = \sum x_n a.$$

Это следует из определения 10.2, пункт 3.

Исследуем непрерывность операции сложения. Имеет место следующая теорема.

Теорема 3. Для любой последовательности $\{x_n\}_{n\in\mathbb{N}}$ элементов замкнутого полукольца и любого элемента a этого полукольца выполняется равенство

$$a + \sum x_n = \sum (a + x_n).\# \tag{10.3}$$

Тождество (10.3) можно рассматривать как свойство непрерывности операции сложения в замкнутом полукольце.
Это свойство аналогично свойству непрерывности операции умножения, которое имеет место по определению.

Свойства частичных сумм.

Рассмотрим последовательность $\{x_n\}_{n\geq 0}$. Назовем k-ой частичной суммой последовательности

$$s_k = \sum_{i=0}^{k} x_i.$$

При $k = 0$ получим $s_0 = x_0$,
при $k = 1$ — $s_1 = x_0 + x_1$,
при $k = 2$ — $s_2 = x_0 + x_1 + x_2$ и т.д.

Рассмотрим последовательность $\{s_k\}_{k\geq 0}$. Эта последовательность является неубывающей, поскольку

$$s_k + s_{k+1} = s_k + (s_k + x_{k+1}) = (s_k + s_k) + x_{k+1} = s_k + x_{k+1} = s_{k+1}$$

и $s_k \leq s_{k+1}$.

Утверждение 10.2.

$$\sum x_n = \sum s_n$$

Одним из важнейших понятий в замкнутых полукольцах является понятие **итерации** (или **замыкания**) **элемента** замкнутого полукольца.

Итерация x^* элемента x определяется как точная верхняя грань последовательности всех *степеней элемента* x, т.е.

$$x^* = \sum_{n=0}^{\infty} x^n,$$

где, по определению, $x^0 = \mathbf{1}$, а $x^n = x^{n-1}x$, $n = 1, 2, \ldots$

Пример 10.7. Идемпотентное полукольцо $\mathcal{B}$ из примера замкнуто согласно теореме о замкнутости конечного идемпотентного кольца.

Точная верхняя грань любой последовательности элементов этого полукольца равна 1, если хотя бы один ее член равен 1, и равна 0 в противном случае.

Итерация любого элемента полукольца $\mathcal{B}$ равна **1**.

Для 1^* это очевидно, а для 0^* имеем

$$0^* = 0^0 + 0^1 + \ldots + 0^k + \ldots = 1 + 0 + \ldots + 0 + \ldots = 1.$$

10.2. Решение линейных уравнений

Всякое замкнутое полукольцо является индуктивным упорядоченным множеством.

Следовательно, согласно теореме о неподвижной точке любое непрерывное отображение f замкнутого полукольца в себя имеет наименьшую **неподвижную точку**.

Т.е. в любом замкнутом полукольце всякое уравнение вида $x = f(x)$, где f — непрерывное отображение носителя этого полукольца в себя, имеет наименьшее решение.

Особенно важными для приложений являются линейные уравнения в замкнутом полукольце, имеющие вид

$$x = ax + b \tag{10.4}$$

или

$$x = xa + b. \tag{10.5}$$

В силу непрерывности операций сложения полукольца и умножения полукольца правые части уравнений (10.4) и (10.5) есть непрерывные отображения.
Поэтому, согласно теореме о неподвижной точке, существуют наименьшие решения этих уравнений.

First Prev Next Last Go Back Full Screen Close Quit

Теорема 4. Наименьшими решениями уравнений (10.4) и (10.5) в замкнутых полукольцах являются соответственно

$$x = a^* b \tag{10.6}$$

и

$$x = b a^*, \tag{10.7}$$

где a^* — итерация элемента a .

◀ Используем формулу $x = \sup\limits_{n\geq 0} f^n(\mathbf{0})$ для вычисления наименьшей неподвижной точки.
Запишем $\sup$ в случае замкнутого полукольца как бесконечную сумму, для уравнения (10.6).

$$x = \sum_{n\geq 0}^{\infty} f^n(\mathbf{0}), \tag{10.8}$$

где $\mathbf{0}$ — нуль полукольца, а $f(x) = ax + b$.
Вычислим $f^0(\mathbf{0}), f^1(\mathbf{0}), \ldots, f^n(\mathbf{0}), \ldots$

$$
\begin{aligned}
f^0(\mathbf{0}) &= \mathbf{0}, \\
f^1(\mathbf{0}) &= b, \\
f^2(\mathbf{0}) &= ab + b = (a + \mathbf{1})b, \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
f^n(\mathbf{0}) &= (a^{n-1} + \ldots + a^2 + a + \mathbf{1})b,
\end{aligned}
$$

Подставим в выражение (10.8) и получим

$$\sum_{n=0}^{\infty} f^n(\mathbf{0}) = \sum_{n=1}^{\infty}(\mathbf{1} + a + \ldots + a^{n-1})b.$$

Используя непрерывность умножения, вынесем b (вправо) за знак бесконечной суммы и получим

$$\sum_{n=1}^{\infty}(\mathbf{1} + a + \ldots + a^{n-1})b = \Big(\sum_{n=1}^{\infty}(\mathbf{1} + a + \ldots + a^{n-1})\Big)b.$$

Сумма $\mathbf{1} + a + \ldots + a^{n-1}$ есть частичная сумма s_{n-1} последовательности $\{a^n\}_{n\geq 0}$.

Используя равенство $\sum\limits_{n\in\mathbb{N}} x_n = \sum\limits_{n\in\mathbb{N}} s_n = \sum\limits_{k\in\mathbb{N}} (x_1 + \ldots + x_k)$, запишем

$$\sum_{n=1}^{\infty} (\mathbf{1} + a + \ldots + a^{n-1}) = \sum_{n=0}^{\infty} a^n = a^*.$$

Окончательно получим

$$\sum_{n=0}^{\infty} f^n(\mathbf{0}) = a^* b.$$

Анологично доказывается равенство (10.7). ►

Формулы (10.6) и (10.7) дают именно **наименьшие** решения уравнений (10.4) и (10.5), а не все возможные их решения.

Пример 10.8. В полукольце $\mathcal{B} = (\{0, 1\}, +, \cdot, 0, 1)$, можно определить только два уравнения: $x = x + 1$ и $x = x + 0$.
Второе уравнение переписывается совсем просто: $x = x$; его решением является любой элемент полукольца — как 0, так и 1.
Но по формуле (10.6) получим $x = 1^*0 = 0$, что, как и доказано выше, есть наименьшее решение уравнения.

В **полукольце**, в котором **итерация любого элемента** равна **единице полукольца**, формулы (10.6) и (10.7) дают один и тот же результат: $x = b$, т.е. в данном случае наименьшее решение совпадает со свободным членом уравнения.

Материал для самостоятельного изучения

Утверждение 10.3.

$$\sum x_n = \sum s_n$$

◀ Покажем, что для любого элемента x_j имеет место $x_j \leq \sum s_n$. Заметим, что для неубывающей последовательности $\{s_n\}$ и любого $j \leq n$

$$\sum_{n=0} s_n = \sum_{n \geq j} s_n,$$

поскольку точная верхняя грань неубывающей последовательности не изменится, если отбросить конечное число первых членов последовательности от 0 до $j-1$.

Для произвольного j рассмотрим $x_j + \sum\limits_{n=0} s_n$. Имеем

$$x_j + \sum_{n=0} s_n = x_j + \sum_{n\geq j} s_n = \sum_{n\geq j}(x_j + s_n) =$$

$$= \sum_{n\geq j}(x_j + \overbrace{x_1 + \ldots + x_j + \ldots + x_n}^{s_n}) =$$

$$= \sum_{n\geq j}(x_1 + \ldots + x_j + x_j + \ldots + x_n) =$$

$$\sum_{n\geq j}(x_1 + \ldots + x_j + \ldots + x_n) = \sum_{n\geq j} s_n = \sum s_n,$$

Следовательно, $x_j \leq \sum s_n$. Поэтому точная верхняя грань последовательности частичных сумм $\sum s_n$ есть верхняя грань последовательности $\{x_n\}_{n\geq 0}$.

Покажем, что $\sum s_n$ есть **точная** верхняя грань последовательности $\{x_n\}_{n\geq 0}$. Пусть b есть некоторая верхняя грань этой последовательности. Тогда для любого j имеем $x_j + b = b$, поскольку $x_j \leq b$.

Тогда

$$b + s_n = b + \sum_{i=0}^{n} x_i = \sum_{i=0}^{n} (b + x_i) = \sum_{i=0}^{n} b = b,$$

и, следовательно,

$$b + \sum s_n = \sum_{n \geq 0} (b + s_n) = \sum_{n \geq 0} b = b.$$

Таким образом, $\sum s_n \leq b$ и $\sum s_n$ есть наименьший элемент множества верхних граней последовательности $\{x_n\}_{n \geq 0}$, т.е. ее точной верхней гранью и $\sum x_n = \sum s_n$. ▶

Пример 10.9. В идемпотентном полукольце $\mathcal{R}^+ = (\mathbb{R}^+ \cup \{+\infty\}, \min, +, +\infty, 0)$, любая последовательность $\{x_n\}_{n\geq 0}$ есть последовательность **неотрицательных** действительных чисел.

Естественный порядок идемпотентного полукольца $\mathcal{R}^+$ является двойственным к **естественному числовому порядку**.

Последовательность $\{x_n\}_{n\geq 0}$ ограничена снизу и имеет точную нижнюю грань $\inf x_n$ относительно естественного числового порядка (известно из курса математического анализа).

Точная нижняя грань представляет собой точную верхнюю грань относительно **естественного порядка идемпотентного полукольца** $\mathcal{R}^+$.

Для любой последовательности x_n элементов полукольца $\mathcal{R}^+$ точная верхняя грань существует.

Докажем непрерывность операции умножения в этом полукольце, опираясь на естественный числовой порядок.

Непрерывность операции умножения в идемпотентном полукольце $\mathcal{R}^+$ эквивалентна выполнению равенства

$$a + \inf\{x_n\}_{n\in\mathbb{N}} = \inf\{a + x_n\}_{n\in\mathbb{N}}, \qquad (10.9)$$

где a — неотрицательное действительное число;
$\inf$ — точная нижняя грань последовательности относительно естественного числового порядка.
Равенство 10.9 доказывается в курсе математического анализа.

Следовательно, идемпотентное полукольцо $\mathcal{R}^+$ является замкнутым.

• First • Prev • Next • Last • Go Back • Full Screen • Close • Quit

Итерация x^* элемента x в полукольце $\mathcal{R}^+$ есть точная верхняя грань последовательности степеней элемента x.

Поскольку операция умножения в этом полукольце определена как операция сложения действительных чисел, то $x^0 = 0$, так как число 0 есть единица полукольца $\mathcal{R}^+$. Далее, $x^2 = x + x = 2x, \ldots, x^n = nx$.

Для каждого $n \geq 0$ выполняется неравенство $x^n \geq 0$ в смысле естественного числового порядка.

Т.о., число 0 есть наименьший в смысле естественного числового порядка член последовательности $\{x^n\}_{n\in\mathbb{N}}$ и, следовательно, $\inf\{x^n\}_{n\in\mathbb{N}} = 0$.

Переходя к двойственному порядку — естественному порядку полукольца $\mathcal{R}^+$, получим, что число 0 является точной верхней гранью последовательности $\{x^n\}_{n\in\mathbb{N}}$, т.е. $x^* = 0$.

В полукольце $\mathcal{R}^+$ итерация произвольного элемента также равна единице полукольца, т.е. числу 0.

First Prev Next Last Go Back Full Screen Close Quit

Ткачев С.Б.
каф. Математического моделирования
МГТУ им. Н.Э. Баумана

ДИСКРЕТНАЯ МАТЕМАТИКА

ИУ5 — 4 семестр, 2015 г.

Лекция 10. ГРАФЫ. ЗАДАЧА О ПУТЯХ ВО ВЗВЕШЕННЫХ ОРИЕНТИРОВАННЫХ ГРАФАХ

Неориентированный граф G задается двумя множествами

$$G = (V,\, E),$$

где V — конечное множество, элементы которого называют **вершинами** или **узлами**;

E —**множество неупорядоченных пар** на V, т.е. подмножество множества двухэлементных подмножеств V, элементы которого называют **ребрами**.

Для каждого ребра $\{u,\, v\} \in E$ считаем, что u и v — различные вершины.

Если ребро $e = (u, v)$, то говорят, что ребро e соединяет вершины u и v, и обозначают это $u \mapsto v$; если необходимо, указывают имя графа G: $u \mapsto_G v$.

Вершины u и v, соединеные ребром ($u \mapsto v$), называют **смежными**, а также **концами ребра** $\{u, v\}$.

Если $u \mapsto v$, говорят, что вершины u и v связаны **отношением непосредственной достижимости**.

Ориентированный граф G задается двумя множествами

$$G = (V,\, E),$$

где V — конечное множество, элементы которого называют вершинами или узлами;
E — множество **упорядоченных пар** на V, т.е. подмножество множества $V \times V$, элементы которого называют **дугами**.

Если дуга $e = (u, v)$, то говорят, что дуга e ведет из вершины u в вершину v, и обозначают это $u \to v$; если необходимо, указывают имя графа G: $u \to_G v$.

Вершины u и v, такие, что из вершины u в вершину v ведет дуга ($u \to v$), называют **смежными**, u называют **началом**, а v — **концом дуги** (u, v).

Дугу, начало и конец которой есть одна и та же вершина, называют **петлей**.

Если $u \to v$, то говорят, что вершины u и v связаны **отношением непосредственной достижимости**.

Цепь в неориентированном графе G — это последовательность вершин (конечная или бесконечная) v_0, v_1, …, v_n, …, такая, что $v_i \longmapsto v_{i+1}$ для любого i, если v_{i+1} существует. (Под конечной последовательностью понимается *кортеж* вершин.)

Для конечной цепи v_0, v_1, …, v_n число n ($n \geq 0$) называют **длиной цепи**. Таким образом, длина цепи есть число ее ребер, т.е. всех ребер, соединяющих вершины v_i и v_{i+1} ($i = \overline{0,\, n-1}$).

Цепь длины 0 — это произвольная вершина графа.

Говорят, что **вершина** v неориентированного графа G **достижима** из вершины u этого графа и обозначают $u \left|==\right|^* v$, если существует цепь v_0, v_1, . . . , v_n, такая, что $u = v_0$, $v_n = v$ (при этом говорят также, что данная цепь соединяет вершины u и v, которые называют **концами цепи**). Таким образом, задано **отношение достижимости** $\left|==\right|^*$ в неориентированном графе.

Оно является **рефлексивно-транзитивным замыканием** отношения $\mapsto$ непосредственной достижимости.

Отношение достижимости в неориентированном графе **рефлексивно**, **симметрично** и **транзитивно**, т.е. является **отношением эквивалентности**.

Путь в ориентированном графе G — это последовательность вершин (конечная или бесконечная) v_0, v_1, $\ldots$, v_n, $\ldots$, такая, что $v_i \to v_{i+1}$ для любого i, если v_{i+1} существует.

Для конечного пути v_0, v_1, $\ldots$, v_n число n называют **длиной пути** $(n \geq 0)$. Тем самым длина пути есть число его дуг, т.е. всех дуг, которые ведут из вершины v_i в вершину v_{i+1} ($i = \overline{0, n-1}$). Путь длины 0 — это произвольная вершина графа.

Говорят, что **вершина** v ориентированного графа G **достижима** из вершины u этого графа и обозначают $u \Rightarrow^* v$, если существует путь v_0, v_1, $\ldots$, v_n, такой, что $u = v_0$, $v = v_n$ (при этом говорят, что данный путь ведет из вершины u в вершину v, называя первую вершину **началом**, а вторую — **концом** данного **пути**).

Таким образом, задано **отношение достижимости** $\Rightarrow^*$ в ориентированном графе. Оно является **рефлексивно-транзитивным замыканием** отношения $\rightarrow$ непосредственной достижимости.

Отношение достижимости в ориентированном графе рефлексивно и транзитивно, но в общем случае не **антисимметрично**: если две вершины ориентированного графа достижимы одна из другой, то из этого вовсе не следует, что они совпадают. Таким образом, отношение достижимости в ориентированном графе есть *отношение предпорядка*.

Если существует **цепь** ненулевой длины, соединяющая u и v, то пишут $u \,|{=}{=}|^{+}\, v$.
Если необходимо явно указать длину цепи, то пишут $u \,|{=}{=}|^{n}\, v$ и говорят, что существует цепь длины n, соединяющая u и v.

Простая цепь — это цепь, все вершины которой, кроме, быть может, первой и последней, попарно различны и все ребра попарно различны.

Простую цепь ненулевой длины с совпадающими концами называют **циклом**.

Произвольную цепь ненулевой длины с совпадающими концами, все ребра которой попарно различны, будем называть **замкнутой цепью**.

Если существует **путь** ненулевой длины, ведущий из u в v , то пишут $u \Rightarrow^{+} v$.

Если необходимо явно указать длину пути, то пишут $u \Rightarrow^{n} v$ и говорят, что существует путь длины n ,ведущий из u в v .

Простой путь — это путь, все вершины которого, кроме, быть может, первой и последней, попарно различны.

Простой путь ненулевой длины, начало и конец которого совпадают, называют **контуром**.

Произвольный путь ненулевой длины, начало и конец которого совпадают, будем называть **замкнутым путем**.

Неориентированный граф, не содержащий циклов, называют **ациклическим графом**.

Ориентированный граф, не содержащий контуров, называют **бесконтурным графом**.

Определение 10.1. **Взвешенным** (или **размеченным**) **ориентированным графом** называют пару $W = (G, \varphi)$, где $G = (V, E)$ — обычный ориентированный граф, $\varphi\colon E \to \mathcal{R}$ — **весовая функция** (или **функция разметки**) со значениями в некотором **идемпотентном полукольце** $\mathcal{R} = (R, +, \cdot, \mathbf{0}, \mathbf{1})$, причем $(\forall e \in E)(\varphi(e) \neq \mathbf{0})$.

Мы будем в этом случае также говорить, что ориентированный граф размечен над идемпотентным полукольцом $\mathcal{R}$.
Часто полукольцо $\mathcal{R}$ является **замкнутым**, хотя это требование необязательно.
$\mathcal{R}$ — обязательно **полукольцо с итерацией**.

Пусть вершины ориентированного графа каким-либо образом пронумерованы. Тогда взвешенный ориентированный граф может быть задан матрицей A, элемент которой a_{ij} равен значению $\varphi((i,\ j))$ весовой функции на дуге $(i,\ j)$, если из вершины i ведет дуга в вершину j, или **нулю полукольца** в противном случае. Эту матрицу будем называть **матрицей меток дуг**.

Важные задачи анализа ориентированных графов.

1. Вычисление для заданного ориентированного графа его **матрицы достижимости**.

Задача построения **транзитивного замыкания ориентированного графа**.

Матрицу достижимости можно рассматривать как матрицу **транзитивного** и **рефлексивного замыкания бинарного отношения непосредственной достижимости** в ориентированном графе.

2. **Задача о кратчайших расстояниях** .

Вычисление наименьших расстояний между всеми парами **вершин** в ориентированном графе.

Расстоянием от вершины v до вершины w **по пути** S называют сумму меток дуг, входящих в этот путь. **Наименьшие** расстояние это **минимальное** из расстояний между вершинами v и w по всем возможным путям.

3. Перечисление всех путей между двумя произвольными вершинами. Эту задачу будем называть **задачей о перечислении путей**. При ее решении требуется для любой заданной пары вершин u и v ориентированного графа получить все пути, для которых u является началом, а v — концом.

Вычисление **итерации** A^* матрицы A дает решение всех сформулированных задач, если для каждой задачи выбирать соответствующее полукольцо.

В случае полукольца $\mathcal{B}$ получаем решение задачи о транзитивном замыкании, в случае полукольца $\mathcal{R}^+$ — решение задачи о кратчайших расстояниях.

Задачу вычисления матрицы A^* для ориентированного графа, размеченного над произвольным полукольцом с итерацией, в частности над **замкнутым полукольцом**, будем называть **общей задачей о путях** во взвешенных ориентированных графах.

Рассмотрим теперь решение общей задачи о путях для произвольного замкнутого полукольца $\mathcal{R}$.

Определение 10.2. Метка пути, ведущего из вершины v_i в вершину v_j, есть **произведение в полукольце** $\mathcal{R}$ меток входящих в путь дуг в порядке их следования (для пути ненулевой длины) и есть $\mathbf{1}$ (**единица полукольца** $\mathcal{R}$) для пути нулевой длины.

Определение 10.3. Стоимость прохождения из вершины v_i в вершину v_j (или между i-й и j-й вершинами) — это **сумма в полукольце** $\mathcal{R}$ меток всех путей, ведущих из вершины v_i в вершину v_j.

Сумма, определяющая стоимость прохождения есть **бесконечная сумма** замкнутого полукольца, т.е. **точная верхняя грань** соответствующей **последовательности меток**.

Если стоимость прохождения между парой вершин по какому-либо множеству путей равна $\mathbf{0}$, то это означает, что не существует пути, принадлежащего данному множеству путей, ведущего из первой вершины рассматриваемой пары во вторую вершину.

Матрица меток дуг является элементом **полукольца матриц над полукольцом** $\mathcal{R}$. В этом полукольце определены операции сложения и умножения матриц, а также возведение матрицы в неотрицательную степень.

10.1. Решение систем линейных уравнений

Рассмотрим множество $\mathbb{M}_{m\times n}(\mathcal{S})$ прямоугольных матриц типа $m\times n$ с элементами из произвольного идемпотентного полукольца $\mathcal{S} = (S, +, \cdot, \mathbf{0}, \mathbf{1})$.

