Доля П.Г.
Харківський національний університет імені В. Н. Каразіна
факультет математики і інформатики
кафедра теоретичної та прикладної інформатики
2015 р.

Дискретная математика.
Конспект лекций.

Оглавление

1. Алгебра высказываний и логика.

1.1 Высказывания и логические операции

Логика — это наука о рассуждениях, которая позволяет определить истинность или ложность того или иного математического утверждения, исходя из совокупности первичных предположений, называемых аксиомами. Блоками формальной логики являются высказывания. Высказывание – это утверждение или повествовательное предложение, о котором можно сказать, что оно истинно или ложно. Истинность или ложность, приписываемые некоторому высказыванию, называются его значением. Пример высказываний:

- *земля плоская*;
- *Вася — студент*;
- *18 — четное число*.

Вот примеры предложений, не являющихся высказываниями:

- *Кто вы?* (вопрос),
- *Прочтите эту главу до следующего занятия* (приказ или восклицание),
- *Это утверждение ложно* (внутренне противоречивое утверждение).

Будем обозначать высказывания буквами латинского алфавита p, q, r, ... Например, p может обозначать утверждение: *Завтра будет дождь*, а q – утверждение: *Квадрат целого числа есть число положительное*.

В обыденной речи для образования сложного предложения из простых используются связки – особые части речи, соединяющие отдельные предложения. Наиболее часто употребляются связки и, или, нет, если ... то, только если, тогда и только тогда. В отличие от обыденной речи, в логике смысл таких связок должен быть определен однозначно. Истинность сложного высказывания однозначно определяется истинностью или ложностью составляющих его частей. Высказывание, не содержащее связок, называется простым. Высказывание, содержащее связки, называется сложным.

Чтобы при формулировке сложных высказываний не возникали неоднозначности трактовки необходимо соблюдать некоторые правила.

Правило тождественности: значения слов использованных в повествовательном предложении (высказывании) должны быть уточнены так, чтобы сказанное в нем было безусловно верным или неверным, но не тем и другим одновременно. Значение слов, сказанных в предложении, в продолжение обсуждения не меняется.

Правило исключающего третьего: Все используемые утверждения или правильные или нет – другие оценки невозможны.

Правило противоречия: утверждение и его отрицание не могут быть истинными одновременно.

Рассмотрим пример. Пусть p и q обозначают высказывания

p: *Николай водит автомобиль,*

q*: У Елены светлые волосы.*

Сложное высказывание

Николай водит автомобиль ***и*** *у Елены светлые волосы.*

состоит из двух частей, объединенных связкой И. Это высказывание может быть символически записано в виде

$$p \text{ и } q$$

или

$$p \wedge q$$

где символ $\wedge$ заменяет союз и на языке символических выражений. Выражение $p \wedge q$ называется конъюнкцией высказываний p и q или логическим умножением.

Другое высказывание

Николай водит автомобиль ***или*** *у Елены светлые волосы.*

символически выражается как

$$p \text{ или } q$$

или

$$p \vee q$$

где символ $\vee$ заменяет союз или в переводе на символический язык. Выражение $p \vee q$ называется дизъюнкцией высказываний p и q или логическим сложением.

Опровержение, или отрицание высказывания p обозначается через ~p. Например, если p есть высказывание *Николай водит автомобиль*, то ~p — это утверждение *Николай не водит автомобиль*.

Пусть есть еще одно утверждение

r: *Саше нравится программирование*.

Тогда предложение

Николай не водит автомобиль и у Елены светлые волосы или Саше нравится программирование символически запишется так $((\sim p)\wedge q)\vee r$. Наоборот, выражение $p\wedge(\sim q)\wedge r$ является символической записью высказывания

Николай водит автомобиль, у Елены волосы не светлые и Саше нравится программирование.

Рассмотрим выражение $p\wedge q$. Если некто говорит: *Николай водит автомобиль и у Елены светлые волосы*, то мы представляем Николая за рулем автомобиля и светловолосую Елену. В любой другой ситуации (например, если Елена имеет темные волосы или Николай не водит автомобиль) мы скажем, что говорящий не прав, т.е. его высказывание ложно.

Всего возможны четыре случая. Высказывание p может быть истинным (T = True = Истина) или ложным (F = False = Ложь). Независимо от значения p высказывание q также может быть истинным (T) или ложным (F). Таблица истинности перечисляет все возможные комбинации значений входящих в него простых высказываний и для каждой из них (комбинаций) сообщает результат сложного высказывания. На следующем рисунке представлены два варианта таблицы истинности для высказывания $p\wedge q$.