Множество всех квадратных матриц порядка n с элементами из полукольца $\mathcal{S}$ обозначим $\mathbb{M}_n(\mathcal{S})$.

Операции сложения и умножения матриц определяют точно так же, как и в числовом случае , — с учетом того, что сложение и умножение элементов матриц понимаются в смысле данного идемпотентного полукольца $\mathcal{S}$:

1) суммой матриц $A = (a_{ij})$ и $B = (b_{ij})$ типа $m \times n$ называют матрицу $C = (c_{ij})$ того же типа с элементами $c_{ij} = a_{ij} + b_{ij}$, $i = \overline{1, m}$, $j = \overline{1, n}$, и используют обозначение $C = A + B$;
2) произведением AB матриц $A = (a_{ij})$ типа $m \times n$ и $B = (b_{ij})$ типа $n \times p$ называют матрицу $C = (c_{ij})$ типа $m \times p$ с элементами

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}.$$

Нулевая (O) и единичная (E) матрицы любого порядка определяются с помощью единицы и нуля полукольца.

На множестве $\mathbb{M}_n(\mathcal{S})$ всех квадратных матриц фиксированного порядка n можно определить **алгебру**

$$\mathrm{M}_n(\mathcal{S}) = (\mathbb{M}_n(\mathcal{S}), +, \cdot, O, E).$$

Теорема 1.
Алгебра $\mathrm{M}_n(\mathcal{S})$ есть идемпотентное полукольцо. Если полукольцо $\mathcal{S}$ замкнуто, то полукольцо $\mathrm{M}_n(\mathcal{S})$ тоже замкнуто. (без доказательства)

Полукольцо $\mathrm{M}_n(\mathcal{S})$ будем называть **полукольцом матриц над полукольцом** $\mathcal{S}$.

Теорема 1 позволяет решать произвольные уравнения вида

$$X = AX + B \tag{10.1}$$

или

$$X = XA + B \tag{10.2}$$

относительно неизвестной матрицы X .

Наименьшие решения этих уравнений есть

$$X = A^* B \tag{10.3}$$

и

$$X = B A^* \tag{10.4}$$

соответственно, где A^* — **итерация** матрицы A в $\mathrm{M}_n(\mathcal{S})$. Итерация A^* матрицы A играет в теории линейных уравнений в замкнутых полукольцах такую же роль, как обратная матрица в классической линейной алгебре.

Основную роль при решении задач теории ориентированных графов и теории формальных языков играют **праволинейные уравнения** вида (10.1), поэтому мы будем, как правило, рассматривать только их.

Леволинейное уравнение (10.2) может быть проанализировано аналогично.

Разработаем технику поиска решений матричных уравнений в матричном полукольце над замкнутым полукольцом.

Пусть X^j — j -й столбец матрицы X , B^j — j -й столбец матрицы B .
Перепишем уравнение (10.1) как систему уравнений относительно неизвестных столбцов матрицы X :

$$X^j = AX^j + B^j, \quad 1 \leq j \leq n. \tag{10.5}$$

Наименьшее решение этой системы, как следует из (10.3), есть

$$X^j = A^* B^j. \quad 1 \leq j \leq n. \tag{10.6}$$

Каждая система вида (10.5) есть матричная форма записи системы линейных уравнений вида

$$\begin{cases} x_1 = a_{11}x_1 + \ldots + a_{1n}x_n + b_1, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_n = a_{n1}x_1 + \ldots + a_{nn}x_n + b_n, \end{cases} \tag{10.7}$$

где все элементы a_{ij} , $1 \leq i,\ j \leq n$, b_i , $1 \leq i \leq n$, есть элементы некоторого **замкнутого полукольца**.

Для поиска решения системы вида 10.7 можно воспользоваться **методом последовательного исключения неизвестных**, аналогичным классическому **методу Гаусса**.

Процедура решения системы уравнений (10.7). Запишем первое уравнение системы так:

$$x_1 = a_{11}x_1 + (a_{12}x_2 + \ldots + a_{1n}x_n + b_1).$$

Из первого уравнения системы выразим x_1 через остальные неизвестные, воспользовавшись формулой $x = a^* b$:

$$x_1 = a_{11}^*(a_{12}x_2 + \ldots + a_{1n}x_n + b_1). \qquad (10.8)$$

В формуле (10.8) выражение в скобках не содержит неизвестного x_1.

Подставляя (10.8) вместо x_1 в остальные уравнения, получаем систему из $n-1$ уравнений, которая уже не содержит x_1 :

$$\begin{cases} x_2 = a_{21}a_{11}^*(a_{12}x_2 + \ldots + a_{1n}x_n + b_1) + \\ \qquad\qquad + a_{22}x_2 + \ldots + a_{2n}x_n + b_2, \\ x_3 = a_{31}a_{11}^*(a_{12}x_2 + \ldots + a_{1n}x_n + b_1) + \\ \qquad\qquad + a_{32}x_2 + \ldots + a_{3n}x_n + b_3, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_n = a_{n1}a_{11}^*(a_{12}x_2 + \ldots + a_{1n}x_n + b_1) + \\ \qquad\qquad + a_{n2}x_2 + \ldots + a_{nn}x_n + b_n. \end{cases} \tag{10.9}$$

Приведем подобные члены в каждом уравнении системы и получим:

$$\begin{cases} x_2 = (a_{21}a_{11}^*a_{12} + a_{22})x_2 + \ldots \\ \qquad\qquad \ldots + (a_{21}a_{11}^*a_{1n} + a_{2n})x_n + a_{21}a_{11}^*b_1 + b_2, \\ x_3 = (a_{31}a_{11}^*a_{12} + a_{32})x_2 + \ldots \\ \qquad\qquad \ldots + (a_{31}a_{11}^*a_{1n} + a_{3n})x_n + a_{31}a_{11}^*b_1 + b_3, \\ \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_n = (a_{n1}a_{11}^*a_{12} + a_{n2})x_2 + \ldots \\ \qquad\qquad \ldots + (a_{n1}a_{11}^*a_{1n} + a_{nn})x_n + a_{n1}a_{11}^*b_1 + b_n. \end{cases} \tag{10.10}$$

Перепишем первое уравнение этой системы:
$x_2 = (a_{21}a_{11}^*a_{12} + a_{22})x_2 + \gamma_2$,
где $\gamma_2 = a_{21}a_{11}^*(a_{13}x_3 + \ldots + a_{1n}x_n + b_1) + a_{23}x_3 + \ldots + a_{2n}x_n + b_2$.
γ_2 не содержит x_1 и x_2.
Выразим x_2 через остальные неизвестные, воспользовавшись формулой $x = a^*b$:

$$x_2 = \alpha_2^*\gamma_2, \tag{10.11}$$

где $\alpha_2 = a_{21}a_{11}^*a_{12} + a_{22}$ не содержит неизвестных. Исключаем x_2 из остальных уравнений.

Действуя подобным образом, на i-м шаге ($1 \leq i \leq n$) получаем

$$x_i = \alpha_i^* \gamma_i, \tag{10.12}$$

где выражение α_i^* не содержит неизвестных, а выражение γ_i может содержать только неизвестные, начиная с $(i+1)$-го, т.е. $x_{i+1}, \ldots, x_n$.

При $i = n$ имеем

$$x_n = \alpha_n^* \gamma_n, \tag{10.13}$$

где выражения α_n^* и γ_n не содержат неизвестных.

Таким образом, исходная система (10.7) преобразована к „треугольному“ виду: правая часть уравнения (10.13) не содержит неизвестных, уравнение (10.12) при $i = n - 1$ в правой части содержит только одно неизвестное x_{n-1} и каждое следующее уравнение при просмотре „снизу вверх“ на одно неизвестное больше, чем предыдущее.

Первое уравнение системы — уравнение (10.8) — в правой части содержит все неизвестные от x_2 до x_n.

На этом завершается первый этап алгоритма, который называют **прямым ходом метода Гаусса**.

Обратный ход метода Гаусса

Второй этап алгоритма состоит в последовательном нахождении значения всех неизвестных $x_1, \ldots, x_{n-1}$, начиная с x_{n-1}.

Найдем x_{n-1}, подставив в выражение для x_{n-1} вместо x_n правую часть (10.13).

Затем определим x_{n-2}, подставив полученные значения x_n и x_{n-1} в правую часть выражения (10.12) при $i = n - 2$, и так далее до тех пор, пока не найдем x_1.

Положив $B = E$ в уравнении (10.1), получим $X = A^*E = A^*$.

Таким образом, чтобы вычислить итерацию матрицы A, достаточно решить матричное уравнение (10.5) для всех $j = \overline{1, n}$ при β_j, равном j-му столбцу единичной матрицы E.

Утверждение 10.1. Если A — матрица, все элементы которой принадлежат некоторому полукольцу с итерацией, то все элементы ее **итерации** A^* также принадлежат этому полукольцу с итерацией.

Лемма 1. Элемент $a_{ij}^{(l)}$ матрицы A^l, $l \geq 0$, равен стоимости прохождения из вершины v_i в вершину v_j по всем путям длины l.

◄ Доказательство проведем индукцией по l.
При $l = 0$ получаем $A^0 = E$, где E — единичная матрица, которая будет матрицей стоимости прохождения по всем путям длины 0. Это согласуется с определением 10.3.
При $l = 1$ получаем $A^1 = A$. Матрица меток дуг A — матрица стоимости прохождения по всем путям длины 1.

Согласно предположению индукции, элемент $a_{ik}^{(l-1)}$ равен стоимости прохождения из вершины v_i в вершину v_k по всем путям длины $l-1$.

Множество всех путей длины l из вершины v_i в вершину v_j, проходящих через фиксированную k-ю вершину так, что вершина v_k связана дугой с вершиной v_j $(v_k \to v_j,)$ образуется путем присоединения дуги (v_k, v_j) к каждому из путей, ведущих из v_i в v_k и имеющих длину $l-1$.

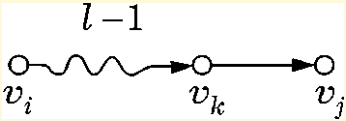

Рис. 1

$a_{ij}^{(l)} = \sum_{k=1}^{n} a_{ik}^{(l-1)} a_{kj}$. Выражение для элемента $a_{ij}^{(l)}$ дает стоимость прохождения из вершины v_i в вершину v_j по всем путям длины l ▶

Стоимость прохождения между парой вершин (v_i, v_j) равна сумме меток всех путей, ведущих из первой вершины во вторую, указанную сумму можно можно получить, суммируя последовательно метки путей длины 0, длины 1, длины 2 и т.д.

Матрица стоимостей взвешенного ориентированного графа с учетом леммы 1 (лекция 15) может быть представлена в виде

$$C = A^0 + A^1 + A^2 + \ldots + A^n + \ldots = \sum_{n \geq 0} A^n = A^*.$$

До сих пор мы рассматривали матрицы над замкнутым полукольцом.
Однако, если элементы матрицы A принадлежат некоторому полукольцу с итерацией, из утверждения 1(лекция 15) следует, что и все элементы матрицы стоимостей $C = A^*$ останутся в этом же полукольце. Таким образом, полученные результаты можно перенести на произвольное полукольцо с итерацией.

Теорема 2. Матрица стоимостей ориентированного графа G, размеченного над полукольцом с итерацией $\mathcal{R}$ (в частности, над замкнутым полукольцом), равна итерации матрицы A меток дуг ориентированного графа G.

Для вычисления $C = A^*$ достаточно решить (т.е. найти наименьшее решение) в $\mathcal{R}$ при всех $j = \overline{1, n}$ систему уравнений

$$X^j = AX^j + E^j,$$

где $E^j \in \mathcal{R}^n$ — j-й единичный вектор, т.е. вектор, все элементы которого, кроме j-го, равны $\mathbf{0}$, а j-й равен единице полукольца $\mathcal{R}$, j-й столбец матрицы E.
Наименьшее решение имеет вид $X^j = A^*E^j$.
Тогда столбец $X^j = A^*E^j$ есть j-й столбец матрицы A^*.

Смысл матрицы стоимостей $C = A^*$ для полуколец $\mathcal{B}$ и $\mathcal{R}^+$.

В полукольце $\mathcal{B}$ метка отдельного пути всегда равна 1 (так как метка дуги в размеченном над полукольцом графе не может, согласно определению, быть нулем полукольца). Следовательно, стоимость $c_{ij} = 1$, если существует хотя бы один путь из i-й вершины в j-ю, и $c_{ij} = 0$, если иначе. Для полукольца $\mathcal{B}$ матрица стоимостей совпадает с матрицей достижимости ориентированного графа.

В полукольце $\mathcal{R}^+$ метка пути — это арифметическая сумма меток его дуг, так как умножение в $\mathcal{R}^+$ — это обычное арифметическое сложение.

Поскольку сложение в $\mathcal{R}^+$ — это взятие наименьшего из слагаемых, то стоимость c_{ij} — это наименьшая из меток пути среди всех путей, ведущих из i-й вершины в j-ю, т.е. это и есть наименьшая длина пути между указанными вершинами.

Таким образом, в полукольце $\mathcal{R}^+$ матрица стоимостей является матрицей кратчайших расстояний, т.е. наименьших длин путей между всеми парами вершин ориентированного графа.

Пример 10.1.

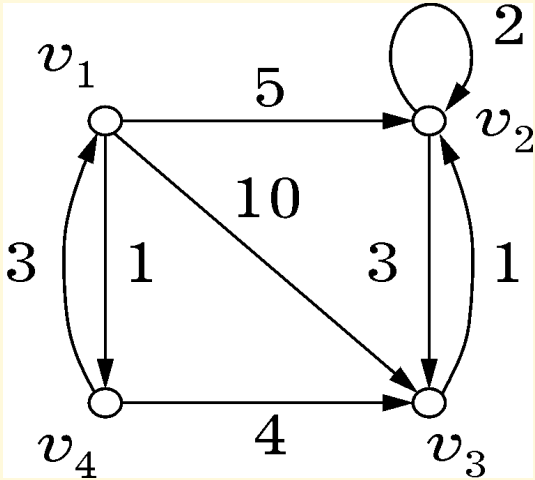

Рис. 2

Вычислим матрицу достижимости изображенного на рисунке графа в полукольце $\mathcal{R}^+$.

1. Вычислим матрицу достижимости в полукольце $\mathcal{B}$.
Считаем, что ориентированный граф размечен над полукольцом $\mathcal{B}$ и метка каждой дуги равна 1 (на числовые метки дуг внимания пока не обращаем).
Ориентированный граф задан матрицей:

$$A = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

Запишем систему уравнений в полукольце $\mathcal{B}$ для определения первого столбца матрицы A^* :

$$\begin{cases} x_1 = & x_2 + x_3 + x_4 + 1, \\ x_2 = & x_2 + x_3 & + 0, \\ x_3 = & x_2 & + 0, \\ x_4 = x_1 & + x_3 & + 0. \end{cases} \tag{10.14}$$

Часто нулевые слагаемые не записывают, как и в системах уравнений в поле действительных чисел.

Для вычисления матрицы достижимости воспользуемся **методом последовательного исключения неизвестных**.

В правой части первого уравнения нет переменной x_1, исключим эту переменную из системы, подставив в остальные уравнения (в 4-ое).

С учетом идемпотентности сложения $(x_3 + x_3 = x_3)$, получим

$$\begin{cases} x_2 = x_2 + x_3 + 0, \\ x_3 = x_2 + 0, \\ x_4 = x_2 + x_3 + x_4 + 1. \end{cases}$$

Из второго уравнения имеем $x_2 = 1^*(x_3 + 0)$.

В полукольце $\mathcal{B}$ итерация любого элемента равна единице полукольца. Поэтому $x_2 = x_3 + 0$.

Исключим x_2 из системы, получим

$$\begin{cases} x_3 = x_3 + 0, \\ x_4 = x_3 + x_4 + 1 \qquad (*). \end{cases}$$

$x_3 = 1^*0 = 1 \cdot 0 = 0$. Подставим $x_3 = 0$ в $(*)$, $x_4 = 1^*1 = 1$.

Далее подставляем $x_3 = 0$ в выражение $x_2 = x_3 + 0$, $x_2 = 0$, затем полученные значения x_2, x_3 и x_4 подставим в первое уравнение $x_1 = x_2 + x_3 + x_4 + 1 = 0 + 0 + 1 + 1;\ x_1 = 1$.

Первый столбец A^* $\begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$.

Второй столбец A^* определим из системы

$$\begin{cases} x_1 = & x_2 + x_3 + x_4 + 0, \\ x_2 = & x_2 + x_3 & + 1, \\ x_3 = & x_2 & + 0, \\ x_4 = x_1 & + x_3 & + 0. \end{cases}$$

Исключим x_1.

$$\begin{cases} x_2 = x_2 + x_3 & + 1, \\ x_3 = x_2 & + 0, \\ x_4 = x_2 + x_3 + x_4 + 0. \end{cases} \qquad (*)$$

Из $(*)$ получим $x_2 = 1^*(x_3 + 1) = x_3 + 1$.

$$\begin{cases} x_3 = (x_3 + 1) + 0, & (**) \\ x_4 = x_3 + x_4 + 1. \end{cases}$$

Из $(**)$ получим $x_3 = x_3 + 1;\ \Rightarrow x_3 = 1^*1 = 1$
$x_4 = 1 + x_4 + 1 \Rightarrow x_4 = 1^*1 = 1;\ x_2 = 1 + 1 = 1;\ x_1 = 1 + 1 + 1 + 0 = 1\,.$

Второй столбец A^* $\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$.

Аналогично вычисляем третий и четвертый столбцы и в результате получаем матрицу A^*:

$$A^* = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$

Анализ этой матрицы показывает , что данный граф связен и имеет две бикомпоненты: $\{v_1,\, v_4\}$ и $\{v_2,\, v_3\}$.

В полукольце $\mathcal{B}$ можно упростить решение систем уравнений, воспользовавшись свойствами полукольца.
Наименьшее решение уравнения

$$x_k = \sum_{i=1}^{n} a_i x_i + 1$$

есть $x_k = 1$ и не зависит от значений переменных в правой части уравнения.
С учетом этого решение системы (10.14) упростится.
Так, из первого уравнения сразу получаем $x_1 = 1$.
Тогда четвертое уравнение принимает вид $x_4 = x_3 + 1$, откуда $x_4 = 1$. Поскольку x_1 и x_4 не входят в оставшиеся два уравнения, их решение нужно искать, используя метод исключения.

2. Вычислим матрицу достижимости в полукольце $\mathcal{R}^+$.

Для упрощения записи ∞ здесь будем понимать как $+\infty$. Взвешенный ориентированный граф задан матрицей:

$$A = \begin{pmatrix} \infty & 5 & 10 & 1 \\ \infty & 2 & 3 & \infty \\ \infty & 1 & \infty & \infty \\ 3 & \infty & 4 & \infty \end{pmatrix}. \qquad (10.15)$$

В полукольце $\mathcal{R}^+$:
элементы 1 и 0 не являются **единицей** и **нулем полукольца**, т.е. $x \neq x + 0$ и $x \neq 1 \cdot x$ в общем случае;
сложение ($\oplus$) — **взятие наименьшего** из двух чисел, **умножение** ($\odot$) — обычное **арифметическое сложение**;
наличие слагаемого 0 в любой сумме означает, что вся сумма равна 0; слагаемое $+\infty$ можно не записывать (как нуль полукольца);
итерация любого элемента равна единице полукольца.

Система для вычисления первого столбца матрицы A^* имеет вид

$$\begin{cases} x_1 = & 5 \odot x_2 \oplus 10 \odot x_3 \oplus 1 \odot x_4 \oplus 0, \\ x_2 = & 2 \odot x_2 \oplus \ 3 \odot x_3 \qquad\quad \oplus \infty, \\ x_3 = & 1 \odot x_2 \qquad\qquad\quad \oplus \infty, \\ x_4 = 3 \odot x_1 & \qquad \oplus \ 4 \odot x_3 \qquad\quad \oplus \infty. \end{cases}$$

Из первого уравнения системы следует, что $x_1 = 0$, так как одно из слагаемых в правой части есть элемент 0 .
Из второго уравнения получаем
$x_2 = 2^* \odot (3 \odot x_3 \oplus \infty) = 3 \odot x_3 \oplus \infty$.

Исключая x_2 из остальных уравнений системы и учитывая, что $x_1 = 0$, получаем

$$\begin{cases} x_2 = 3 \odot x_3 \oplus \infty, \\ x_3 = 1 \odot (3 \odot x_3) \oplus \infty, \\ x_4 = 3 \odot 0 \oplus 4 \odot x_3 \oplus \infty. \end{cases}$$

Далее, из второго уравнения имеем

$$x_3 = (1 \odot 3) \odot x_3 \oplus \infty = 4 \odot x_3 \oplus \infty,$$

откуда $x_3 = 4^* \odot \infty = \infty$, и поэтому

$$x_4 = 3 \odot 0 \oplus 4 \odot \infty \oplus \infty = 3 \oplus \infty = 3.$$

Подставляя найденное значение x_3 в выражение для x_2, получаем $x_2 = \infty$.
Первый столбец матрицы A^*:

$$\begin{pmatrix} 0 \\ \infty \\ \infty \\ 3 \end{pmatrix}.$$

Этот столбец содержит кратчайшие расстояния от всех вершин графа до вершины v_1. Наличие в нем нулей полукольца во второй и третьей строках говорит о том, что вершина v_1 не достижима из вершин v_2 и v_3.
Аналогично вычисляются остальные столбцы матрицы A^*.

Результат: $A^* = \begin{pmatrix} 0 & 5 & 5 & 1 \\ \infty & 0 & 3 & \infty \\ \infty & 1 & 0 & \infty \\ 3 & 5 & 4 & 0 \end{pmatrix}$.

Для данного простого ориентированного графа легко сопоставить полученный алгебраический результат с результатом „визуального“ анализа ориентированного графа.