№	p	q	$p\wedge q$
1	T	T	T
2	T	F	F
3	F	T	F
4	F	F	F

p \ q	T	F
T	T	F
F	F	F

Если некто скажет: *Николай водит автомобиль или у Елены светлые волосы*, то он будет не прав только тогда, когда Николай не сможет управлять автомобилем, а Елена не будет светловолосой. Для того чтобы все высказывание было истинным, достаточно, чтобы одна из двух составляющих его компонент была истинной. Поэтому $p\vee q$ имеет таблицу истинности

Случай	p	q	$p\vee q$
1	T	T	T
2	T	F	T
3	F	T	T
4	F	F	F

p \ q	T	F
T	T	T
F	T	F

Таблица истинности для отрицания p имеет вид

Случай	p	~p
1	T	F
2	F	T

Истинностное значение ~p всегда противоположно значению p. В таблицах истинности отрицание всегда выполняется первым (имеет приоритет перед операциями $\wedge$ и $\vee$), поэтому высказывание $\sim p\vee q$ интерпретируется как $(\sim p)\vee q$, так что отрицание применяется только к p. Если мы хотим отрицать все высказывание $p\vee q$, то следует использовать скобки $\sim(p\vee q)$.

Символы $\wedge$ и $\vee$ называют бинарными связками (или бинарными операциями), поскольку они связывают два высказывания. Символ ~ является унарной связкой, так как применяется только к одному высказыванию.

Иногда используют еще одну бинарную связку – исключающее или, которое мы будем обозначать $\underline{\vee}$. Высказывание $p \underline{\vee} q$ истинно, когда истинно p или q, но не оба одновременно. Эта связка имеет таблицу истинности

Случай	p	q	$p \underline{\vee} q$
1	T	T	F
2	T	F	T
3	F	T	T
4	F	F	F

p \ q	T	F
T	F	T
F	T	F

В повествовательном предложении, используя союз или, мы можем иметь в виду исключающее или. Например, когда мы говорим, *Сергей сдаст экзамен по дискретной математике или не сдаст этот экзамен*, мы, конечно, предполагаем, что Сергей сделает что-то одно. В этом примере каждое из простых высказываний

p: *Сергей сдаст экзамен по дискретной математике*

q: *Сергей не сдаст экзамен по дискретной математике*

рассматриваемых автономно (без рассмотрения другого) может принимать значение ИСТИНА, но результат сложного высказывания **p** или **q** будет ЛОЖЬ. Поэтому в ячейке таблицы истинности, соответствующей значениям **p** = T, **q** = T, мы должны поставить F (ЛОЖЬ).

Аналогично, можно составить сложное высказывание

На улице идет дождь или на улице дождь не идет

которое мы можем записать символически $p \underline{\vee} q$ с использованием исключающего или. Здесь элементарными высказываниями являются предложения

p: *На улице идет дождь*

q: *На улице нет дождя*

Но высказывание

На улице идет дождь или на улице снег не идет

мы должны записать символически $p \vee q$, где

p: *На улице идет дождь*

q: *На улице нет снега*

Поскольку в случае истинности обоих простых высказываний сложное высказывание будет также истинным, то следует использовать обычное или.

Рассмотрим высказывание: *Николай сдаст экзамен по дискретной математике или Николай будет исключен из университета и пойдет работать.*

Для простых высказываний введем обозначения

p: *Николай сдаст экзамен по дискретной математике*

q: *Николай останется студентом*

r: *Николай пойдет работать*
Тогда наше сложное высказывание можно представить в виде

$$p \vee ((\sim q) \wedge r),$$

где скобки использованы, чтобы показать, какие именно высказывания являются компонентами каждой операции.

Таблица истинности дает возможность однозначно указать те ситуации, когда высказывание $p \vee ((\sim q) \wedge r)$ является истинным; при этом мы должны быть уверены, что учтены все случаи. Поскольку сложное высказывание содержит три простых высказывания p, q и r, то возможны восемь вариантов

Случай	p	q	r	~q	$(\sim q) \wedge r$	$p \vee ((\sim q) \wedge r)$
1	T	T	T	F	F	T
2	T	T	F	F	F	T
3	T	F	T	T	T	T
4	T	F	F	T	F	T
5	F	T	T	F	F	F
6	F	T	F	F	F	F
7	F	F	T	T	T	T
8	F	F	F	T	F	F

Для нахождения значений столбца $(\sim q) \wedge r$ мы используем столбцы для ~q и r, а также таблицу истинности для $\wedge$. Для нахождения значений столбца $p \vee ((\sim q) \wedge r)$ используем значения столбца p и значения уже заполненного столбца $(\sim q) \wedge r$.