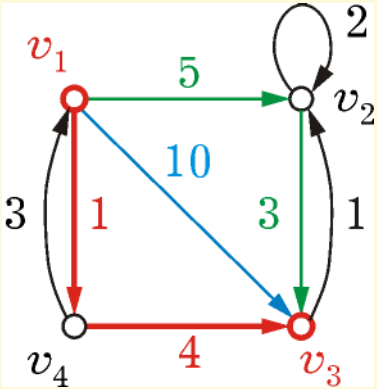

Рис. 3

Рассмотрим вершины $(v_1,\ v_3)$. Из вершины v_1 в вершину v_3 есть различные пути.

Пути, содержащие контуры и петли рассматривать не будем.

Вычислим метки по **простым путям**. По пути $v_1 \to v_4 \to v_3$ сумма меток равна 5, по пути $v_1 \to v_3$ — 10, по пути $v_1 \to v_2 \to v_3$ — 8.

Кратчайшее расстояние — 5, совпадает с ответом, полученным алгебраически: элемент a^*_{13} также равен 5.

Ткачев С.Б.

каф. Математического моделирования
МГТУ им. Н.Э. Баумана

ДИСКРЕТНАЯ МАТЕМАТИКА

ИУ5 — 4 семестр, 2015 г.

Лекция 12. ДЕРЕВЬЯ. МЕТОДЫ СИСТЕМАТИЧЕСКОГО ОБХОДА ВЕРШИН ГРАФА

12.1. Представление графа матрицей смежности вершин.

Матрица смежности вершин, или **булева матрица** графа — это квадратная матрица B порядка n, элементы которой определяют следующим образом:
для неориентированного графа

$$\begin{cases} b_{ij} = 1, i\text{-я и } j\text{-я вершины } \textit{смежные}; \\ 0, \text{иначе}; \end{cases}$$

для ориентированного графа

$$\begin{cases} b_{ij} = 1, \text{из } i\text{-й вершины в } j\text{-ю ведет дуга}; \\ 0, \text{иначе}; \end{cases}$$

В k -й строке матрицы ориентированного графа количество единиц равно **полустепени исхода** $\mathrm{dg}^+ v_k$ вершины v_k , а количество единиц в k -м столбце — **полустепени захода** $\mathrm{dg}^- v_k$.

Для неориентированного графа **матрица** смежности вершин **симметрическая**.

Эта матрица есть матрица бинарного отношения непосредственной достижимости на множестве вершин V .

Пример 12.1. Для ориентированного графа $G = (V,\, E)$, где $V = \{v_1,\, v_2,\, v_3\}$,
$E = \{\{v_1, v_2\},\, \{v_1, v_3\},\, \{v_2, v_3\},\, \{v_3, v_2\}\}$.

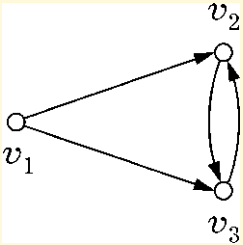

Рис. 1

Матрица смежности вершин:

$$\begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

12.2. Списки смежности

В ориентированном графе для задания множества **вершин**, **непосредственно достижимых** из вершины v, используют **линейный однонаправленный список**.

Рис. 2

Задать для вершины v ее список смежности означает в произвольном порядке в **данные** элементов списка поместить номера вершин u, для которых в ориентированном графе есть дуга из v в u ($v \to u$). **Список смежности вершины** v обозначают $L(v)$.

Если количество вершин ориентированного графа известно заранее, то ориентированный граф удобно задавать в виде структуры, называемой **массивом лидеров**.

Пример 12.2. Ориентированный граф $G = (V,\ E)$: $V = \{v_1,\ v_2,\ v_3,\ v_4,\ v_5\}$, $E = \{(v_1,\ v_2), (v_2,\ v_2), (v_2,\ v_3), (v_2,\ v_4), (v_2,\ v_5), (v_3,\ v_1), (v_5,\ v_2)\}$. Списки смежности, собранные в массив лидеров:

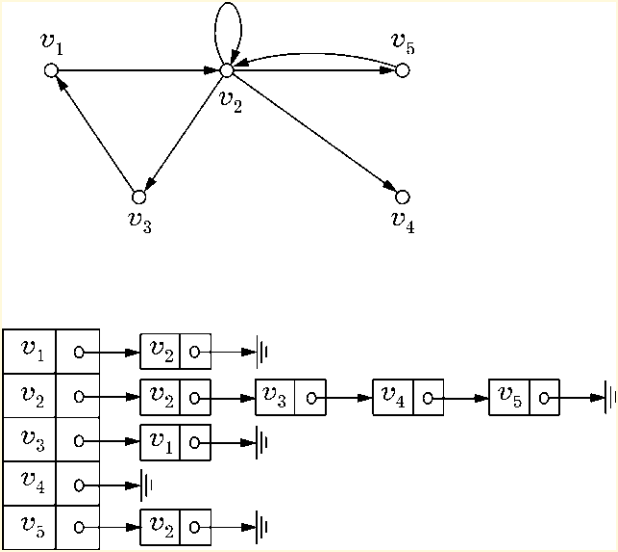

Рис. 3

Неориентированный граф задать с помощью списков смежности можно так же, как и ориентированный. Здесь в список смежности вершины v войдут все вершины, смежные с ней, а списки смежности могут быть собраны в массив лидеров.

12.3. Деревья

Определение 12.1. Неориентированным деревом называют **связный** и **ациклический неориентированный граф**.

Определение 12.2. Ориентированным деревом называют **бесконтурный ориентированный граф**, у которого **полустепень захода** любой вершины не больше 1 и существует ровно одна вершина, называемая **корнем ориентированного дерева**, **полустепень захода** которой равна 0.

В ориентированном дереве любая вершина **достижима** из корня.
Требование бесконтурности ориентированного графа в определении 12.2 является обязательным.

Определение 12.3. Вершину v ориентированного дерева называют **потомком (подлинным потомком)** вершины u, если существует путь из u в v (путь ненулевой длины из u в v). В этом же случае вершину u называют **предком (подлинным предком)** вершины v, а если длина пути из u в v равна 1, то вершину v называют **сыном** вершины u, которая при этом вполне естественно именуется **отцом** вершины v.

Вершину, не имеющую потомков, называют **листом**.

Пример 12.3.

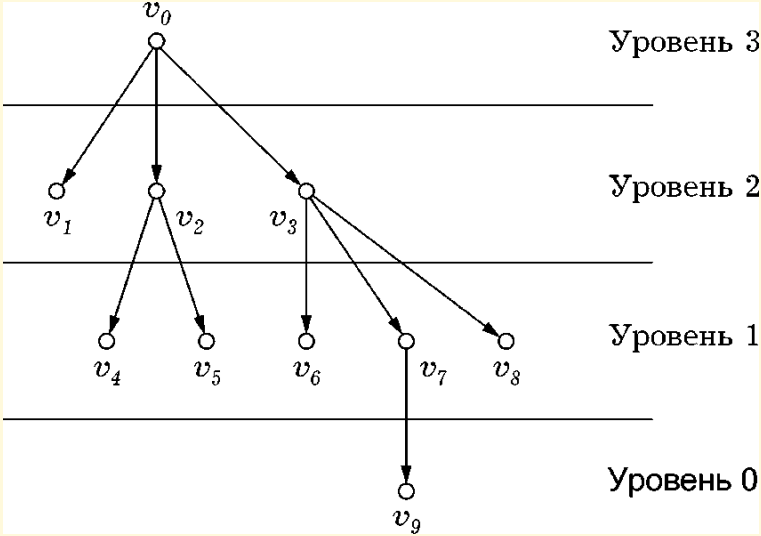

Рис. 4

Вершины v_4 и v_5 — сыновья вершины v_2, которая, в свою очередь, является сыном вершины v_0 — корня дерева. Вершины v_4 и v_5 являются подлинными потомками вершин v_0 и v_2, которые соответственно будут их подлинными предками. Вершины v_1, v_4, v_5, v_6, v_9, v_8 — листья дерева. Взаимно недостижимые вершины ориентированного дерева (v_2 и v_9) не являются ни предком, ни потомком одна другой. Каждая вершина будет сама для себя предком и потомком, но не подлинным.

Определение 12.4. Ориентированное дерево, у которого каждая вершина, отличная от корня, есть лист, называют **кустом**.

Определение 12.5. Подграф неориентированного (ориентированного) дерева, являющийся неориентированным (ориентированным) деревом, называют **поддеревом** исходного дерева.

Компонентами ориентированного дерева являются его **подграфы**, порожденные множеством вершин, расположенных на некотором пути из корня в лист.

Подграф, порожденный множеством вершин $\{v_3,\ v_6,\ v_7,\ v_8,\ v_9\}$, является **поддеревом** ориентированного дерева.

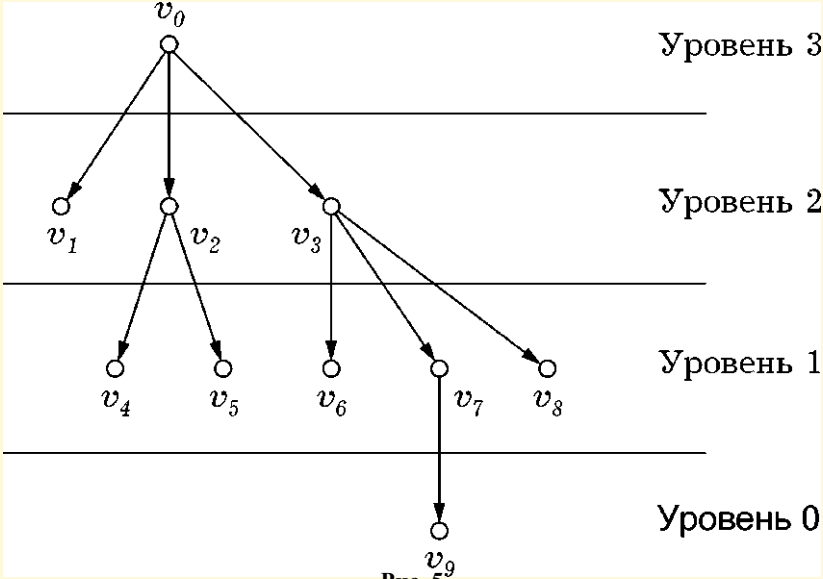

Рис. 5

Определение 12.6. Произвольный ациклический граф называют **неориентированным лесом**.
Если каждая **слабая компонента** ориентированного графа является ориентированным деревом, то такой граф называют **ориентированным лесом**.

Неориентированный лес — это неориентированный граф, каждая компонента которого является неориентированным деревом.

Примеры неориентированного и ориентированного леса:

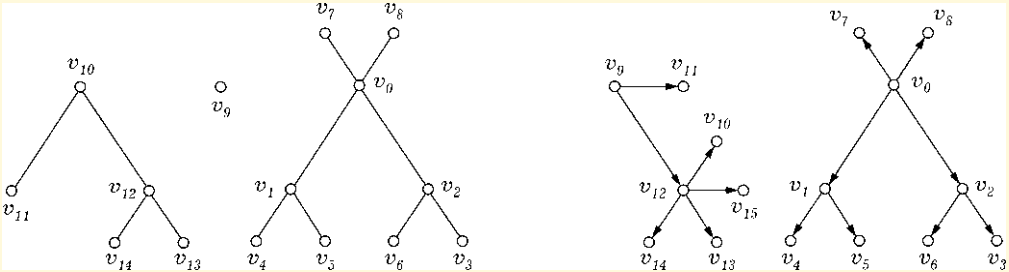

Рис. 6

Остовным лесом (**деревом**) неориентированного (ориентированного) графа называют любой его остовный подграф, являющийся лесом (деревом).

Определение 12.7. Высота ориентированного дерева — это наибольшая длина пути из корня в лист.
Глубина $\mathrm{d}(v)$ **вершины ориентированного дерева** v — это длина пути из корня в эту вершину.
Высота $\mathrm{h}(v)$ **вершины ориентированного дерева** v — это наибольшая длина пути из данной вершины в лист.
Уровень вершины ориентированного дерева — это разность между высотой ориентированного дерева и глубиной данной вершины.

Уровень корня равен высоте ориентированного дерева, но уровни различных листьев,так же как и их глубины, могут быть различными; **высота** любого **листа** равна **нулю**.

Определение 12.8. Ориентированное дерево называют **бинарным**, если полустепень исхода любой его вершины не больше 2.
Бинарное ориентированное дерево называют **полным**, если из любой его вершины, не являющейся листом, исходят ровно две дуги, а уровни всех листьев совпадают.

Теорема 1 (теорема о высоте бинарного ориентированного дерева с заданным числом листьев). Бинарное ориентированное дерево с n листьями имеет высоту, не меньшую $\log_2 n$.

◀ Покажем, что в **полном** бинарном ориентированном дереве высоты h ровно 2^h листьев. Используем метод математической индукции.

Ориентированное дерево высоты 0 имеет $2^0 = 1$ лист.

Полное бинарное ориентированное дерево высоты 1 имеет $2^1 = 2$ листа.

Пусть полное бинарное ориентированное дерево имеет высоту k и соответственно 2^k листьев.

Рассмотрим полное бинарное ориентированное дерево высоты $k + 1$. Поскольку в полном бинарном ориентированном дереве уровни всех листьев совпадают, ориентированное дерево высоты $k + 1$ можно получить из полного бинарного ориентированного дерева высоты k, если из каждого листа последнего провести по две дуги.

Тогда количество листьев в ориентированном дереве высоты $k+1$ будет в 2 раза больше, чем в ориентированном дереве высоты k, т.е. $2^k \cdot 2 = 2^{k+1}$.
В произвольном бинарном дереве листьев может быть только меньше, чем в полном.
Следовательно, в произвольном бинарном дереве высоты h не более 2^h листьев $(n \leq 2^h)$.
Таким образом $h \geq \log_2 n$. ▶

12.4. Задача сортировки:

необходимо расположить строго по возрастанию элементы конечного линейно упорядоченного множества $\{a_1, \ldots, a_n\}$.

Эту задачу называют **задачей сортировки**, а любой алгоритм, ее решающий, — **алгоритмом сортировки**.

Все сравнения, которые могут быть проведены в процессе работы некоторого алгоритма, изображаются наглядно в виде ориентированного дерева, называемого **деревом решений**.

Пример 12.4. Дерево решений для трехэлементного множества $\{a, b, c\}$.

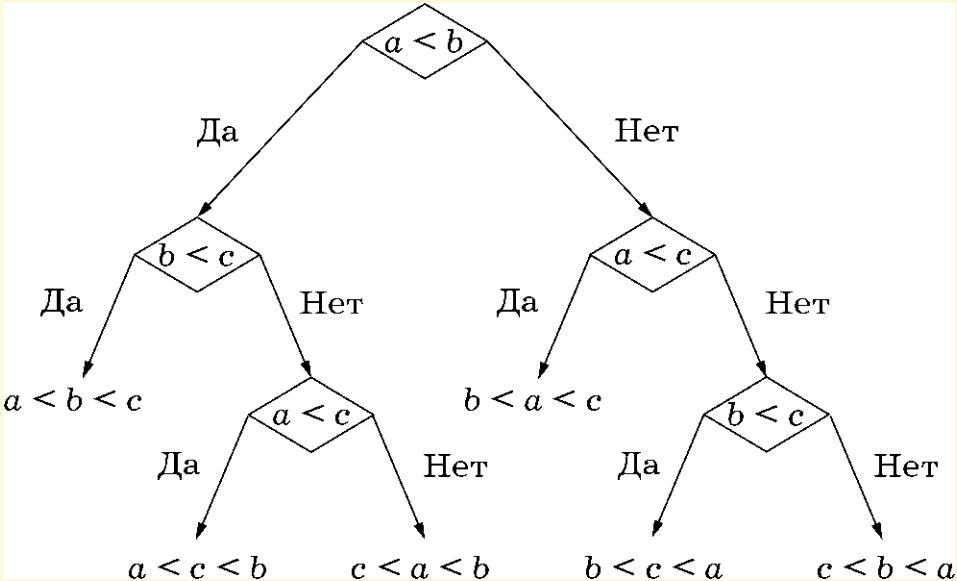

Рис. 7

С математической точки зрения алгоритм сортировки должен найти такую **перестановку** $\{a_{p_1}, \ldots, a_{p_n}\}$ элементов множества, которая была бы согласована с заданным на нем отношением $\leq$ линейного порядка, т.е. для любых k, l из справедливости неравенства $p_k < p_l$ должно следовать $a_{p_k} \leq a_{p_l}$.

Первоначально сортируемые элементы могут быть расположены в произвольном порядке, т.е. исходной может быть любая перестановка элементов сортируемого множества, и мы не имеем никакой априорной информации об этой перестановке.

Поскольку в результате сортировки может получиться любая перестановка исходного множества и каждой такой перестановке соответствует лист дерева решений, в общем случае количество листьев будет равно $n!$ — количеству перестановок n -элементного множества.
Следовательно, сортируя входную последовательность, алгоритм обязательно пройдет какой-то путь от корня дерева решений к одному из листьев, и, таким образом, число операций сравнения (число шагов алгоритма сортировки) будет величиной, пропорциональной высоте дерева решений, не меньшей чем $\log_2 n!$, в силу теоремы 1.

Используя для оценки факториала при больших n **формулу Стирлинга**

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n,$$

получаем, что дерево решений имеет высоту порядка $n \log_2 n$.

Таким образом, в общем случае задачу сортировки с помощью попарных сравнений нельзя решить быстрее, чем за указанное число шагов.

12.5. Методы систематического обхода вершин графа

Необходимо уметь обходить все вершины графа таким образом, чтобы каждая вершина была отмечена ровно один раз.

Обычно такое „путешествие“ по графу сопровождается нумерацией вершин графа в том порядке, в котором они отмечаются, а также определенной „маркировкой“ *ребер* (или *дуг*) графа.

Существуют две основные стратегии таких обходов: **поиск в глубину** и **поиск в ширину**.

Алгоритм поиска в глубину в неор. графе

Граф задан списками смежности, собранными в массив лидеров.

При поиске вершины графа нумеруются в порядке их посещения. Номер вершины v графа, присваиваемый ей при поиске в глубину, обозначим $D[v]$ и будем называть **D -номером**.

В процессе обхода будем находить **фундаментальные циклы** графа.

Пусть в неориентированном графе $G = (V, E)$ произвольно фиксирован **максимальный остовный лес**. Для связного графа это будет максимальное остовное дерево. Множество его ребер обозначим T. Все ребра из T назовем **древесными**, а ребра исходного графа G, не принадлежащие T, — **обратными**.

Любой цикл графа G, содержащий только одно обратное ребро, назовем фундаментальным.

Максимальный остовный лес, находимый с помощью алгоритма поиска в глубину, называют **глубинным остовным лесом**.

Классификация ребер зависит от хода работы алгоритма, который определяется стартовой вершиной и расположением вершин в списках смежности.

Для организации работы алгоритма поиска в глубину используется способ хранения данных, называемый **стеком**. Элементы в стеке упорядочиваются в порядке поступления. В стек можно добавлять новые элементы и из него можно извлекать элементы. При этом доступен только последний добавленный элемент — **вершина стека**.

В алгоритме поиска в глубину используется стек вершин.

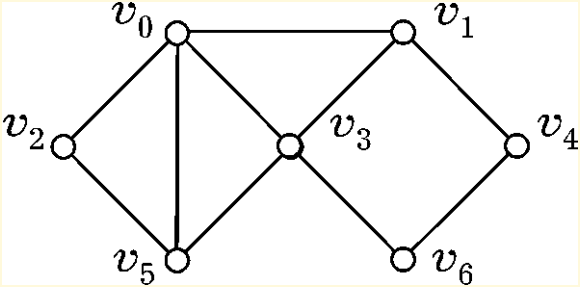

Рис. 8

Списки смежности

$V_0 \to (V_1, V_2, V_3, V_5)$
$V_1 \to (V_0, V_3, V_4)$
$V_2 \to (V_0, V_5)$
$V_3 \to (V_0, V_1, V_5, V_6)$
$V_4 \to (V_1, V_6)$
$V_5 \to (V_0, V_2, V_3)$
$V_6 \to (V_3, V_4)$

Рис. 9

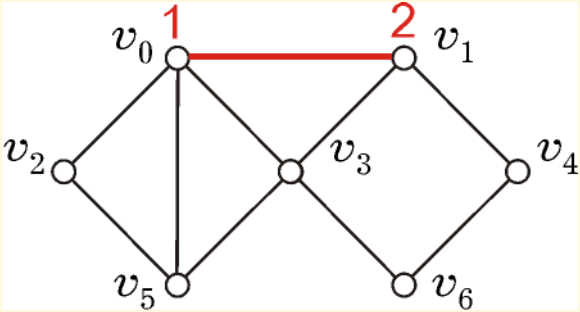

Рис. 10

Списки смежности

$V_0 \to (V_1, V_2, V_3, V_5)$
$V_1 \to (V_0, V_3, V_4)$
$V_2 \to (V_0, V_5)$
$V_3 \to (V_0, V_1, V_5, V_6)$
$V_4 \to (V_1, V_6)$
$V_5 \to (V_0, V_2, V_3)$
$V_6 \to (V_3, V_4)$

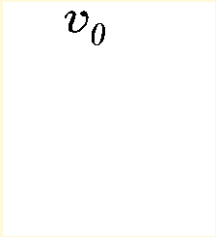

Рис. 11

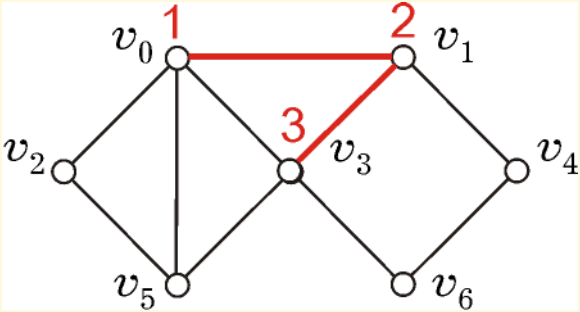

Рис. 12

Списки смежности

$V_0 \to (\not{V}_1, V_2, V_3, V_5)$
$V_1 \to (\not{V}_0, \not{V}_3, V_4)$
$V_2 \to (V_0, V_5)$
$V_3 \to (V_0, V_1, V_5, V_6)$
$V_4 \to (V_1, V_6)$
$V_5 \to (V_0, V_2, V_3)$
$V_6 \to (V_3, V_4)$

v_0
v_1

Рис. 13

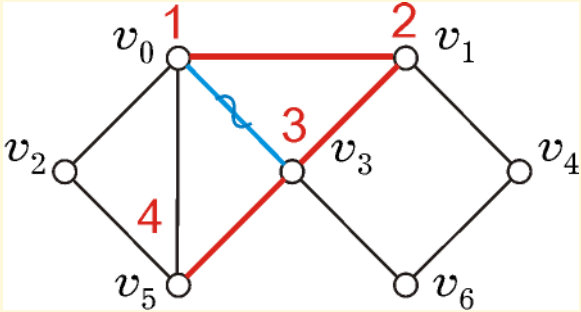

Рис. 14

Списки смежности

$V_0 \to (\not{V}_1, V_2, \not{V}_3, V_5)$
$V_1 \to (\not{V}_0, \not{V}_3, V_4)$
$V_2 \to (V_0, V_5)$
$V_3 \to (\not{V}_0, \not{V}_1, \not{V}_5, V_6)$
$V_4 \to (V_1, V_6)$
$V_5 \to (V_0, V_2, V_3)$
$V_6 \to (V_3, V_4)$

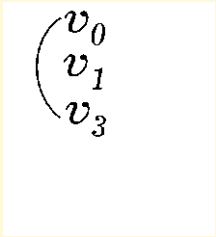

Рис. 15

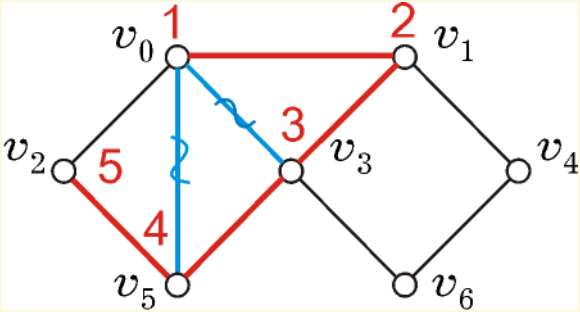

Рис. 16

Списки смежности

$V_0 \to (\not{N}_1, \not{N}_2, \not{N}_3, \not{N}_5)$
$V_1 \to (\not{N}_0, \not{N}_3, V_4)$
$V_2 \to (V_0, V_5)$
$V_3 \to (\not{N}_0, \not{N}_1, \not{N}_5, V_6)$
$V_4 \to (V_1, V_6)$
$V_5 \to (\not{N}_0, \not{N}_2, \not{N}_3)$
$V_6 \to (V_3, V_4)$

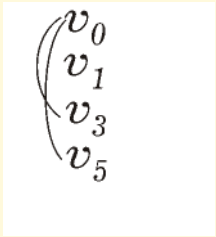

Рис. 17

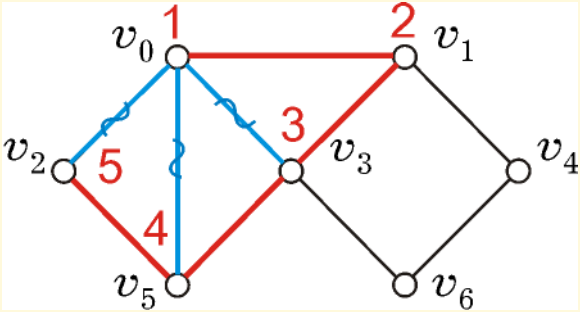

Рис. 18

Списки смежности

$V_0 \to (\not{V}_1, \not{V}_2, \not{V}_3, \not{V}_5)$
$V_1 \to (\not{V}_0, \not{V}_3, V_4)$
$V_2 \to (\not{V}_0, \not{V}_5)$
$V_3 \to (\not{V}_0, \not{V}_1, \not{V}_5, V_6)$
$V_4 \to (V_1, V_6)$
$V_5 \to (\not{V}_0, \not{V}_2, \not{V}_3)$
$V_6 \to (V_3, V_4)$

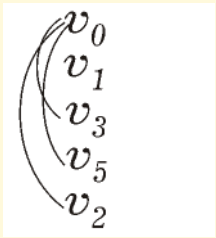

Рис. 19

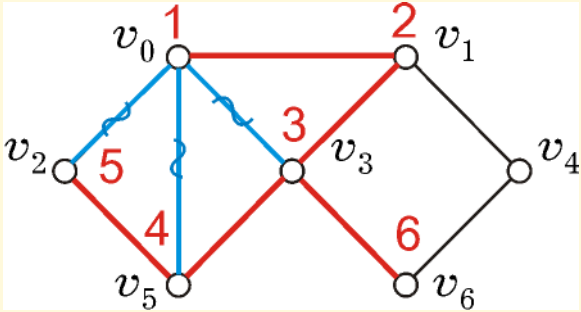

Рис. 20

Списки смежности

$V_0 \to (\not{V}_1, \not{V}_2, \not{V}_3, \not{V}_5)$
$V_1 \to (\not{V}_0, \not{V}_3, V_4)$
$V_2 \to (\not{V}_0, \not{V}_5)$
$V_3 \to (\not{V}_0, \not{V}_1, \not{V}_5, \not{V}_6)$
$V_4 \to (V_1, V_6)$
$V_5 \to (\not{V}_0, \not{V}_2, \not{V}_3)$
$V_6 \to (V_3, V_4)$

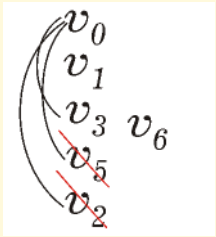

Рис. 21

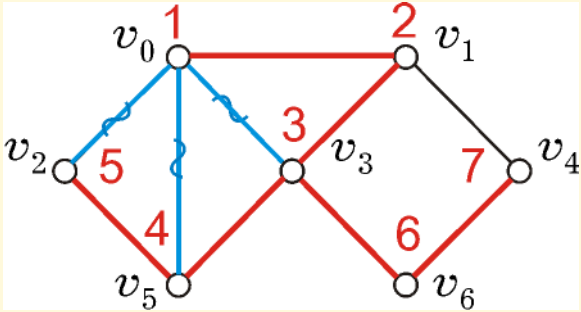

Рис. 22

Списки смежности

$V_0 \to (\not{V}_1, \not{V}_2, \not{V}_3, \not{V}_5)$
$V_1 \to (\not{V}_0, \not{V}_3, V_4)$
$V_2 \to (\not{V}_0, \not{V}_5)$
$V_3 \to (\not{V}_0, \not{V}_1, \not{V}_5, \not{V}_6)$
$V_4 \to (V_1, V_6)$
$V_5 \to (\not{V}_0, \not{V}_2, \not{V}_3)$
$V_6 \to (\not{V}_3, \not{V}_4)$

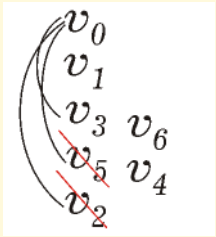

Рис. 23

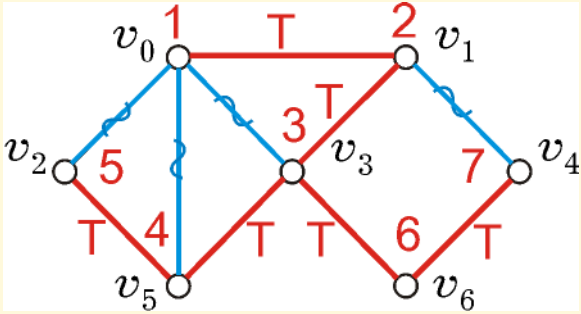

Рис. 24

Списки смежности

$V_0 \to (\not{N}_1, \not{N}_2, \not{N}_3, \not{N}_5)$
$V_1 \to (\not{N}_0, \not{N}_3, \not{N}_4)$
$V_2 \to (\not{N}_0, \not{N}_5)$
$V_3 \to (\not{N}_0, \not{N}_1, \not{N}_5, \not{N}_6)$
$V_4 \to (\not{N}_1, \not{N}_6)$
$V_5 \to (\not{N}_0, \not{N}_2, \not{N}_3)$
$V_6 \to (\not{N}_3, \not{N}_4)$

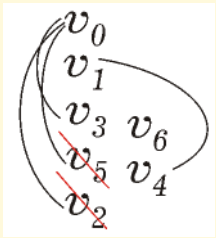

Рис. 25

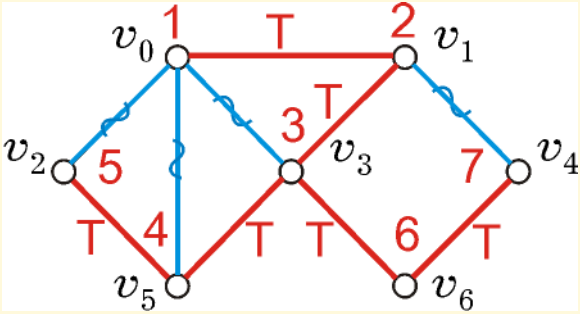

Рис. 26

Списки смежности

$V_0 \to (\not{V}_1, \not{V}_2, \not{V}_3, \not{V}_5)$
$V_1 \to (\not{V}_0, \not{V}_3, \not{V}_4)$
$V_2 \to (\not{V}_0, \not{V}_5)$
$V_3 \to (\not{V}_0, \not{V}_1, \not{V}_5, \not{V}_6)$
$V_4 \to (\not{V}_1, \not{V}_6)$
$V_5 \to (\not{V}_0, \not{V}_2, \not{V}_3)$
$V_6 \to (\not{V}_3, \not{V}_4)$

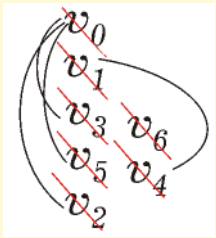

Рис. 27

Алгоритм поиска в глубину в ор. графе

В ориентированном графе вершинам также присваиваются D -номера. Классификация дуг:

1) **древесные дуги** — каждая такая *дуга ведет* от *отца* к *сыну* в глубинном остовном лесу;

2) **прямые дуги** — каждая такая дуга ведет от *подлинного предка* к *подлинному потомку* (но не от отца к сыну) в глубинном остовном лесу;

3) **обратные дуги** — от *потомков* к *предкам* (включая все петли);

4) **поперечные дуги** — все дуги, не являющиеся ни древесными, ни прямыми, ни обратными.

Результат работы алгоритма: множества $Tree$ — древесных дуг, $Back$ — обратных дуг, $Forward$ — прямых дуг, C — поперечных дуг и массив D , содержащий D -номера вершин.

Особенности алгоритма.

Так, если очередная вершина w, извлеченная из списка смежности текущей вершины v, новая, то дуга (v, w) является древесной.

Если вершина w не новая $(w \notin V_0)$, то дуга (v, w) будет либо прямой, либо обратной, либо поперечной.

Если D-номер вершины v строго меньше D-номера вершины w $(D[v] < D[w])$, то дуга (v, w) является прямой.

Если D-номер вершины v не меньше D-номера вершины w $(D[v] \geq D[w])$, необходимо проверить, есть ли в стеке $STACK$ вершина w. Если вершина w находится в стеке, то дуга (v, w) является обратной.Если вершины w в стеке нет, то дуга является поперечной.

Если стек пуст, но не все вершины ориентированного графа обработаны, поиск продолжают из любой необработанной вершины.

В случае ориентированного графа поиск контуров на базе поиска в глубину существенно сложнее.

Ориентированный граф является бесконтурным тогда и только тогда, когда при поиске в глубину от некоторой начальной вершины множество обратных дуг оказывается пустым.

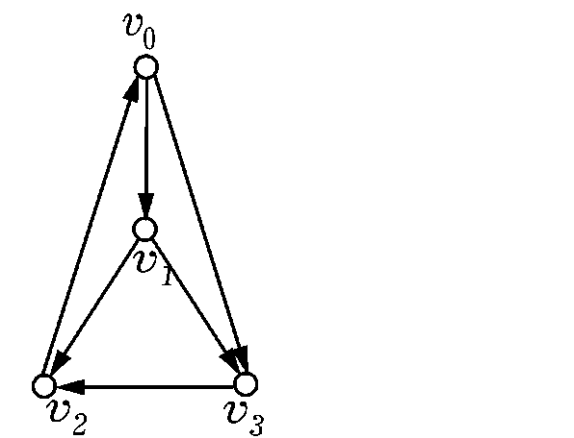

Рис. 28

Списки смежности

$V_0 \to (V_1, V_3)$
$V_1 \to (V_2, V_3)$
$V_2 \to (V_0)$
$V_3 \to (V_2)$

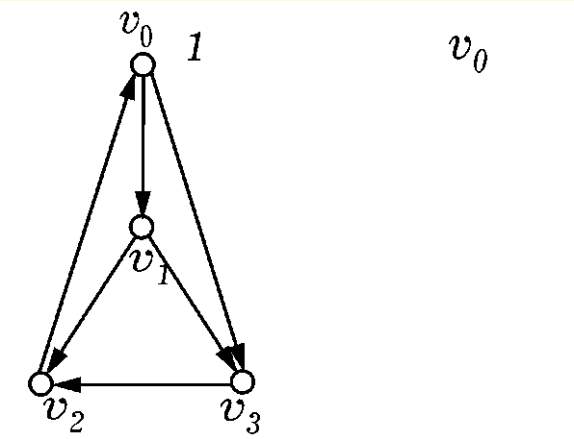

Рис. 29

Списки смежности

$V_0 \to (V_1, V_3)$
$V_1 \to (V_2, V_3)$
$V_2 \to (V_0)$
$V_3 \to (V_2)$

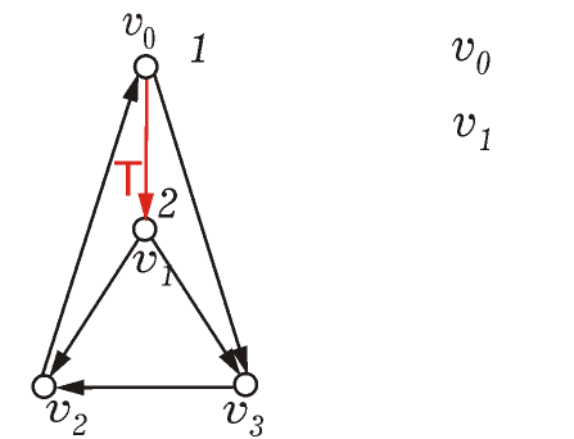

Рис. 30

Списки смежности

$V_0 \to (V_1, V_3)$
$V_1 \to (V_2, V_3)$
$V_2 \to (V_0)$
$V_3 \to (V_2)$

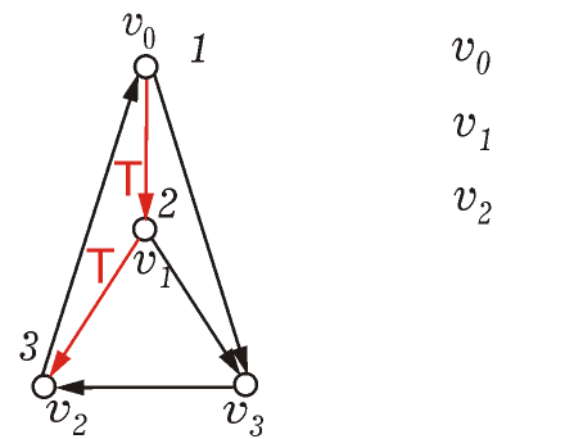

Рис. 31

Списки смежности

$V_0 \to (\not{N}_1, V_3)$
$V_1 \to (\not{N}_2, V_3)$
$V_2 \to (V_0)$
$V_3 \to (V_2)$

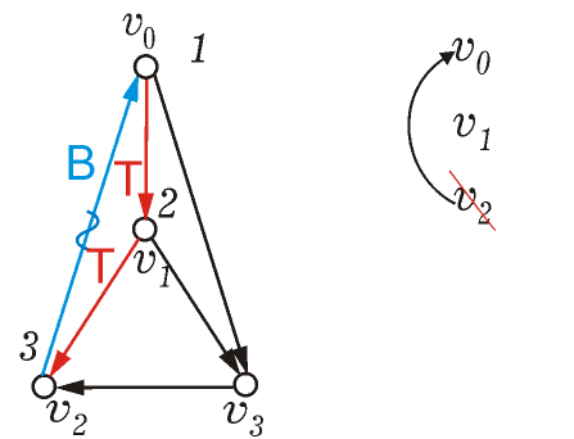

Рис. 32

Списки смежности

$V_0 \to (\not{N}_1, V_3)$
$V_1 \to (\not{N}_2, V_3)$
$V_2 \to (\not{N}_0)$
$V_3 \to (V_2)$

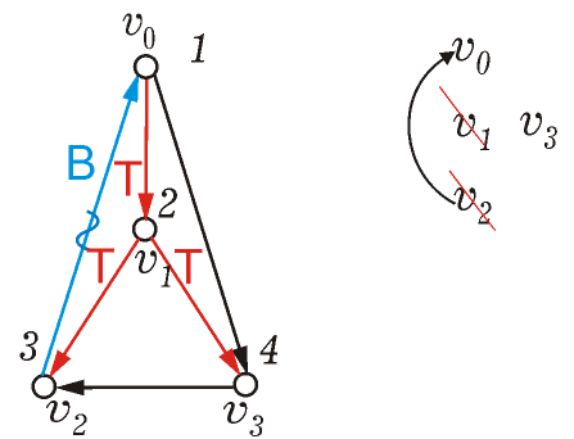

Рис. 33

Списки смежности

$V_0 \to (\not{V}_1, V_3)$
$V_1 \to (\not{V}_2, \not{V}_3)$
$V_2 \to (\not{V}_0)$
$V_3 \to (V_2)$

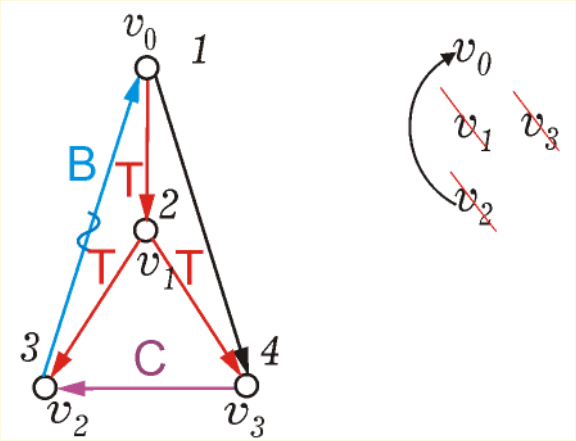

Рис. 34

Списки смежности

$V_0 \to (\not{V}_1, V_3)$
$V_1 \to (\not{V}_2, \not{V}_3)$
$V_2 \to (\not{V}_0)$
$V_3 \to (\not{V}_2)$

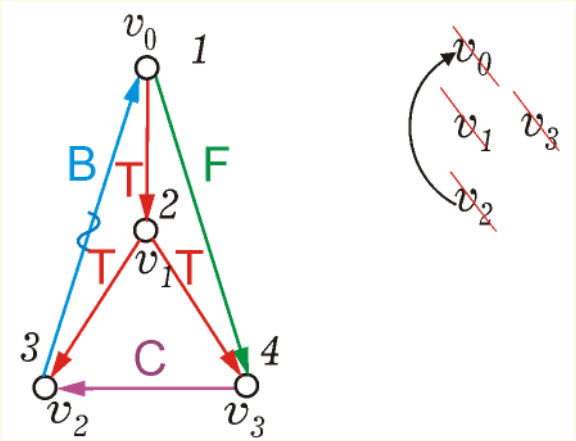

Рис. 35

Списки смежности

$V_0 \to (\not{N}_1, \not{N}_3)$
$V_1 \to (\not{N}_2, \not{N}_3)$
$V_2 \to (\not{N}_0)$
$V_3 \to (\not{N}_2)$

12.6. Алгоритм поиска в ширину в ор. графе

Вход. Граф $G = (V, E)$, заданный списками смежности; v_0 — начальная вершина (не обязательно первый элемент массива лидеров).

Выход. Массив M меток вершин, где каждая метка равна длине пути от v_0 до v.

0. Очередь Q положить пустой $(Q := \varnothing)$. Все вершины пометить как недостижимые из вершины v_0, присваивая элементам массива M значение $+\infty$ $(M[v_i] := +\infty,\ i = \overline{1,\ N})$.
Стартовую вершину v_0 пометить номером 0, т.е. длину пути от стартовой вершины v_0 до самой себя положить равной 0 ($M[v_0] := 0$). Поместить вершину v_0 в очередь Q. Перейти на шаг **1**.
1. Если очередь Q не пуста $(Q \neq \varnothing)$, то из „головы“ очереди извлечь (с удалением из очереди) вершину u и перейти на шаг **2**. Если очередь пуста, перейти на шаг **3**.

2. Если список смежности $L(u)$ вершины u пуст, вернуться на шаг **1**.
Если список смежности $L(u)$ вершины u не пуст, для каждой вершины w из списка смежности, где $M[w] = +\infty$, т.е. вершины, которую еще не посещали, положить длину пути из стартовой вершины v_0 до вершины w равной длине пути от v_0 до вершины u плюс одна дуга $(M[w] := M[u] + 1)$, т.е. отметить вершину w и поместить ее в очередь Q. После просмотра всех вершин списка смежности $L(u)$ вернуться на шаг **1**.
3. Распечатать массив M. Закончить работу.