1.2 Условные высказывания.

Допустим, некто утверждает, что если случится одно событие, то случится и другое. Предположим, декан говорит студенту: *Если ты сдашь все экзамены на «отлично», то получишь повышенную стипендию*. Высказывание имеет вид: если p, то q, где p – высказывание: *Ты сдашь все экзамены на «отлично»*, а q – высказывание: *получишь повышенную стипендию*. Такие сложные высказывания обозначают символически $p \Rightarrow q$ (или $p \to q$) и называются импликацией. Спрашивается, при каких условиях декан говорит правду? Если проанализировать все случаи, а их четыре, то единственный случай, когда высказывание декана ложно - это когда он дал обещание и не выполнил его. Таким образом, таблица истинности для высказывания $p \Rightarrow q$ имеет вид

№	p	q	$p \Rightarrow q$
1	T	T	T
2	T	F	F
3	F	T	T
4	F	F	T

Может показаться, что $p \Rightarrow q$ носит характер причинно – следственной связи, но это не является необходимым. Чтобы увидеть отсутствие причины и следствия в импликации, вернемся к примеру, в котором p есть высказывание

Николай водит автомобиль, а q – утверждение *У Елены светлые волосы.* Тогда утверждение *Если Николай водит автомобиль, то у Елены светлые волосы* запишется как

если p, то q

или символически

$$p \Rightarrow q$$

То, что *Николай водит автомобиль,* никак причинно не связано с тем, что Елена светловолосая. Истинность или ложность сложного высказывания зависит только от истинности составляющих его частей и не зависит от наличия или отсутствия между ними какой-либо связи.

Таким образом, импликацией называется формализация сложного союза если … то … , который записывается в виде $p \Rightarrow q$. Знак $\Rightarrow$ называется знаком импликации, или условной связкой.

Условные высказывания могут выражаться в виде различных языковых конструкций, но символически все они записываются как $p \Rightarrow q$. Вот несколько примеров таких конструкций: *если p, то q ; p достаточно для q; p является достаточным условием для q; q необходимо для p; q является необходимым условием для p; p, только если q (или только если q, то p).*

Таблица для $p \Rightarrow q$ показывает, что если $p \Rightarrow q$ истинно и p истинно, тогда q должно быть истинным, т.е. истинность p достаточна для истинности q. Поэтому p достаточно для q имеет тот же смысл, что и $p \Rightarrow q$. Аналогично, если q ложно и q необходимо для p, тогда p должно быть ложно. Поэтому, если ~q истинно, тогда ~p должно быть истинно и $\sim q \Rightarrow \sim p$. Следовательно, q необходимо для p имеет то же значение, что и $\sim q \Rightarrow \sim p$. Но это эквивалентно $p \Rightarrow q$. Действительно выражения $p \Rightarrow q$, $\sim q \Rightarrow \sim p$, имеют одинаковые таблицы истинности

p	q	$p \Rightarrow q$	~q	~p	$\sim q \Rightarrow \sim p$
F	F	T	T	T	T
F	T	T	F	T	T
T	F	F	T	F	F
T	T	T	F	F	T

Логические выражения, имеющие различное строение, но являющиеся истинными в одних и тех же случаях, называются эквивалентными. Т.о. выражения $p \Rightarrow q$ и $\sim q \Rightarrow \sim p$ эквивалентны.

Анализ значения p только если q проводится аналогично. Получаем, что p может быть истинным, только если q истинно. Если q не истинно, то p не может быть истинным. Но это эквивалентно утверждению, что если ~q истинно, то ~p должно быть истинно и $\sim q \Rightarrow \sim p$. Поэтому p только если q имеет то же значение, что и $p \Rightarrow q$.

С импликацией $p \Rightarrow q$ связаны три типа высказываний: конверсия, инверсия и контрапозиция. Они определяются следующим образом:

$p \Rightarrow q$ импликация;

$q \Rightarrow p$ конверсия высказывания $p \Rightarrow q$;

$\sim p \Rightarrow \sim q$ инверсия высказывания $p \Rightarrow q$;

$\sim q \Rightarrow \sim p$ контрапозиция высказывания $p \Rightarrow q$.