Алгоритм поиска в ширину может быть дополнен процедурой „обратного хода“, определяющей номера вершин, лежащих на кратчайшем пути из вершины v_0 в данную вершину u.

Для этого необходимо завести массив PR размера $|V|$, каждый элемент $PR[w]$ которого содержит номер той вершины, из которой был осуществлен переход в вершину w при ее пометке.

Если вершина w находится в списке смежности $L(u)$ вершины u, заполнение элемента массива $PR[w]$ происходит при изменении метки вершины w $M[w]$ с $+\infty$ на единицу. При этом в элементе $PR[w]$ сохраняется номер вершины u $(PR[w] := u)$. Для начальной вершины $PR[v_0]$ можно положить равным 0, в предположении, что начальная вершина v_0 имеет номер 0 и остальные вершины пронумерованы от 1 до N.

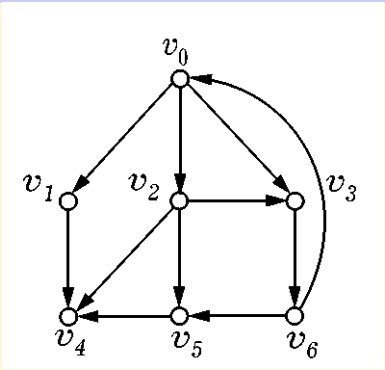

Рис. 36

Списки смежности

$v_0 \longrightarrow (v_1, v_2, v_3)$
$v_1 \longrightarrow (v_4)$
$v_2 \longrightarrow (v_4, v_5, v_3)$
$v_3 \longrightarrow (v_6)$
$v_4 \longrightarrow ()$
$v_5 \longrightarrow (v_4)$
$v_6 \longrightarrow (v_5, v_0)$

	v_0	v_1	v_2	v_3	v_4	v_5	v_6
1.	0	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$

$Q = (v_0)$
$PR(v_0) = \varnothing$

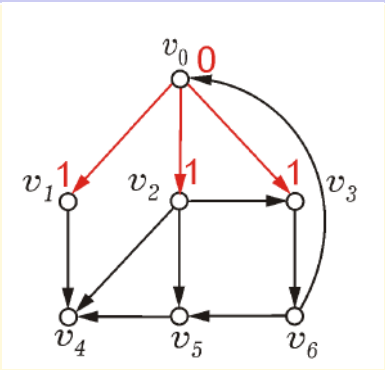

Рис. 37

Списки смежности

$v_0 \longrightarrow (v_1, v_2, v_3)$
$v_1 \longrightarrow (v_4)$
$v_2 \longrightarrow (v_4, v_5, v_3)$
$v_3 \longrightarrow (v_6)$
$v_4 \longrightarrow ()$
$v_5 \longrightarrow (v_4)$
$v_6 \longrightarrow (v_5, v_0)$

	v_0	v_1	v_2	v_3	v_4	v_5	v_6
1.	0	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$
2.	0	1	1	1	$+\infty$	$+\infty$	$+\infty$

$Q = (\not{v}_0, v_1, v_2, v_3)$
$PR(v_0) = \varnothing, PR(v_1) = v_0, PR(v_2) = v_0,$
$PR(v_3) = v_0$

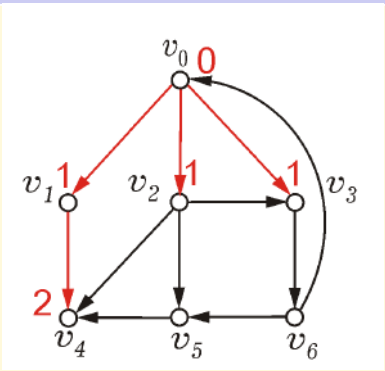

Рис. 38

Списки смежности

$v_0 \longrightarrow (v_1, v_2, v_3)$
$v_1 \longrightarrow (v_4)$
$v_2 \longrightarrow (v_4, v_5, v_3)$
$v_3 \longrightarrow (v_6)$
$v_4 \longrightarrow ()$
$v_5 \longrightarrow (v_4)$
$v_6 \longrightarrow (v_5, v_0)$

	v_0	v_1	v_2	v_3	v_4	v_5	v_6
1.	0	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$
2.	0	1	1	1	$+\infty$	$+\infty$	$+\infty$
3.	0	1	1	1	2	$+\infty$	$+\infty$

$Q = (\not{v}_0\ \not{v}_1, v_2, v_3, v_4)$
$PR(v_0) = \varnothing, PR(v_1) = v_0, PR(v_2) = v_0,$
$PR(v_3) = v_0, PR(v_4) = v_1$

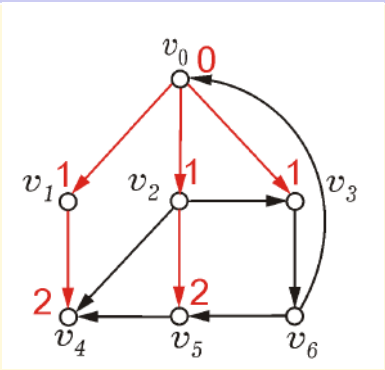

Рис. 39

Списки смежности

$v_0 \longrightarrow (v_1, v_2, v_3)$
$v_1 \longrightarrow (v_4)$
$v_2 \longrightarrow (v_4, v_5, v_3)$
$v_3 \longrightarrow (v_6)$
$v_4 \longrightarrow ()$
$v_5 \longrightarrow (v_4)$
$v_6 \longrightarrow (v_5, v_0)$

	v_0	v_1	v_2	v_3	v_4	v_5	v_6
1.	0	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$
2.	0	1	1	1	$+\infty$	$+\infty$	$+\infty$
3.	0	1	1	1	2	$+\infty$	$+\infty$
4.	0	1	1	1	2	2	$+\infty$

$Q = (\not{v}_0\ \not{v}_1\ \not{v}_2, v_3, v_4, v_5)$
$PR(v_0) = \varnothing, PR(v_1) = v_0, PR(v_2) = v_0,$
$PR(v_3) = v_0, PR(v_4) = v_1, PR(v_5) = v_2$

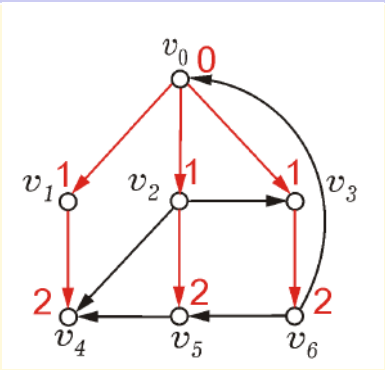

Рис. 40

Списки смежности

$v_0 \longrightarrow (v_1, v_2, v_3)$

$v_1 \longrightarrow (v_4)$

$v_2 \longrightarrow (v_4, v_5, v_3)$

$v_3 \longrightarrow (v_6)$

$v_4 \longrightarrow ()$

$v_5 \longrightarrow (v_4)$

$v_6 \longrightarrow (v_5, v_0)$

	v_0	v_1	v_2	v_3	v_4	v_5	v_6
1.	0	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$
2.	0	1	1	1	$+\infty$	$+\infty$	$+\infty$
3.	0	1	1	1	2	$+\infty$	$+\infty$
4.	0	1	1	1	2	2	$+\infty$
5.	0	1	1	1	2	2	2

$Q = (\not{v}_0\ \not{v}_1\ \not{v}_2\ \not{v}_3, v_4, v_5, v_6)$

$PR(v_0) = \varnothing, PR(v_1) = v_0, PR(v_2) = v_0,$
$PR(v_3) = v_0, PR(v_4) = v_1, PR(v_5) = v_2,$
$PR(v_6) = v_3$

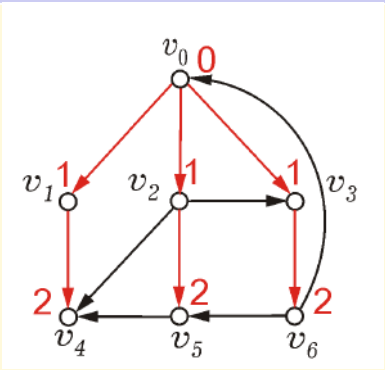

Рис. 41

Списки смежности

$v_0 \longrightarrow (v_1, v_2, v_3)$

$v_1 \longrightarrow (v_4)$

$v_2 \longrightarrow (v_4, v_5, v_3)$

$v_3 \longrightarrow (v_6)$

$v_4 \longrightarrow ()$

$v_5 \longrightarrow (v_4)$

$v_6 \longrightarrow (v_5, v_0)$

	v_0	v_1	v_2	v_3	v_4	v_5	v_6
1.	0	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$
2.	0	1	1	1	$+\infty$	$+\infty$	$+\infty$
3.	0	1	1	1	2	$+\infty$	$+\infty$
4.	0	1	1	1	2	2	$+\infty$
5.	0	1	1	1	2	2	2

$Q = (\not{v_0}\ \not{v_1}\ \not{v_2}\ \not{v_3}\ \not{v_4}\ \not{v_5}\ \not{v_6})$

$PR(v_0) = \varnothing, PR(v_1) = v_0, PR(v_2) = v_0,$
$PR(v_3) = v_0, PR(v_4) = v_1, PR(v_5) = v_2,$
$PR(v_6) = v_3$

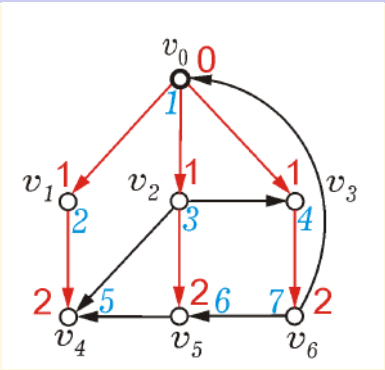

Рис. 42

Списки смежности

$v_0 \longrightarrow (v_1, v_2, v_3)$
$v_1 \longrightarrow (v_4)$
$v_2 \longrightarrow (v_4, v_5, v_3)$
$v_3 \longrightarrow (v_6)$
$v_4 \longrightarrow ()$
$v_5 \longrightarrow (v_4)$
$v_6 \longrightarrow (v_5, v_0)$

	v_0	v_1	v_2	v_3	v_4	v_5	v_6
1.	0	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$
2.	0	1	1	1	$+\infty$	$+\infty$	$+\infty$
3.	0	1	1	1	2	$+\infty$	$+\infty$
4.	0	1	1	1	2	2	$+\infty$
5.	0	1	1	1	2	2	2

$Q = \varnothing$
$PR(v_0) = \varnothing, PR(v_1) = v_0, PR(v_2) = v_0,$
$PR(v_3) = v_0, PR(v_4) = v_1, PR(v_5) = v_2,$
$PR(v_6) = v_3$

Материал для самостоятельного изучения

Матрица достижимости.

Это квадратная матрица C порядка $|V|$, каждый элемент c_{ij} которой равен 1, если j-я *вершина достижима* из i-й вершины, и равен 0 если иначе.

Согласно определению достижимости, элементы $c_{ii} = 1$. Матрица достижимости несет важную информацию об ориентированном графе. Ее анализ позволяет, например, найти его **бикомпоненты**.

Пример 12.5.

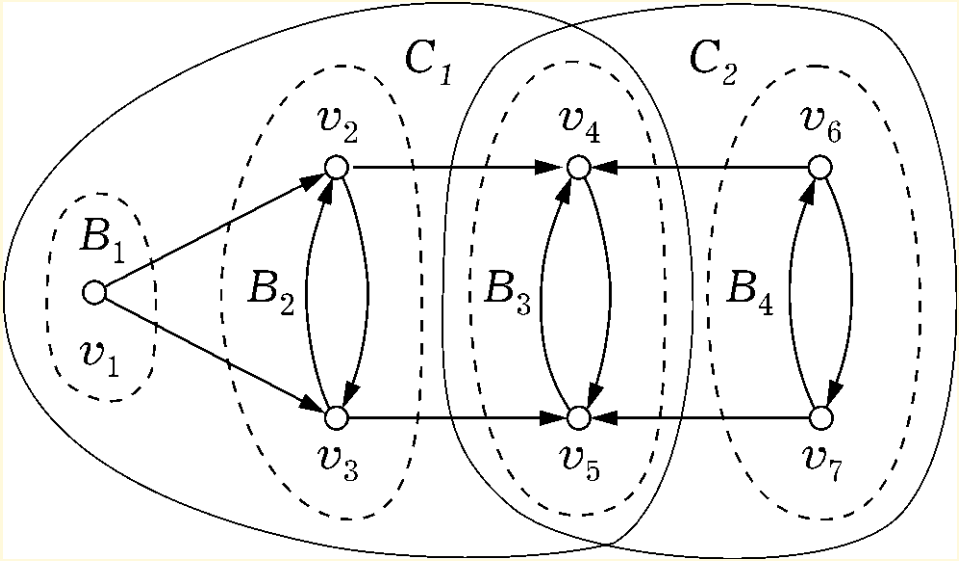

Рис. 43

Матрица достижимости:

$$C = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

Бикомпоненты.
Для первой вершины множество $P_1 = \{1\}$ включает только ее саму.
Для второй вершины — $P_2 = \{2,\, 3\}$,
для четвертой — $P_4 = \{4,\, 5\}$, для шестой — $P_6 = \{6,\, 7\}$.
Получили B_1, B_2, B_3, B_4.

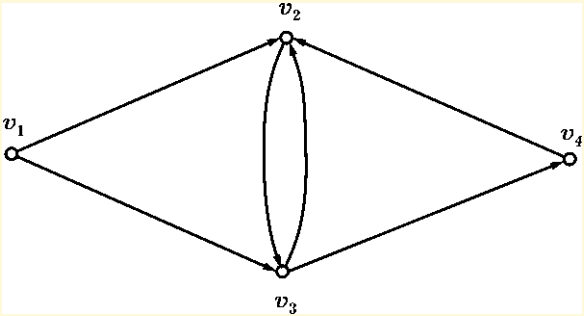

Рис. 44

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix}$$

Найдем бикомпоненты графа изображенного на рисунке. Для первой вершины множество $P_1 = \{1\}$ включает только ее саму. Для второй вершины имеем $P_2 = \{2, 3, 4\}$. Соответственно полученные множества вершин порождают бикомпоненты B_1, B_2.

12.7. Остовное дерево наименьшего веса

Неориентированный (ориентированный) граф, у которого каждому ребру (дуге) сопоставлено некоторое действительное число, называют **взвешенным** или **размеченным** графом.

Это число называют **весом** или **меткой** ребра (дуги).

Алгоритм Краскала вычисляет для заданного взвешенного неориентированного графа G *остовное дерево* с наименьшей суммой весов ребер — **остовное дерево наименьшего веса**.

При описании алгоритма будем использовать способ хранения данных, называемый **очередью**.

Элементы данных в очереди упорядочиваются по времени поступления. Элементы можно добавлять в очередь и извлекать из очереди.

В каждый момент времени доступен только один элемент, который был помещен в очередь раньше других, — „голова“ очереди.

При добавлении новый элемент помещается в „хвост“ очереди, т.е. работа ведется по обычному для очереди правилу — „первым пришел — первым вышел“.

Чтобы извлечь из очереди некоторый элемент, не доступный в текущий момент, надо извлечь все ранее поступившие элементы, начиная с „головы“ очереди.

Обычно очередь реализуется в виде списка.

Алгоритм Краскала

Рассмотрим алгоритм нахождения остовного дерева наименьшего веса. Пусть дан связный неориентированный граф $G = (V, E)$ с числовыми неотрицательными весами ребер. Вес ребра e обозначим $\varphi(e)$.

В результате работы алгоритма получим остовное дерево $T = (V, H)$ графа G, такое, что сумма $\sum_{e \in H} \varphi(e)$ является наименьшей.

Отсортируем все ребра исходного графа по возрастанию весов и сформируем из них очередь так, чтобы в „голове“ очереди находилось ребро с наименьшим весом, а в „хвосте“ — с наибольшим и веса ребер не убывали от „головы“ очереди к „хвосту“.

Метод состоит в „сшивании“ искомого дерева из *компонент* остовного леса. Первоначально *остовный лес* представляет собой множество изолированных вершин исходного графа, т.е. его множество ребер пусто. На первом шаге из очереди извлекается ребро наименьшего веса и добавляется к множеству ребер исходного дерева.

На последующих шагах алгоритма из очереди извлекается по одному ребру. Если это ребро соединяет вершины, принадлежащие разным компонентам текущего остовного леса, то оно добавляется к текущему множеству ребер искомого дерева, а указанные компоненты сливаются в одну. Иначе ребро отбрасывается. Процесс повторяется до тех пор, пока число компонент остовного леса не окажется равным 1. Можно показать, что эта компонента и будет искомым остовным деревом наименьшего веса.

1. Множество ребер H искомого остовного дерева полагаем пустым ($H = \varnothing$).
2. Формируем множество $V_S = \{\{v_1\}, \ldots, \{v_n\}\}$, элементами которого являются множества вершин, соответствующих компонентам исходного остовного леса. Каждая такая компонента состоит из единственной вершины.
3. Сортируем множество ребер E исходного графа по возрастанию весов и формируем очередь Q, элементами которой являются ребра графа G.
4. Если множество V_S содержит более одного элемента (т.е. остовный лес состоит из нескольких компонент) и очередь Q не пуста, переходим на шаг **5**, если иначе — на шаг **7**.

5. Извлекаем из очереди Q ребро e. Если *концы* ребра e принадлежат различным множествам вершин V_i и V_j из V_S, то переходим на шаг **6**, если иначе, то отбрасываем извлеченное ребро и возвращаемся на шаг **4**.

6. Объединяем множества вершин V_i и V_j (полагая $W = V_i \cup V_j$), удаляем множества V_i и V_j из множества V_S и добавляем в V_S множество W. Добавляем ребро e в множество H. Возвращаемся на шаг **4**.

7. Прекращаем работу. Множество H есть множество ребер полученного остовного дерева.

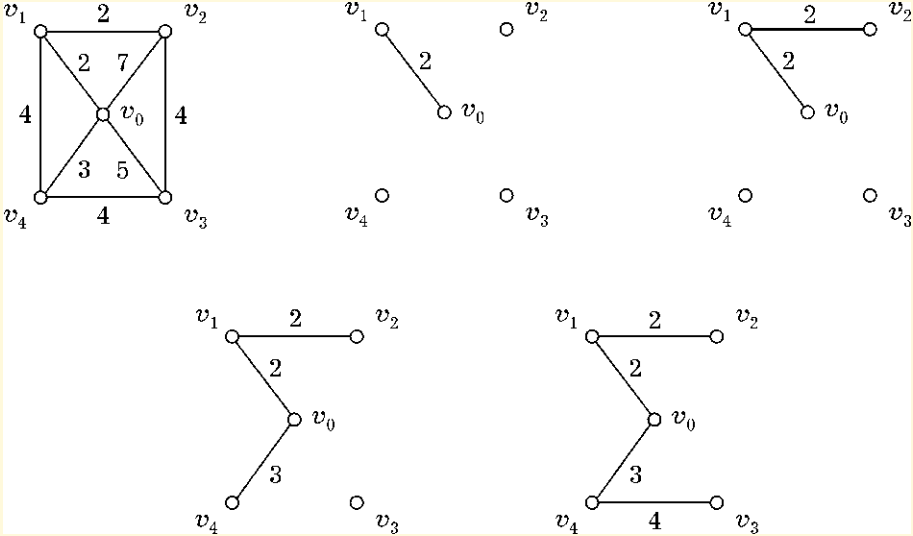
v_1
2
v_2
2
7
4
v_0
4
3
5
v_4
4
v_3
v_1
v_2
2
v_0
v_4
v_3
v_1
2
v_2
2
v_0
v_4
v_3
v_1
2
v_2
2
v_0
3
v_4
v_3
v_1
2
v_2
2
v_0
3
v_4
4
v_3

Рис. 45

Исходный граф изображен на рисунке *а*. На рисунке *б* проиллюстрирован результат выполнения первого шага алгоритма.

На *в* показан результат добавления следующего ребра $\{v_1, v_2\}$ с весом 2 из очереди. На *г* приведен результат добавления ребра $\{v_0, v_4\}$ с весом 3.

Если следующим в очереди ребром будет $\{v_1, v_4\}$, оно будет отброшено.

Дальнейший ход работы алгоритма зависит от того, в каком порядке в очереди размещены ребра $\{v_2,\, v_3\}$ и $\{v_3,\, v_4\}$ с весами 4. Любое из них может быть добавлено в множество ребер остовного дерева, и на этом алгоритм закончит работу. На *д* приведено остовное дерево, полученное после добавления ребра $\{v_3,\, v_4\}$.

Для приведенного графа оба ребра с весом 2 войдут в остовное дерево независимо от порядка их расположения в очереди после сортировки, а ребро $\{v_1,\, v_4\}$ не войдет ни в какое остовное дерево наименьшего веса.

ДИСКРЕТНАЯ МАТЕМАТИКА

ИУ5 - 4 семестр, 2015 г.

Лекция 13. РЕГУЛЯРНЫЕ ЯЗЫКИ И КОНЕЧНЫЕ АВТОМАТЫ

Формальный язык.

Термин „формальный язык“ обычно означает искусственный язык, придуманный людьми для специальных целей, например — язык программирования.

Непреодолимой преграды между специально придуманными искусственными (формальными) языками и стихийно возникающими и развивающимися естественными языками нет.

Естественные языки характеризуются сложными грамматическими правилами, т.е. довольно жестко формализованы, а специально разработанный язык программирования содержит „темные места“, однозначное понимание которых является проблемой.