Например, пусть дано высказывание – импликация: *Если он играет в футбол, то он популярен*. Для этой импликации имеем

конверсия: *Если он популярен, то он играет в футбол*

инверсия: *Если он не играет в футбол, то он не популярен*

контрапозиция: *Если он не популярен, то он не играет в футбол*

Таблица истинности для выражений $p \Rightarrow q$ и $\sim q \Rightarrow \sim p$, построенная выше, говорит о том, что импликация и ее контрапозиция эквивалентны.

В качестве примера найдем таблицу истинности для выражения $(p \Rightarrow q) \wedge (q \Rightarrow r)$. Используя таблицу истинности для $\Rightarrow$, приведенную выше, построим сначала таблицы (столбцы) истинности для $p \Rightarrow q$ и $q \Rightarrow r$, учитывая, что импликация ложна только в случае, когда $T \Rightarrow F$. Затем заполняем столбец $(p \Rightarrow q) \wedge (q \Rightarrow r)$

№	p	q	r	$p \Rightarrow q$	$q \Rightarrow r$	$(p \Rightarrow q) \wedge (q \Rightarrow r)$
1	T	T	T	T	T	T
2	T	T	F	T	F	F
3	T	F	T	F	T	F
4	T	F	F	F	T	F
5	F	T	T	T	T	T
6	F	T	F	T	F	F
7	F	F	T	T	T	T
8	F	F	F	T	T	T

Высказывание вида $(p \Rightarrow q) \wedge (q \Rightarrow p)$ обозначается через $p \Leftrightarrow q$ (или $p \leftrightarrow q$) и называется эквиваленцией. Очевидно, таблица истинности для $(p \Rightarrow q) \wedge (q \Rightarrow p)$ определяет таблицу истинности для $p \Leftrightarrow q$.

№	p	q	$p \Rightarrow q$	$q \Rightarrow p$	$(p \Rightarrow q) \wedge (q \Rightarrow r)$
1	T	T	T	T	T
2	T	F	F	T	F
3	F	T	T	F	F
4	F	F	T	T	T

Непосредственно из определения вытекает, что эквиваленция истинна только в случае, когда p и q имеют одинаковые значения. Эквиваленция является формализацией сложного союза `тогда и только тогда`. Если есть два высказывания p и q, то высказывание `p тогда и только тогда когда q` записывается в символьном виде $p \Leftrightarrow q$, а знак $\Leftrightarrow$ называется знаком эквиваленции или равнозначности. Записи $p \Leftrightarrow q$ и $q \Leftrightarrow p$ означают одно и то же.

Конъюнкцию $\wedge$, дизъюнкцию $\vee$, отрицание ~, исключающее или $\underline{\vee}$, импликацию $\Rightarrow$ и эквиваленцию $\Leftrightarrow$ называют логическими союзами, логическими связками, логическими операторами или операциями. Можно сконструировать и другие логические операции, но мы здесь пока не рассматриваем. Возникает вопрос о том, как интерпретировать сложные выражения, например такие как $\sim p \vee q$, $p \wedge q \vee r$, $p \wedge q \Rightarrow r$, $p \wedge q \Leftrightarrow q \vee r$, в которых отсутствуют скобки. Во избежание неоднозначности лучше всегда использовать скобки. Однако в логике, как и в алгебре, имеется приоритет выполнения операций. Они выполняются в следующей последовательности: $\sim, \wedge, \vee, \Rightarrow, \Leftrightarrow$. Поэтому указанные выражения следует интерпретировать следующим образом: $(\sim p) \vee q$, $(p \wedge q) \vee r$, $(p \wedge q) \Rightarrow r$, $(p \wedge q) \Leftrightarrow (q \vee r)$.

1.3 Основные теоремы алгебры логики.

Особый интерес представляют сложные высказывания, имеющие различное строение, но являющиеся истинными в одних и тех же случаях. Как говорилось ранее, такие высказывания называются логически эквивалентными. Эквивалентность двух высказываний легко установить посредством сравнения их таблиц истинности. Например, пусть p и q обозначают высказывания

p : *Сегодня шел дождь,*

q : *Сегодня шел снег.*

Рассмотрим сложные высказывания

Неверно, что сегодня шел дождь или снег,

или символически

$$\sim(p \vee q)$$

и

Сегодня не шел дождь и сегодня не шел снег,

или символически

$$\sim p \wedge \sim q$$

Построим таблицы истинности для обоих высказываний.