Основные аспекты изучения языков.

1. Синтаксис языка.

Слово „синтаксис“ происходит от древнегреческих слов „syn“ — „вместе“ и „taxis“ — „порядок, строй“.

Язык — это какое-то множество „слов“, где „слово“ есть определенная конечная последовательность „букв“ — символов какого-то заранее фиксированного алфавита.

Термины „буква“ и „слово“ могут пониматься по-разному (математическое определение этих терминов будет дано позже).

„Буквами“ могут быть действительно буквы алфавита какого-нибудь естественного или формального языка, например русского языка или языка программирования „Паскаль“.
„Словами“ будут конечные последовательности „букв“.
Но „буквой“ может быть и некоторое „слово“ целиком. Например: „if“ „then“ „else“ Тогда „слова“ — это предложения языка программирования: „if a then b else c.
Если фиксировано какое-то множество „букв“, то не каждая их последовательность будет „словом“, а только такая последовательность, которая подчиняется определенным правилам.

Синтаксис языка и представляет собой систему правил, в соответствии с которыми можно строить „правильные“ последовательности „букв“. Каждое слово языка характеризуется определенной структурой, специфичной именно для данного языка.

Необходимо разработать механизмы порождения слов с заданной структурой и механизмы проверки того, что данное слово принадлежит данному языку.

Именно эти механизмы и изучает классическая теория формальных языков.

2. Семантика языка.

Семантика („semanticos“ — „обозначающий“) предполагает сопоставление словам языка некоего „смысла“, „значения“. Например, записывая математическую формулу, мы должны соблюдать определенные синтаксические правила (расстановка скобок, правописание символов, порядок символов и т.п.), но, кроме этого, формула имеет вполне определенный смысл, что-то обозначает.

Обсуждение семантики выходит за рамки курса.

13.1. Алфавит, слово, язык

Алфавит — это произвольное *непустое конечное множество*
$V = \{a_1, \ldots, a_n\}$, *элементы* которого называют **буквами** или **символами**.

Определение 13.1. **Словом** или **цепочкой** в алфавите V называют произвольный кортеж из множества V^k (k-й *декартовой степени* алфавита V) для различных
$k = 0, 1, 2, \ldots$

Например, если $V = \{a, b, c\}$, то
(a), (b), (c), (a, b), (a, b, c), (c, b, a, a, c) и т.д. есть слова в алфавите V.

При $k = 0$ получаем **пустой кортеж**, называемый **пустым словом** или **пустой цепочкой** и обозначаемый λ.
Пустое слово — это слово, не имеющее символов.
Слова будем записывать без скобок и запятых. Такая запись согласуется с пониманием слова как цепочки следующих друг за другом символов.
Например: a, b, c, ab, abc, $cbaac$.
Длина слова w — число компонент кортежа, т.е. если $w \in V^r$, то длина слова w равна r. Длину слова можно понимать как число составляющих это слово букв.
Длину слова w обозначим $|w|$.
Для пустого слова $|\lambda| = 0$.

Будем использовать следующую запись непустого слова x в алфавите V по буквам:

$$x = x(1)x(2)\dots x(k),$$

где $x(i)$, $1 \le i \le k$, — i-я буква слова x.
Множество всех слов в алфавите V обозначают V^*, а множество всех непустых слов в V — как V^+.

Определение 13.2. Языком в алфавите V называется произвольное подмножество множества V^*.

Определим на множестве 2^{V^*} всех языков в произвольном (но фиксированном!) алфавите V **алгебраическую структуру**.

Языки — это множества, следовательно, над ними можно производить все теоретико-множественные операции: **объединение, пересечение, разность, дополнение** и т.п. **Универсальное множество** в данном случае есть множество слов V^*, которое называют **универсальным языком**.

Специальные операции над языками.

Определение 13.3. Соединением или **конкатенацией** слов $x = x(1)x(2)\dots x(k)$ и $y = y(1)y(2)\dots y(m)$ называют слово

$$xy = x(1)x(2)\dots x(k)y(1)y(2)\dots y(m).$$

Соединение иногда обозначают точкой ($\cdot$).
Из определения следует, что соединение слов **ассоциативно**, т.е. для произвольных трех слов x, y, z имеет место $x(yz) = (xy)z$,
и $x\lambda = \lambda x = x$ для любого x.

Множество V^* всех слов в алфавите V с операцией соединения образует **моноид** $(V^*, \cdot, \lambda)$. **Единица моноида** — пустое слово. Этот моноид есть **свободный моноид**, порожденный алфавитом V. Для него используют то же обозначение, что и для множества всех слов в алфавите V, т.е. V^*.

Неформально соединение xy получается приписыванием слова y справа к слову x. Для любых двух слов $x \in V^k$ и $y \in V^m$ конкатенация $xy \in V^{k+m}$ ($k, m \geq 0$). Следовательно, $|xy| = |x| + |y|$.

Определение 13.4. Соединением (конкатенацией) языков L_1 и L_2 называют язык L_1L_2, состоящий из всех возможных соединений слов xy, в которых слово x принадлежит первому, а слово y — второму языку, т.е.

$$L_1L_2 = \{xy\colon x \in L_1 \text{ и } y \in L_2\}.$$

Соединение конечных языков легко вычислить.

Пример 13.1. $V = \{a, b, c\}$, $L_1 = \{ab, bcc, cab\}$, $L_2 = \{ca, bcc\}$

$$L_1L_2 = \{abca, abbcc, bccca, bccbcc, cabca, cabbcc\},$$
$$L_2L_1 = \{caab, cabcc, cacab, bccab, bccbcc, bcccab\}.$$

Можно формально написать так:

$$(ab + bcc + cab)(ca + bcc) =$$
$$= abca + abbcc + bccca + bccbcc + cabca + cabbcc.$$

Соединение языков не коммутативно: $L_1L_2 \neq L_2L_1$. В общем случае $L_1L_2 \cap L_2L_1$ не пусто. В примере 13.1 пересечение равно $\{bccbcc\}$.

Операция соединения языков позволяет определить операцию возведения языка в произвольную натуральную степень.

По определению, $L^0 = \{\lambda\}$ для любого $L \subseteq V^*$, а $L^n = L^{n-1}L$ при $n > 1$.

Итерацией языка L называют объединение всех его степеней:

$$L^* = \bigcup_{n=0}^{\infty} L^n.$$

Рассматривая объединение всех степеней языка L, начиная с первой, получим **позитивную итерацию**

$$L^+ = \bigcup_{n=1}^{\infty} L^n.$$

Основное алгебраическое свойство множества всех языков в алфавите V.

Теорема 1. Алгебра $\mathcal{L}(V) = \left(2^{V^*}, \cup, \cdot, \varnothing, \{\lambda\}\right)$ есть **замкнутое полукольцо**.

◀ Необходимо проверить **аксиомы полукольца**, т.е. показать, что:
1) алгебра $\left(2^{V^*}, \cup, \varnothing\right)$ — **коммутативный** и **идемпотентный моноид**;
2) алгебра $\left(2^{V^*}, \cdot, \{\lambda\}\right)$ — моноид;
3) операция соединения дистрибутивна относительно объединения;
4) выполняется аннулирующее свойство нуля.

1. Рассмотрим алгебру $(2^{V^*}, \cup)$, где 2^{V^*} — множество всех языков.

Пусть L_1, L_2, L_3 — произвольные подмножества множества 2^{V^*}, т.е. произвольные языки в алфавите V.

Ассоциативность ($L_1 \cup (L_2 \cup L_3) = (L_1 \cup L_2) \cup L_3$),
коммутативность ($L_1 \cup L_2 = L_2 \cup L_1$) и
идемпотентность ($L_1 \cup L_1 = L_1$) операции $\cup$ доказана ранее
(см. свойства операции объединения множеств).
Нейтральный элемент по операции объединения — пустое множество ($\varnothing$). Этот элемент есть **нуль полукольца**.

Следовательно, по операции объединения множество языков в алфавите V образует **коммутативный** и **идемпотентный моноид**.

2. Рассмотрим алгебру $(2^{V^*}, \cdot)$ и покажем, что по операции соединения множество языков образует **моноид**.

Используя **метод двух включений**, докажем ассоциативность операции соединения языков:

$(L_1 \cdot L_2) \cdot L_3 = L_1 \cdot (L_2 \cdot L_3)$.

Пусть слово w принадлежит языку $(L_1 \cdot L_2) \cdot L_3$. Тогда согласно определению соединения языков, слово w может быть представлено в виде $w = vz$, где $v \in L_1 \cdot L_2$ и $z \in L_3$.

Слово v может быть представлено в виде $v = xy$, где $x \in L_1$ и $y \in L_2$. Следовательно, w может быть представлено в виде $w = (xy)z$.

Согласно определению операции соединения слов и в силу ее ассоциативности имеем $w = (xy)z = x(yz)$. Слово $w = x(yz)$, согласно определению соединения языков, принадлежит языку $L_1 \cdot (L_2 \cdot L_3)$.

Обратно:

$w \in L_1 \cdot (L_2 \cdot L_3) \Rightarrow w = xu,$ где $x \in L_1, u \in (L_2 \cdot L_3) \Rightarrow$
$\Rightarrow u = yz,$ где $y \in L_2, z \in L_3 \Rightarrow w = x(yz) = (xy)z \Rightarrow$
$\Rightarrow w \in (L_1 \cdot L_2) \cdot L_3.$

Рассмотрим пустой язык $\{\lambda\}$, состоящий из одного пустого слова. Для любого языка L выполняется тождество $\{\lambda\}L = L\{\lambda\} = L$. Это следует из тождества $\lambda x = x\lambda = x$ для любого слова x.
Следовательно, пустой язык $\{\lambda\}$ есть нейтральный элемент по операции соединения языков — **единица полукольца**.

3. Свойство **дистрибутивности** операции соединения относительно объединения имеет вид:

$$L_1 \cdot (L_2 \cup L_3) = L_1 \cdot L_2 \cup L_1 \cdot L_3, \qquad (1)$$

$$(L_1 \cup L_2) \cdot L_3 = L_1 \cdot L_3 \cup L_2 \cdot L_3 \qquad (2)$$

Докажем тождество (1).

Пусть слово x принадлежит языку $L_1 \cdot (L_2 \cup L_3)$.

Тогда, согласно определению соединения языков, это слово может быть представлено в виде $x = yz$, где $y \in L_1$, а $z \in L_2 \cup L_3$, т.е. $z \in L_2$ или $z \in L_3$.

Если $z \in L_2$, то $yz \in L_1 \cdot L_2$, а если $z \in L_3$, то $yz \in L_1 \cdot L_3$, т.е. $x = yz \in L_1 \cdot L_2 \cup L_1 \cdot L_3$.

Пусть теперь $x \in L_1 \cdot L_2 \cup L_1 \cdot L_3$. Тогда $x = yz$, где $y \in L_1$, а $z \in L_2$ или $z \in L_3$, т.е. $x \in L_1 \cdot (L_2 \cup L_3)$.

Первое тождество доказано.

Тождество (2) доказывается аналогично.

4) Аннулирующее свойство нуля $L \cdot \varnothing = \varnothing \cdot L = \varnothing$ выполняется в силу определения операции соединения языков.

Замкнутость полукольца $\mathcal{L}(V)$ всех языков в алфавите V .

Вспомним определение.

Полукольцо $\mathcal{S} = (S, +, \cdot, \mathbf{0}, \mathbf{1})$ называют **замкнутым**, если:

1) оно идемпотентно;

2) любая последовательность элементов множества S имеет точную верхнюю грань относительно естественного порядка $\leq$ этого идемпотентного полукольца;

3) операция умножения полукольца $\mathcal{S}$ сохраняет точные верхние грани последовательностей, т.е. для любого $a \in S$ и любой последовательности $X = \{x_n\}_{n\in\mathbb{N}}$ элементов множества S

$$a \sup X = \sup aX, \quad (\sup X)a = \sup(Xa).$$

В полукольце $\mathcal{S}$ **отношение порядка** вводится следующим образом: для любых $x, y \in S$ по определению полагают $x \leq y$ тогда и только тогда, когда $x + y = y$.
В полукольце всех языков в алфавите V операция **сложения** — это операция объединения множеств и отношение порядка $\leq$ есть отношение **теоретико-множественного включения** $\subseteq$ (включение $L_1 \subseteq L_2$ равносильно тому, что $L_1 \cup L_2 = L_2$).

Замкнутость полукольца $\mathcal{L}(V)$ следует из существования **объединения** любого **семейства множеств** (в частности, бесконечной последовательности множеств) и из следующих тождеств (для любого языка L и любого семейства языков P_i, $i \in I$):

$$L\Big(\bigcup_{i\in I} P_i\Big) = \bigcup_{i\in I}(LP_i), \quad \Big(\bigcup_{i\in I} P_i\Big)L = \bigcup_{i\in I} P_iL. \qquad (13.1)$$

Объединение семейства множеств служит **точной верхней гранью** этого семейства (относительно теоретико-множественного включения).

Тождества гарантируют выполнение **непрерывности** операции **умножения** данного полукольца, т.е. непрерывности операции соединения.

Эти тождества доказываются точно так же, как тождества обычной дистрибутивности.

Докажем второе тождество из (13.1), используя **метод двух включений**.

Если $x \in \left(\bigcup\limits_{i\in I} P_i\right)L$, то $x = yz$, где $y \in \bigcup\limits_{i\in I} P_i$, а $z \in L$.

Согласно определению объединения семейства множеств, найдется такое $i \in I$, что $y \in P_i$, и тогда $yz = x \in P_iL$, т.е. $x \in \bigcup\limits_{i\in I} P_iL$.

Обратное включение:

из $x \in \bigcup\limits_{i\in I} P_iL$ следует, что для некоторого $i \in I$ $x \in P_iL$, т.е. $x = yz$, где $y \in P_i$, а $z \in L$, откуда $y \in \bigcup\limits_{i\in I} P_i$, и, следовательно, $yz = x \in \left(\bigcup\limits_{i\in I} P_i\right)L$. ►

Следствие 13.1. Для любого языка L верно тождество $L^+ = L^*L = LL^*$.

◄ Вычислим соединение LL^*: $LL^* = L\Big(\bigcup\limits_{n=0}^{\infty} L^n\Big)$.

Применим тождество $L\Big(\bigcup_{i\in I} P_i\Big) = \bigcup_{i\in I}(LP_i)$ (13.1).

Получим

$L\bigcup\limits_{n=0}^{\infty} L^n = \bigcup\limits_{n=0}^{\infty} LL^n = \bigcup\limits_{n=1}^{\infty} L^n$, т.е. $L^+ = LL^*$.

Тождество $L^+ = L^*L$ доказывается аналогично. ►

В общем случае нельзя утверждать, что позитивная итерация языка L получается выбрасыванием из обычной итерации пустого слова. Это верно в том и только в том случае, когда язык L не содержит пустого слова. Если же $\lambda \in L$, то $L^+ = L^*$, так как тогда $L^0 = \{\lambda\} \subseteq L$.

13.2. Регулярные языки и регулярные выражения

Опишем с помощью индуктивной порождающей процедуры множество регулярных языков в алфавите

$$V = \{a_1, \ldots, a_n\} :$$

1) пустой язык $\varnothing$, язык $\{\lambda\}$, состоящий из пустого слова, и однобуквенные языки $\{a_i\}$ для каждого $a_i \in V$ считаем регулярными языками в алфавите V;
2) если P и Q — регулярные языки в алфавите V, то **объединение** $P \cup Q$ и **соединение** PQ есть регулярные языки в алфавите V;
3) если P — регулярный язык в алфавите V, то его итерация P^* есть регулярный язык в алфавите V;
4) других нет.

В замкнутом полукольце $\mathcal{L}(V)$ всех языков в алфавите V рассмотрим подалгебру, обозначаемую $\mathcal{R}(V)$, порожденную множеством регулярных языков.
Согласно порождающей процедуре, эта подалгебра **замкнута относительно итерации**.

Эта подалгебра есть полукольцо с итерацией. Оно играет важнейшую роль в теории формальных языков. Это полукольцо называют **полукольцом регулярных языков**.

Теорема 2. Язык в алфавите V регулярен тогда и только тогда, когда он является элементом полукольца $\mathcal{R}(V)$. #

Алгебраические операции над регулярными языками удобно представлять с помощью так называемых **регулярных выражений**.
Каждое регулярное выражение задает некоторый однозначно определяемый регулярный язык.

Язык	регулярное выражение
$\varnothing$	$\varnothing$
$\{\lambda\}$	λ
$\{a\}$, где $a \in V$	a
P	p
Q	q
$P \cup Q$	$p + q$
$P \cdot Q$	$p \cdot q$
P^*	p^*

В регулярных выражениях для обозначения операции объединения языков используют знак „ $+$ “ (плюс), а для операции соединения — знак умножения „ $\cdot$ “, который как правило опускают.
Соглашение о приоритетах: самый высокий приоритет имеет операция итерации, затем — соединения и, наконец, — объединения.

Пример 13.2. Регулярное выражение $a^* + (bc)^*$ обозначает множество цепочек, состоящее из цепочек вида a^n, $n \geq 0$, и цепочек вида $(bc)^n$, $n \geq 0$, где a, b, $c \in V$.
Слова (цепочеки) языка, заданного регулярным выражением $a^* + (bc)^*$: $a, aa, \ldots, a^n \ldots$ или $bc, bcbc, \ldots, (bc)^n, \ldots$, где $n \geq 0$.
Вместо регулярного выражения мы должны были бы использовать более громоздкую формулу: $\{a\}^* \cup (\{b\} \cdot \{c\})^*$.

Соответствие между регулярными языками и регулярными выражениями не является взаимно однозначным: один и тот же регулярный язык может представляться многими регулярными выражениями.
Например,

$$(a + b)^* = (a^* b^*)^*.$$

Для регулярного выражения $\alpha\alpha^*$ или $\alpha^*\alpha$ будем использовать обозначение α^+ и называть это выражение **позитивной итерацией** выражения α.
Из определения регулярного языка, теоремы 2 и следствия 13.1 следует, что **позитивная итерация** регулярного языка регулярна.

Конечные автоматы

Индуктивная процедура построения регулярных языков можно рассматривать как порождающую модель для регулярных языков.

Распознающей моделью для регулярных языков служат конечные автоматы.

Неформально конечный автомат можно описать как устройство, состоящее из **блока управления**,**входной ленты** и **головки автомата**

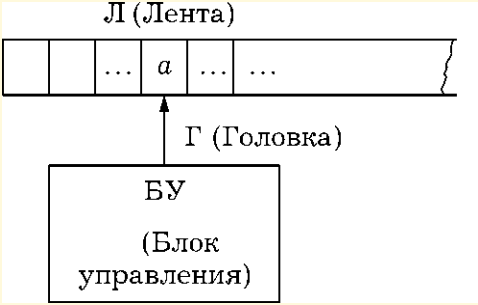

Рис. 1

Блок управления автомата может в каждый момент времени находиться в одном из конечного множества **состояний** Q. Головка автомата может быть установлена в точности на одну ячейку входной ленты.

Входная лента — это неограниченная справа полубесконечная лента, разделенная на ячейки. На ней записываются цепочки во **входном алфавите** V так, что буквы цепочки занимают ячейки ленты. Буквы записываются последовательно, начиная с первой без пропусков, по одной букве в каждой ячейке.

Цепочку, записанную на входной ленте автомата, называют **входной цепочкой**.

Автомат, читая входную цепочку, работает по шагам (или по тактам).

Пусть в некоторый момент времени автомат читает содержимое текущей ячейки ленты, блок управления находится в некотором состоянии $q \in Q$.

Такт работы автомата состоит в том, что в зависимости от содержимого читаемой ячейки, состояния q, а также содержимого внутренней памяти автомат сдвигает головку вправо на одну ячейку либо оставляет ее на прежнем месте, его блок управления переходит в некоторое новое состояние r (это состояние может совпасть с исходным состоянием q).

Пусть на входной ленте автомата записана некоторая цепочка $x \in V^*$.

Среди состояний блока управления выделено некоторое специальное состояние, называемое **начальным**, и некоторое подмножество состояний, которые называются **заключительными**.

В начальный момент времени блок управления находится в начальном состоянии, головка читает содержимое первой (крайней левой) ячейки ленты, в которой записан первый символ входной цепочки x.

Автомат читает цепочку x и делает один такт за другим до тех, пор пока он не прочитает последнюю букву цепочки.

После чтения последней буквы цепочки блок управления окажется в некотором состоянии q'. Если это состояние является заключительным, то тогда говорят, что автомат допустил цепочку x.

Из каждого состояния автомат может переходить в другое состояние, читая символ входной цепочки.

Автомат может переходить из одного состояния в другое по пустому слову, не читая ленту и не продвигая головку, если такие переходы предусмотрены описанием автомата. Такой такт можно рассматривать как переход из состояния в состояние по **пустой цепочке**. Его называют λ-тактом.

Принимается, что переход по пустой цепочке и переход по входному символу исключают друг друга.

Поведение конечного автомата определяется его системой команд, в которой каждая команда задается записью

$$qa \to r, \tag{13.2}$$

(„из состояния q по символу $a \in V$ можно перейти в состояние r “),
или

$$q\lambda \to r, \tag{13.3}$$

(„из состояния q по пустой цепочке можно перейти в состояние r “).
Возможно, что $q = r$.
Если для любой пары $(q,\, r)$ состояний конечного автомата существует команда (13.3), то для той же пары состояний нет ни одной команды (13.2) при $a \in V$ и наоборот.

Конечный автомат допускает интерпретацию в терминах **размеченных ориентированных графов**.

Будем рассматривать **состояния блока** управления конечного автомата как **вершины** ориентированного графа, множество **дуг** которого определяется системой команд следующим образом: дуга ведет из состояния q в состояние r (для данных состояний q и r) тогда и только тогда, когда в системе команд автомата есть команда (13.2) или команда (13.3), т.е. возможен переход из состояния q в состояние r.

Метка дуги (q, r) есть пустая цепочка λ, если из q в r можно перейти по пустой цепочке; в противном случае метка дуги (q, r) есть множество всех входных символов, по которым возможет переход из состояния q в состояние r.

Пример 13.3. Конечный автомат: входной алфавит $\{a, b, c\}$, множество состояний $\{q_0, q_1, q_2, q_3\}$, система команд:

$$q_0\lambda \to q_1, \quad q_0\lambda \to q_3, \quad q_1a \to q_2,$$
$$q_1b \to q_2, \quad q_1a \to q_3, \quad q_2b \to q_1,$$
$$q_3b \to q_2, \quad q_3c \to q_2, \quad q_3c \to q_3.$$

По этой системе команд построим размеченный ориентированный граф.

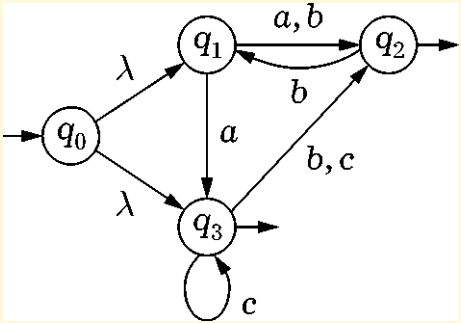

Рис. 2

Среди состояний автомата выделены начальное состояние q_0, помеченное стрелкой $\rightarrow \bigcirc$, и два заключительных состояния q_2 и q_3, также помеченные стрелкой $\bigcirc \rightarrow$.

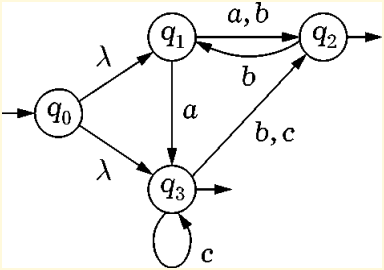

Рис. 3

Последовательности $abac$ отвечает **путь** в ориентированном графе, ведущий из вершины q_0 в вершину q_3:

$q_0 \rightarrow_\lambda q_1 \rightarrow_a q_2 \rightarrow_b q_1 \rightarrow_a q_3 \rightarrow_c q_3$

Обозначение пустой цепочки λ, в виде метки дуги ориентированного графа, который представляет конечный автомат, можно интерпретировать как **регулярное выражение**, т.е. как регулярный язык, состоящий из одной пустой цепочки.
Метка дуги — это множество букв $\{b_1, \ldots, b_m\} \subseteq V$.
Метка дуги может быть также записана в виде регулярного выражения, как сумма $b_1 + \ldots + b_m$.
Метку каждой дуги можно считать регулярным выражением определенного вида или регулярным языком.

Это позволяет формально определить конечный автомат как **ориентированный граф, размеченный над полукольцом** $\mathcal{R}(V)$ регулярных языков.
Такое математическое определение конечного автомата дает возможность применить при изучении конечных автоматов алгебраические методы анализа размеченных ориентированных графов.

Математическое определение конечного автомата.

Определение 13.5. Конечный автомат — это ориентированный граф, размеченный над полукольцом $\mathcal{R}(V)$ регулярных языков в алфавите V с выделенной вершиной q_0, которая называется **начальной**, и с выделенным подмножеством вершин F, каждый элемент которого называется **заключительной вершиной**.
На **функцию разметки** при этом накладываются следующие ограничения: **метка** каждой **дуги** есть либо язык $\{\lambda\}$, либо непустое подмножество **алфавита** V.

Вершины графа называют обычно в этом случае **состояниями конечного автомата**, начальную вершину — **начальным состоянием**, а заключительную вершину — **заключительным состоянием конечного автомата**.

Пример 13.4.

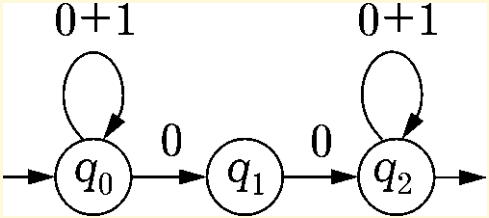

Рис. 4

На рисунке изображен конечный автомат, для которого алфавит $V = \{0, 1\}$.
Начальное состояние показано входной стрелкой, заключительное — выходной.
Метки дуг обычно пишут без фигурных скобок. Разрешена запись меток дуг и в виде *регулярных выражений*.

Конечный автомат может быть задан как пятерка:

$$M = (\mathcal{Q},\ E,\ \varphi,\ q_0,\ F),$$

где

$\mathcal{Q}$ — множество состояний автомата;

E — множество дуг;

φ — функция разметки (весовая функция), причем для каждой дуги $e \in E$ ее метка $\varphi(e) = \{\lambda\}$ или $\varphi(e) \subseteq V$, при этом подмножество $\varphi(e)$ не пусто; $q_0 \in Q$ — начальное состояние;

$F \subseteq Q$ — подмножество заключительных состояний.

Алфавит V называется **входным алфавитом автомата** M, а его буквы — **входными символами** (или **входными буквами**) данного **автомата**.

Если $e = (q,\, r)$ — **дуга** автомата M и ее метка $\varphi(e)$ есть регулярное выражение λ, то в этом случае будем говорить, что в автомате M возможен **переход из состояния** q **в состояние** r по пустой цепочке, и писать $q \to_\lambda r$.
Дугу с меткой λ будем называть λ**-переходом** (или **пустой дугой**).
Если же метка дуги e есть множество, содержащее входной символ a, то будем говорить, что в автомате M возможен переход из состояния q в состояние r по символу a, и писать $q \to_a r$.

Для конечного автомата удобно ввести в рассмотрение **функцию переходов**, определив ее как отображение

$$\delta\colon Q \times (V \cup \{\lambda\}) \to 2^Q,$$

такое, что

$$\delta(q, a) = \{r\colon q \to_a r\},$$

т.е. значение функции переходов на упорядоченной паре (состояние, входной символ или пустая цепочка) есть множество всех состояний, в которые из данного состояния возможен переход по данному входному символу или пустой цепочке. В частности, это может быть пустое множество.

Замечание 13.1. Хотя конечный автомат определен как ориентированный граф, размеченный над полукольцом регулярных языков, метка дуги задается не как произвольный регулярный язык, а как язык, являющийся подмножеством **букв** входного алфавита, или язык, состоящий из одной пустой цепочки. Это связано с приведенным ранее описанием автомата, как устройства.

Рассмотрим автомат из примера 13.4

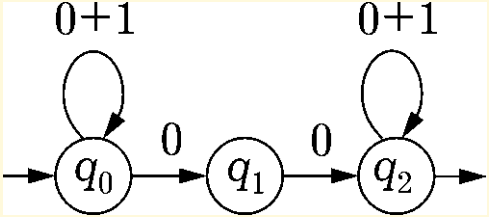

Рис. 5

Функции переходов конечного автомата:

$\delta(q_0, 0) = \{q_0,\ q_1\}$
$\delta(q_0, 1) = \{q_0\}$
$\delta(q_1, 0) = \{q_2\}$
$\delta(q_1, 1) = \varnothing$
$\delta(q_2, 0) = \{q_2\}$
$\delta(q_2, 1) = \{q_2\}.$

Система команд есть конечное множество **команд** вида
$qa \to r$,
где q и r — состояния автомата;
a — входной символ или пустая цепочка, причем указанная команда тогда и только тогда содержится в системе команд, когда $q \to_a r$.
Стрелка $(\to)$, есть „метасимвол“, он не содержится алфавите V.

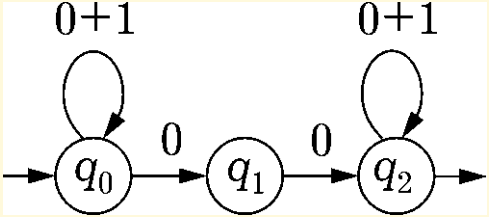

Рис. 6

Система команд автомата из примера 13.4:

$$
\begin{aligned}
q_0 0 &\to q_0,\\
q_0 0 &\to q_1,\\
q_0 1 &\to q_0,\\
q_1 0 &\to q_2,\\
q_2 0 &\to q_2,\\
q_2 1 &\to q_2.
\end{aligned}
$$

Используя функцию переходов, конечный автомат можно задать как упорядоченную пятерку:

$M = (V,\ Q,\ q_0,\ F,\ \delta)$,

где V — входной алфавит;

Q — множество состояний;

q_0 — начальное состояние;

F — множество заключительных состояний;

δ — функция переходов, заданная в виде системы команд.

Согласно определению **метки пути** в размеченном ориентированном графе, **метка пути** в конечном автомате есть **умножение** меток входящих в этот путь дуг (в порядке их прохождения). **Умножением полукольца** $\mathcal{R}(V)$ является **соединение языков**.

Таким образом, метка любого **пути конечной длины** в конечном автомате есть регулярный язык.

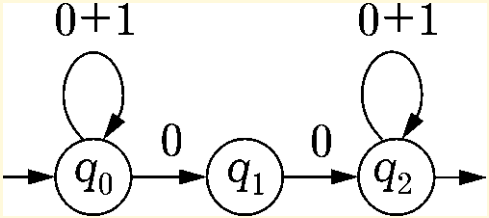

Рис. 7

Метка пути q_0, q_0, q_1, q_2 равна соединению языков $\{0,\ 1\} \cdot \{0\} \cdot \{0\} = \{000,\ 100\}$.

Это можно записать в виде регулярного выражения $(0+1)00$.

Метка пути q_0, q_1, q_2, q_2, q_2 может быть задана таким регулярным выражением:

$$00(0+1)(0+1) = 00(00+01+10+11) =$$
$$= 0000 + 0001 + 0010 + 0011,$$

Эта метка есть множество цепочек $\{0000,\ 0001,\ 0010,\ 0011\}$.

Если цепочка x принадлежит метке некоторого пути W — пути, ведущего из вершины q в вершину r конечного автомата M, то говорят, что цепочка x читается на пути W в M.

Пишем $q \Rightarrow_x^* r$, если x читается на некотором пути из q в r.

В том случае, когда явно надо указать **длину** n **пути**, на котором читается цепочка x, записываем $q \Rightarrow_x^n r$.

Если нужно подчеркнуть, что цепочка x читается на некотором пути ненулевой длины из q в r, то используем запись $q \Rightarrow_x^+ r$.

Стоимость прохождения из состояния q в состояние r есть (согласно общему определению этого понятия в размеченных ориентированных графах) **объединение** меток всех путей ведущих из q в r, т.е. множество всех таких x, что $q \Rightarrow_x^* r$. Это значит, что элемент c_{qr} **матрицы стоимостей** есть язык

$$c_{qr} = \{x\colon q \Rightarrow_x^* r\}.$$

Здесь объединение понимается как **бесконечная сумма** замкнутого полукольца $\mathcal{L}(V)$ всех языков в алфавите V, т.е. как **точная верхняя грань** последовательности языков относительно теоретико-множественного включения.

Определение 13.6. Язык $L(M)$ конечного автомата M есть множество всех цепочек во входном алфавите, читаемых в M на некотором пути из начального состояния в одно из заключительных состояний. Другими словами,

$$L(M) = \{x\colon q_0 \Rightarrow_x^* q_f,\ q_f \in F\} = \bigcup_{q_f \in F} c_{q_0 q_f}, \qquad (13.4)$$

где F — множество заключительных состояний.

Т. о. язык конечного автомата есть объединение тех элементов матрицы стоимостей автомата, которые находятся в строке, соответствующей начальному состоянию q_0, и в столбцах, соответствующих всем заключительным состояниям $q_f \in F$.

Факт, что стоимость прохождения между заданной парой вершин является регулярным языком, требует доказательства.

В конечном автомате метка произвольного пути конечной длины есть регулярный язык, поскольку он вычисляется как соединение конечного семейства регулярных языков.
Однако множество путей, ведущих из одной вершины в другую, может быть бесконечным.
Пусть, например, некоторый конечный путь, ведущий из вершины p в вершину q, содержит **контур**.
Тогда по этому контуру можно пройти любое количество раз. Следовательно множество путей, ведущих из вершины p в q, будет бесконечным.

Стоимость прохождения из вершины p в вершину q будет равна бесконечному объединению меток путей, ведущих из вершины p в q , и необязательно будет регулярным языком. Это бесконечное объединение можно рассматривать как операцию вычисления точной верхней грани в замкнутом полукольце всех языков в данном алфавите.

О языке $L(M)$ говорят, что он допускается КА M. О любой цепочке, принадлежашей языку $L(M)$, говорят, что она допускается КА M. Такую цепочку называют также **допустимой цепочкой** данного **конечного автомата**.

Определение 13.7. Два конечных автомата M_1 и M_2 называют **эквивалентными**, если их языки совпадают:

$$L(M_1) = L(M_2).$$

ДИСКРЕТНАЯ МАТЕМАТИКА

ИУ5 - 4 семестр, 2014 г.

Лекция 14. ТЕОРЕМА КЛИНИ. ДЕТЕРМИНИЗАЦИЯ

Была сформулирована следующая теорема: ” Язык в алфавите V регулярен тогда и только тогда, когда он является элементом полукольца $\mathcal{R}(V)$.”

Это позволило нам называть элементы полукольца $\mathcal{R}(V)$ регулярными языками.

Докажем, что язык допускается конечным автоматом с входным алфавитом V тогда и только тогда, когда он есть элемент полукольца $\mathcal{R}(V)$.

Теорема 1 (теорема Клини). Пусть $V = \{a_1, \ldots, a_n\}$ — произвольный алфавит. Язык $L \subseteq V^*$ является элементом полукольца $\mathcal{R}(V)$ тогда и только тогда, когда он допускается некоторым конечным автоматом.

◄ **1. Докажем, что всякий язык из $\mathcal{R}(V)$ допускается некоторым конечным автоматом.**

Для доказательства этого утверждения воспользуемся методом индукции по построению языка из $\mathcal{R}(V)$ как элемента **замыкания** множества $\{\varnothing, \{\lambda\}, \{a_1\}, \ldots, \{a_n\}\}$.

Этот метод состоит в следующем: сначала утверждение доказывается для языков исходного множества (замыкание которого строится), а затем в предположении, что утверждение доказано для языков L и K из $\mathcal{R}(V)$, оно доказывается для $L \cup K$, LK и L^*.

Пусть V — некоторый фиксированном алфавите .
Конечные автоматы для языков $\varnothing$, $\{\lambda\}$, $\{a\}$, где $a \in V$:

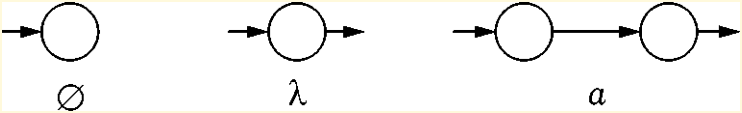

Рис. 1

Пусть конечные автоматы

$$M_1 = (V,\, Q_1,\, q_{01},\, F_1,\, \delta_1) \quad \text{и} \quad M_2 = (V,\, Q_2,\, q_{02},\, F_2,\, \delta_2)$$

для языков L и K полукольца $\mathcal{R}(V)$ соответственно уже построены.
Входные алфавиты этих автоматов совпадают и автоматы не имеют ни общих вершин, ни общих дуг.

Построим конечные автоматы, допускающие языки $L \cup K$, LK и L^*.

Автомат для **объединения языков** получается путем добавления нового начального состояния s_0, проведения из него пустых дуг в каждое из начальных состояний (q_{01} и q_{02}) объединяемых автоматов M_1 и M_2.

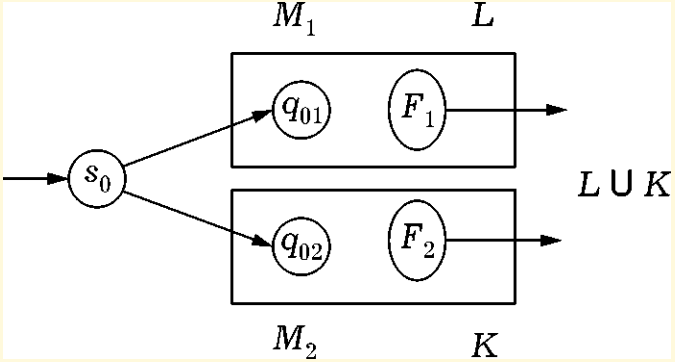

Рис. 2

Все дуги и состояния автоматов M_1 и M_2 сохраняются.

Объединение множеств заключительных состояний (F_1 и F_2) автоматов M_1 и M_2 объявляются множеством заключительных состояний конечного автомата, допускающего язык $L \cup K$.

Получается „параллельное соединение“ автоматов для языков L и K .

Новый конечный автомат как некое распознающее устройство может из своего начального состояния перейти в начальное состояние первого или в начальное состояние второго автомата.

Любая цепочка x, читаемая на некотором пути из состояния s_0 в какое-то из состояний множества $F_1 \cup F_2$, может быть представлена так: $x = \lambda x_1 = x_1$ или $x = \lambda x_2 = x_2$.

В первом случае — это переход по пустой цепочке из s_0 q_{01} и дальнейшее чтение произвольной цепочки x_1, допускаемой автоматом M_1.

Во втором случае — переход по пустой цепочке из s_0 q_{02} и дальнейшее чтение произвольной цепочки x_2, допускаемой автоматом M_2.

При построении конечного автомата для **соединения** новым начальным состоянием будет начальное состояние первого автомата (q_{01}).

Множество заключительных состояний —это множество заключительных состояний второго автомата (F_2).

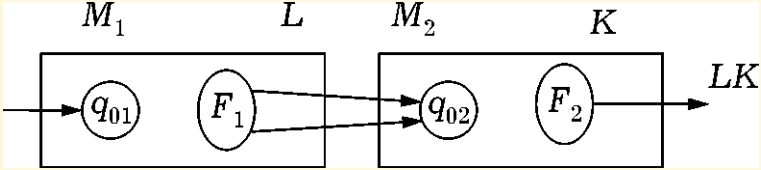

Рис. 3

Из каждого заключительного состояния первого автомата провести пустую дугу в начальное состояние второго автомата.

Получится „последовательное соединение“ автоматов.

Конечный автомат для **итерации языка** L :

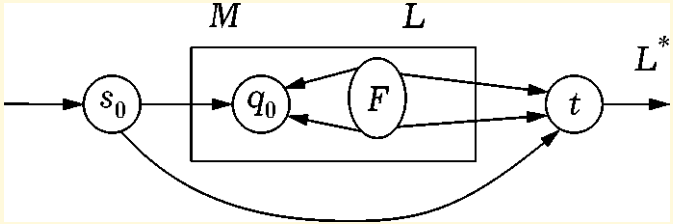

Рис. 4

1. Вводятся новое начальное состояние (s_0) и новое заключительное состояние (f).

2. Проводятся пустые дуги из нового начального состояния s_0 в прежнее начальное состояние q_0 автомата, допускающего язык L .

3. Проводятся пустые дуги из каждого заключительного состояния множества F автомата, допускающего язык L в новое заключительное состояние и прежнее начальное состояние.

Вывод. Каждый язык полукольца $\mathcal{R}(V)$ допускается некоторым конечным автоматом.

2. Докажем, что язык произвольного конечного автомата есть элемент полукольца $\mathcal{R}(V)$.

Язык конечного автомата, как следует из формулы

$$L(M) = \{x\colon q_0 \Rightarrow_x^* q_f,\ q_f \in F\} = \bigcup_{q_f \in F} c_{q_0 q_f},$$

— это конечное объединение языков, являющихся определенными элементами матрицы стоимостей автомата.

Матрица стоимостей есть **итерация матрицы** меток дуг, задающей автомат.

Метка каждой дуги — регулярное выражение, обозначающее язык из полукольца $\mathcal{R}(V)$.

Матрица стоимостей является итерацией матрицы, все элементы которой могут быть определены регулярными выражениями, т.е. принадлежат полукольцу $\mathcal{R}(V)$.
Полукольцо $\mathcal{R}(V)$ есть полукольцо с итерацией.
Вспомним утверждение: "Если A — матрица, все элементы которой принадлежат некоторому полукольцу с итерацией, то все элементы ее **итерации** A^* также принадлежат этому полукольцу с итерацией."
В силу этого утверждения матрица стоимостей конечного автомата будет состоять из языков полукольца $\mathcal{R}(V)$.
Отсюда следует, что язык конечного автомата есть элемент этого полукольца. ►

Замечание В общем случае при построении итерации нельзя обойтись без добавления новых начального и заключительного состояния.

Построим КА для итерации языка, допускаемого следующим КА.

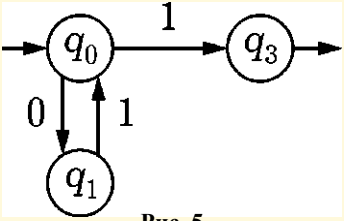

Рис. 5

Язык L: $(01)^*1$.

Итерация языка L есть $L^* = ((01)^*1)^*$.

Любая цепочка из итерации исходного языка есть либо цепочка 1^n, где $n \geq 0$, либо цепочка, оканчивающаяся подцепочкой 11.

Построим автомат для итерации, не вводя новое начальное состояние.

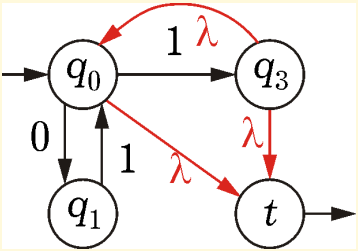

Рис. 6

Этот автомат будет допускать цепочки, описываемые регулярным выражением $(01)^*$. Однако эти цепочки не содержатся в языке L^* (итерации исходного языка L). В частности, $01 \notin ((01)^*1)^*$.