Случай	p	q	$\sim(p \vee q)$	$\sim p \wedge \sim q$
1	T	T	F	F
2	T	F	F	F
3	F	T	F	F
4	F	F	T	T

Итак, во всех четырех строках истинностные значения для $\sim(p \vee q)$ и для $\sim p \wedge \sim q$ совпадают. Это означает, что два рассматриваемых высказывания логически эквивалентны, т.е.

$$\sim(p \vee q) \equiv \sim p \wedge \sim q$$

ТЕОРЕМА (законы логики). Используя таблицы истинности, можно доказать следующие логические эквивалентности:

а) Свойства коммутативности

$$p \wedge q \equiv q \wedge p;$$
$$p \vee q \equiv q \vee p.$$

б) Свойства ассоциативности

$$p \wedge (q \wedge r) \equiv (p \wedge q) \wedge r;$$
$$p \vee (q \vee r) \equiv (p \vee q) \vee r.$$

в) Свойства дистрибутивности

$$p \wedge (q \vee r) \equiv (p \wedge q) \vee (p \wedge r);$$
$$p \vee (q \wedge r) \equiv (p \vee q) \wedge (p \vee r).$$

г) Законы идемпотентности

$$p \wedge p \equiv p;$$
$$p \vee p \equiv p.$$

д) Закон двойного отрицания

$$\sim(\sim p) \equiv p$$

е) Законы де Моргана

$$\sim(p \vee q) \equiv \sim p \wedge \sim q;$$
$$\sim(p \wedge q) \equiv \sim p \vee \sim q.$$

ж) Закон контрапозиции

$$p \Rightarrow q \equiv \sim q \Rightarrow \sim p,$$

з) Другие полезные свойства

$$p \Rightarrow q \equiv \sim p \vee q$$
$$p \Leftrightarrow q \equiv (p \Rightarrow q) \wedge (q \Rightarrow p)$$

В предыдущем параграфе мы показали, что импликация и ее контрапозиция (пункт *ж*) эквивалентны. В начале этого параграфа был доказан закон де Моргана (пункт e_1). Другие пункты доказываются аналогично.

Высказывание, истинное во всех случаях, называется логически истинным, или тавтологией. Например, в прямоугольном треугольнике *сумма квадратов длин катетов равна квадрату длины гипотенузы*. Теоремы в математике являются примерами тавтологий. Высказывание *Он сдаст или не сдаст экзамен* есть пример тавтологии, поскольку либо одно событие, либо другое обязательно должно иметь место. Высказывание *Быть или не быть*, это и есть тавтология. Каждое высказывание вида $p \vee \sim p$ — тавтология.

Высказывание, построенное так, что оно ложно в любом случае, называется логически ложным, или противоречием. Высказывание *Он сдаст зачет и не сдаст зачет* всегда ложно. Следовательно, это противоречие.

Рассмотрим высказывание вида

$$(p \wedge (p \Rightarrow q)) \Rightarrow q$$

Ему соответствует таблица истинности

№	p	q	$p \Rightarrow q$	$p \wedge (p \Rightarrow q)$	$(p \wedge (p \Rightarrow q)) \Rightarrow q$
1	T	T	T	T	T
2	T	F	F	F	T
3	F	T	T	F	T
4	F	F	T	F	T

Все значения в столбце $(p \wedge (p \Rightarrow q)) \Rightarrow q$ равны T. Высказывание истинно во всех четырех возможных случаях, следовательно, оно является тавтологией.
Имея логически истинное высказывание-тавтологию, легко построить логически ложное высказывание - противоречие. Для этого достаточно взять отрицание логически истинного высказывания. Поэтому высказывание $\sim((p \wedge (p \Rightarrow q)) \Rightarrow q)$ логически ложно. Ложным также будет высказывание $\sim(p \vee \sim p)$.