Исходный автомат:

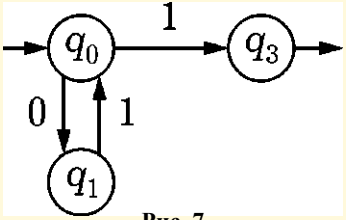

Рис. 7

Правильно построенный автомат для итерации языка:

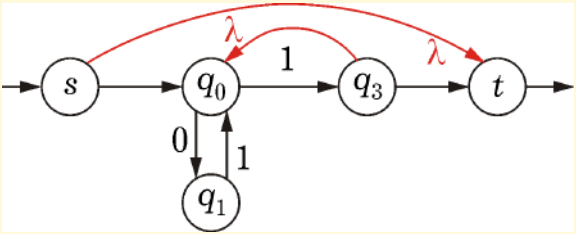

Рис. 8

Вычисление языка конечного автомата.

Вычислим матрицу стоимостей $C = A^*$ автомата как размеченного ориентированного графа.

Надо решить $n = |Q|$ систем вида

$$X^j = AX^j + B^j,$$

где A — квадратная матрица n-го порядка.
Элемент a_{ij} является регулярным выражением, служащим меткой дуги из вершины (состояния) q_i в вершину (состояние) q_j, если такая дуга существует, и равен регулярному выражению $\varnothing$, если нет дуги из q_i в q_j;
B^j — j-й столбец единичной матрицы, т.е. столбец, у которого все компоненты, кроме j-й, равны нулю полукольца $\mathcal{R}(V)$ — $\varnothing$, j-я компонента равна единице полукольца $\mathcal{R}(V)$ — λ.

Язык $L(M)$ конечного автомата M есть множество всех цепочек во входном алфавите, читаемых в M на некотором пути из начального состояния в одно из заключительных состояний:

$$L(M) = \{x\colon q_0 \Rightarrow_x^* q_f,\ q_f \in F\} = \bigcup_{q_f \in F} c_{q_0 q_f},$$

где F — множество заключительных состояний.
Поэтому язык КА будет задаваться регулярным выражением, равным сумме элементов вида c_{st_i}, где s — номер начального, t_i, $i = 1, \ldots, m$ — номера заключительных состояний:

$$\sum_{i=1}^{m} c_{st_i}.$$

Для нахождения языка конечного автомата достаточно решить одну систему линейных уравнений:

$$X^j = AX^j + \beta, \tag{14.1}$$

где β — столбец, все компоненты которого равны $\varnothing$ (нулю полукольца $\mathcal{R}(V)$), кроме компонент с номерами $t_1, \ldots, t_m$, которые являются номерами заключительных состояний. Эти компоненты равны λ (единице полукольца $\mathcal{R}(V)$).

Другими словами, ко всем уравнениям системы, соответствующим заключительным состояниям, добавляется слагаемое λ.

Решение системы (14.1) будет иметь вид

$$X^j = A^*\beta = A^*(\varnothing, \ldots, \varnothing, \lambda, \varnothing, \ldots, \varnothing, \lambda, \varnothing, \ldots, \varnothing)^{\mathrm{T}} \tag{14.2}$$

(элементы λ находятся в строках с номерами $t_1, \ldots, t_m$). Умножая в (14.2) матрицу A^*, равную матрице C стоимостей, на столбец β, получим столбец, s-я компонента которого x_s будет равна произведению s-й строки матрицы C $(c_{s1}, \ldots, c_{st_1}, \ldots, c_{st_m}, \ldots, c_{sn})$ на столбец β в формуле (14.2), т.е.

$$x_s = c_{st_1} + \ldots + c_{st_m},$$

но это и есть регулярное выражение, обозначающее язык конечного автомата.

Пример 14.1.

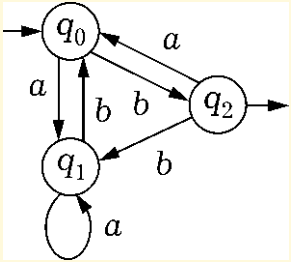

Рис. 9

Запишем для этого автомата матрицу A меток дуг и систему уравнений:

$$A = \begin{pmatrix} 0 & a & b \\ b & a & 0 \\ a & b & 0 \end{pmatrix}, \qquad \begin{cases} x_0 = ax_1 + bx_2, \\ x_1 = bx_0 + ax_1, \\ x_2 = ax_0 + bx_1 + \lambda \end{cases}$$

Слагаемое λ добавлено в уравнение для x_2, так как вершина q_2 является заключительной.

Исключая x_0, получаем

$$\begin{cases} x_1 = b(ax_1 + bx_2) + ax_1, \\ x_2 = a(ax_1 + bx_2) + bx_1 + \lambda, \end{cases}$$

откуда

$$\begin{cases} x_1 = (ba + a)^* b^2 x_2, \\ x_2 = (a^2 + b)(ba + a)^* b^2 x_2 + abx_2 + \lambda. \end{cases}$$

Тогда

$$\begin{cases} x_2 = ((a^2+b)(ba+a)^*b^2+ab)^*, \\ x_1 = (ba+a)^*b^2((a^2+b)(ba+a)^*b^2+ab)^*. \end{cases}$$

Отсюда получим регулярное выражение, обозначающее язык конечного автомата, как значение переменной x_0 :

$$x_0 = a(ba+a)^*b^2((a^2+b)(ba+a)^*b^2+ab)^*+b((a^2+b)(ba+a)^*b^2$$

Полученное регулярное выражение весьма сложно. Найти его, не располагая заранее разработанным алгоритмом, было бы затруднительно.

14.1. Детерминизация конечных автоматов

Функция переходов— это отображение
$\delta\colon Q \times (V \cup \{\lambda\}) \to 2^Q$,
такое, что $\delta(q, a) = \{r\colon q \to_a r\}$,
т.е. значение функции переходов на упорядоченной паре (состояние, входной символ или пустая цепочка) есть множество всех состояний, в которые из данного состояния возможен переход по данному входному символу или пустой цепочке. В частности, это может быть пустое множество.

Используя функцию переходов, конечный автомат можно задавать как упорядоченную пятерку:
$M = (V, Q, q_0, F, \delta)$,
где V — входной алфавит; Q — множество состояний; q_0 — начальное состояние; F — множество заключительных состояний; δ — функция переходов, заданная в виде системы команд.

Конечный автомат называется **детерминированным**, если в нем нет дуг с меткой λ и из любого состояния по любому входному символу возможен переход в точности в одно состояние, т.е.
$(\forall q \in Q)(\forall a \in V)(|\delta(q, a)| = 1)$.

Конечный автомат называется **квазидетерминированным**, если в нем нет дуг с меткой λ и из любого состояния по любому символу возможен переход не более чем в одно состояние, т.е.
$(\forall q \in Q)(\forall a \in V)(|\delta(q, a)| \leq 1)$.

Для детерминированного конечного автомата значением функции переходов для любой пары (состояние, входной символ) является одноэлементное подмножество множества состояний.

Для решения **задачи синтеза конечных автоматов** важное значение имеет следующая теорема.

Теорема 2 (теорема о детерминизации). Для любого конечного автомата может быть построен **эквивалентный** ему **детерминированный** конечный автомат.

(без доказательства)

Алгоритм построения детерминированного автомата.

Преобразование произвольного конечного автомата к эквивалентному детерминированному осуществляется в два этапа:

1. Строится новый КА, не содержащий **дуг** с **меткой** λ.

2. По построенному автомату строится детерминированный КА, эквивалентный исходному автомату.

1. Построение КА без λ-переходов.

Переход от исходного КА $M = (V, Q, q_0, F, \delta)$ к эквивалентному КА $M_1 = (V, Q_1, q_0, F_1, \delta_1)$ без λ-переходов осуществляется следующим образом:

а. Определим множество состояний Q_1 нового КА M_1. Включим в множество состояний начальное состояние q_0 исходного автомата M и все состояния из множества состояний Q исходного КА M, в которые заходит хотя бы одна дуга, помеченная буквой входного алфавита.

Все состояния из Q, в которые заходят **только дуги с меткой** λ, не включаются в множество состояний Q_1.

б. Определим множество дуг конечного автомата M_1 и их меток (функцию переходов M_1) следующим образом.
Для любых двух состояний $p, r \in Q_1$ $p \to_a r$ имеет место тогда и только тогда, когда $a \in V$, в графе M либо существует дуга из p в r, метка которой содержит символ a, либо существует такое состояние q, что $p \Rightarrow_\lambda^+ q$ и $q \to_a r$.

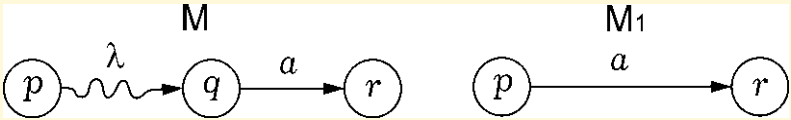

Рис. 10

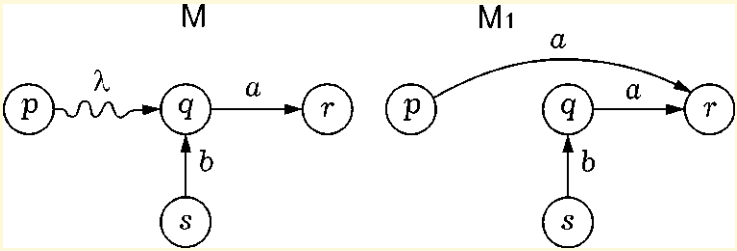

Рис. 11

в. Включим в множество заключительных состояний F_1 конечного автомата M_1 все состояния $q \in Q_1$, которые являются заключительными состояниями исходного КА M, или из которых ведет путь ненулевой длины по дугам с меткой λ в одно из заключительных состояний КА M.

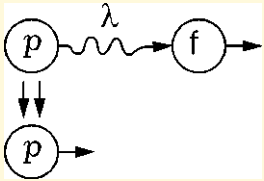

Рис. 12

2. Построение детерминированного КА.

Пусть $M_1 = (Q_1, V, q_0, F_1, \delta_1)$ — конечный автомат без λ -переходов.

Построим эквивалентный M_1 детерминированный КА M_2 , в котором из любого состояния КА по любому входному символу возможен переход ровно в одно состояние.

Каждое отдельное состояние $S = \{q_{s1}, \ldots, q_{sk}\}$ КА M_2 определено как некоторое подмножество множества состояний КА $M_1 : S \subseteq Q_1$.

Множество состояний КА M_2 есть подмножество множества всех подмножеств множества состояний КА M_1 $Q_2 \subseteq 2^Q$.

Начальным состоянием КА M_2 является одноэлементное подмножество $\{q_0\}$, содержащее начальное состояние КА M_1.

Заключительными состояниями КА M_2 являются все подмножества Q_1, которые содержат хотя бы одну заключительную вершину КА M_1.

Состояния КА будем иногда называть состояниями-множествами. Каждое такое состояние-множество есть отдельное состояние КА M_2, но не множество его состояний.

Для исходного КА M_1 это именно множество его состояний. Каждое подмножество состояний старого КА M_1 „свертывается“ в одно состояние нового конечного автомата.

Функция переходов нового КА:
из состояния-множества S по входному символу a КА M_2 переходит в состояние-множество, представляющее собой объединение всех множеств состояний КА M_1, в которые этот КА переходит по символу a из каждого состояния множества S

$$\delta_2(S, a) = \bigcup_{q \in S} \delta(q, a).$$

Таким образом, КА M_2 является детерминированным по построению.

Формально конечный автомат M_2 определяется так:

$$M_2 = (Q_2,\, V,\, \{q_0\},\, F_2,\, \delta_2),$$

где

$$\begin{cases} Q_2 \subseteq 2^Q, \quad F_2 = \{T\colon T \cap F_1 \neq \varnothing,\, T \in Q_2\}, \\ (\forall S \subseteq Q_2)(\forall a \in V)(\delta_2(S,a) = \bigcup\limits_{q \in S} \delta(q,a)). \end{cases} \tag{14.3}$$

Среди состояний нового КА может быть состояние $\varnothing$, причем $\delta_2(\varnothing, a) = \varnothing$ для любого входного символа a.

Попав в такое состояние, КА M_2 уже его не покинет.
Поглощающим состоянием конечного автомата называют такое состояние q КА, что для любого входного символа a $\delta(q, a) = q$.

Таким образом, состояние $\varnothing$ детерминированного конечного автомата M_2 является поглощающим.

$\delta_2(S, a) = \varnothing$ тогда и только тогда, когда в старом КА M_1 для каждого состояния $q \in S$ из множества состояний S $\delta_1(q, a) = \varnothing$, т.е. в графе M_1 из каждого такого состояния q не выходит ни одна дуга, помеченная символом a.

Пример 14.2. Детерминизируем конечный автомат M (рис.13)

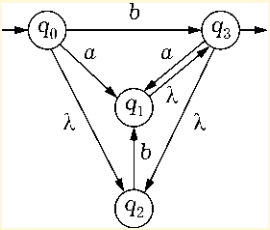

Рис. 13

Построим КА M_1 без λ-переходов, эквивалентный КА M .

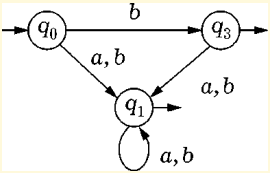

Рис. 14

Состояние q_2 не вошло в множество состояний нового автомата M_1, так как в него заходят только дуги с метками λ в КА M.

Из состояния q_1 по дуге с меткой λ можно попасть в заключительное состояние q_3, следовательно, состояние q_1 войдет в множества заключительных состояний.

Чтобы детерминизировать полученный автомат воспользуемся **методом вытягивания**. В результате получим детерминированный КА M_2, все состояния которого достижимы из начальной вершины $\{q_0\}$.

Метод вытягивания.

В КА без λ (пустых) дуг M_1 определяем все множества состояний, достижимых из начального.

1. Для **каждого символа входного алфавита** a находим множество $\delta(q_0, a)$. Каждое такое множество в новом автомате является состоянием-множеством, непосредственно достижимым из начального.

2. Для каждого из определенных на предыдущем шаге состояний-множеств S и каждого символа входного алфавита a находим множество $\bigcup\limits_{q \in S} \delta(q, a)$.

Все полученные на этом шаге состояния будут состояниями нового детерминированного автомата M_2, достижимыми из начальной вершины $\{q_0\}$ по пути длины 2.

3. Повторяем шаг 2 до тех пор, пока не перестанут появляться новые состояния-множества (включая пустое). Получим все состояния нового детерминированного автомата M_2, достижимые из начальной вершины $\{q_0\}$ по пути длины $n, n \geq 2$.

4. Выделяем множество заключительных состояний. Состояние-множество нового детерминированного конечного автомата (ДКА) будет заключительным, оно включает хотя бы одно заключительное состояние автомата M_1

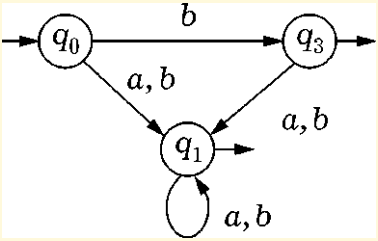

Рис. 15. Повторим рисунок 2

Для конечного автомата M_1 имеем:

$$\delta_1(\{q_0\}, a) = \{q_1\}; \qquad \delta_1(\{q_0\}, b) = \{q_1,\ q_3\};$$
$$\delta_1(\{q_1\}, a) = \{q_1\}; \qquad \delta_1(\{q_1\}, b) = \{q_1\};$$
$$\delta_1(\{q_1,\ q_3\}, a) = \delta(q_1, a) \cup \delta(q_3, a) = \{q_1\} \cup \{q_1\} = \{q_1\};$$
$$\delta_1(\{q_1,\ q_3\}, b) = \delta(q_1, b) \cup \delta(q_3, b) = \{q_1\} \cup \{q_1\} = \{q_1\}.$$

Так как новых состояний-множеств больше не появилось, процедура „вытягивания“ на этом заканчивается. Получаем следующий граф:

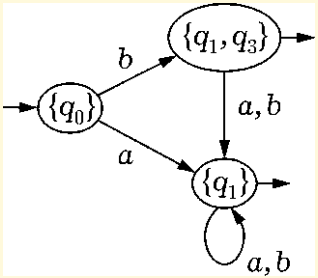

Рис. 16

Пусть L — регулярный язык в алфавите V. Тогда дополнение языка L (как множества слов) есть язык $\overline{L} = V^* \setminus L$. Важным следствием теоремы о детерминизации является следующая теорема.

Теорема 3. Дополнение регулярного языка есть регулярный язык.

◄ Согласно **теореме о детерминизации**, для регулярного языка L может быть построен ДКА M, допускающий язык L.
В детерминированном автомате из каждого состояния по каждому входному символу определен переход в точности в одно состояние.
Для любой цепочки x в алфавите V найдется единственный путь в КА M, на котором читается эта цепочка x, начинающийся в начальном состоянии q_0 и заканчивающийся в некотором состоянии p.

Цепочка x допускается автоматом M ($x \in L(M)$) тогда и только тогда, когда последнее состояние p указанного пути является заключительным ($p \in F$).
Отсюда следует, что цепочка $x \notin L(M)$ тогда и только тогда, когда последнее состояние указанного пути не будет заключительным ($p \notin F$).
Следовательно, КА M_1, который допускает цепочку x тогда и только тогда, когда ее не допускает исходный КА M, можно получить, превращая каждое заключительное состояние КА M в не заключительное и наоборот.
В результате получим детерминированный КА, допускающий дополнение языка L. ▶

Доказанная теорема позволяет строить конечный автомат, не допускающий определенное множество цепочек, следующим методом:

1) строим конечный автомат, допускающий данное множество цепочек;

2) детерминизируем построенный КА;

3) строим КА для дополнения согласно методу, приведенному в доказательстве теоремы 3.

Пример 14.3. Построим конечный автомат, допускающий все цепочки в алфавите $\{0, 1\}$, кроме цепочки 101.
1. Строим конечный автомат , допускающий единственную цепочку 101.

Рис. 17

Этот автомат **квазидетерминированный**, но не детерминированный.

2. Проведем детерминизацию КА, получим ДКА, эквивалентный исходному.

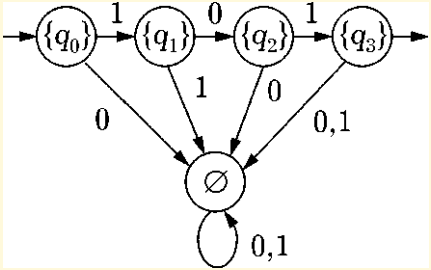

Рис. 18

3. Перейдем к дополнению ДКА и переименуем состояния. Получим ДКА, допускающий все цепочки в алфавите $\{0, 1\}$, кроме цепочки 101.

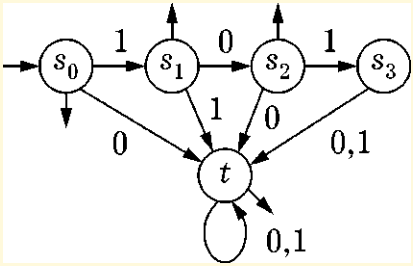

Рис. 19

В полученном ДКА все вершины, кроме s_3, - заключительные.

ДКА на рис.19 допускает все цепочки, содержащие вхождение цепочки 101, но не совпадающие с самой этой цепочкой. Например, на пути s_0, s_1, s_2, s_3, t читается цепочка 1011.

Построение дополнения языка L можно проводить только с использованием ДКА, допускающего язык L.
Если поменяем ролями заключительные и незаключительные вершины в исходном КА (рис.17), то получим КА, допускающий язык $\{\lambda, 1, 10\}$, который не является множеством **всех** цепочек, отличных от цепочки 101.

Из свойства замкнутости класса регулярных языков относительно дополнения (см. теорему 3) вытекает замкнутость этого класса относительно пересечения, теоретико-множественной и симметрической разности.

Из того, что дополнение регулярного языка есть регулярный язык, вытекают следующие утверждения.

Следствие 14.1. Для любых двух регулярных языков L_1 и L_2 справедливы следующие утверждения:
1) пересечение $L_1 \cap L_2$ регулярно;
2) разность $L_1 \setminus L_2$ регулярна;
3) симметрическая разность $L_1 \triangle L_2$ регулярна.

◄ Справедливость утверждений вытекает из тождеств:
1) $L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}}$;
2) $L_1 \setminus L_2 = L_1 \cap \overline{L_2}$;
3) $L_1 \triangle L_2 = (L_1 \cup L_2) \setminus (L_1 \cap L_2)$. ►

Полученные результаты позволяют утверждать, что множество регулярных языков замкнуто относительно операций объединения, пересечения и дополнения.

Эти свойства регулярных языков позволяет решить важную проблему распознавания **эквивалентности** двух произвольных КА.

Конечные автоматы эквивалентны , если допускаемые ими языки совпадают.

Чтобы убедиться в эквивалентности автоматов M_1 и M_2, достаточно доказать, что симметрическая разность языков $L(M_1)$ и $L(M_2)$ пуста.

Для этого необходимо построить КА, допускающий эту разность, и убедиться в том, что допускаемый им язык пуст.

В общем случае проблему распознавания того, что язык конечного автомата пуст, называют **проблемой пустоты для конечного автомата**.

Чтобы решить **проблему непустоты для КА** , достаточно найти множество заключительных состояний автомата, достижимых из начального состояния.

Так как КА — это ориентированный граф, то решить такую проблему можно, например, с помощью, поиска в ширину из начальной вершины.

Язык, допускаемый конечным автоматом, пуст тогда и только тогда, когда множество заключительных состояний, достижимых из начального состояния, пусто.

Практически эквивалентность конечных автоматов предпочтительнее распознавать, используя алгоритм **минимизации**.