Пусть символ T обозначает высказывание, которое есть тавтология и поэтому имеет таблицу истинности, состоящую из одних T. Символом F обозначим противоречие, т.е. высказывание, таблица истинности которого содержит F во всех строках. Используя таблицы истинности, легко проверить, что

$$\begin{array}{llll} p & \wedge & \boldsymbol{T} & \equiv p\,, \\ p & \wedge & \boldsymbol{F} & \equiv \boldsymbol{F}, \\ p & \vee & \boldsymbol{T} & \equiv \boldsymbol{T}, \\ p & \vee & \boldsymbol{F} & \equiv p\,, \\ p & \wedge & \sim p & \equiv \boldsymbol{F}, \\ p & \vee & \sim p & \equiv \boldsymbol{T}, \\ p & \Rightarrow & p & \equiv \boldsymbol{T}. \end{array}$$

В высказывании можно заменить любую его компоненту на логически эквивалентное ей высказывание. Полученное в результате такой замены высказывание будет логически эквивалентно исходному, поскольку истинностное значение высказывания зависит исключительно от истинностных значений составляющих его компонент (но не от их формы или сложности). Например,

$$\begin{array}{lll} & (q \vee r) \vee (p \wedge \sim r) \;\equiv & \\ \equiv & q \vee (r \vee (p \wedge \sim r) \equiv & \text{свойство ассоциативности} \\ \equiv & q \vee ((r \vee p) \wedge (r \vee \sim r)) \equiv & \text{свойство дистрибутивности} \\ \equiv & q \vee ((r \vee p) \wedge \boldsymbol{T}) \equiv & \text{эквивалентность} \\ \equiv & q \vee (r \vee p) \equiv & \text{эквивалентность} \\ \equiv & q \vee (p \vee r) \equiv & \text{свойство коммутативности} \\ \equiv & (q \vee p) \vee r \equiv & \text{свойство ассоциативности} \\ \equiv & (p \vee q) \vee r\,. & \text{свойство коммутативности} \end{array}$$

Здесь мы использовали законы логики для образования новых эквивалентных высказываний.

1.4 Полнота в логике высказываний.

Соберем вместе таблицы истинности для рассмотренных ранее бинарных логических операций и унарной операции отрицания.

И $p \wedge q$

p \ q	T	F
T	T	F
F	F	F

Или $p \vee q$

p \ q	T	F
T	T	T
F	T	F

Исключающее или $p \veebar q$

p \ q	T	F
T	F	T
F	T	F

Импликация $p \Rightarrow q$ Эквиваленция $p \Leftrightarrow q$ Отрицание ~p

p \ q	T	F
T	T	F
F	T	T

p \ q	T	F
T	T	F
F	F	T

p	~p
T	F
F	T

Мы также знаем таблицы истинности для тавтологии (состоит из одних T) и противоречия (состоит из одних F). Существуют ли другие бинарные логические операции. Поскольку в каждой клетке таблицы 2×2 может стоять только два значения T или F, то всего различных таблиц будет $2^4 = 16$. Каждая таблица истинности определяет свою бинарную логическую операцию и для них, конечно, придуманы свои названия и обозначения. Но мы на них останавливаться не будем, поскольку для конструирования любого сложного выражения нам всегда будет достаточно нескольких из выше приведенных логических операций.

Допустим, что задана произвольная таблица истинности. Существует простой способ найти высказывание, которому она соответствует. Например, предположим, что имеется таблица истинности

Случай	***p***	***q***	**Операция**
1	*F*	*F*	*T*
2	*F*	*T*	*F*
3	*T*	*F*	*T*
4	*T*	*T*	*T*

$\sim p \wedge \sim q$	$p \wedge \sim q$	$p \wedge q$
T	F	F
F	F	F
F	T	F
F	F	T

Известно, что $\sim p \wedge \sim q$ истинно в случае 1 и ложно во всех остальных случаях. Аналогично, $\sim p \wedge q$ истинно только в случае 2, $p \wedge \sim q$ истинно только в случае 3, а $p \wedge q$ истинно только в случае 4. Указанное в примере высказывание должно быть истинным в случаях 1, 2 и 4. Возьмем высказывания, истинные только в этих случаях, и соединим их связкой или $\vee$. Тогда мы получим высказывание, истинное только в требуемых случаях (строках). В нашем случае рассматриваемая таблица истинности соответствует высказыванию

$$(p \wedge q) \vee (p \wedge \sim q) \vee (\sim p \wedge \sim q)$$

Таблица истинности каждой из скобок этого выражения записана в столбцах справа. Операция ИЛИ между ними даст столбец заданной «операции». Конечно, полученная форма высказывания не является простейшей. Используя законы логики, почти всегда можно получить более короткую запись.

Очевидно, что каким бы ни был столбец «операции», используя выражения $\sim p \wedge \sim q$, $\sim p \wedge q$, $p \wedge \sim q$, $p \wedge q$ и комбинируя их с помощью операции ИЛИ, мы сможем получить этот столбец . Фактически мы показали, что трех логических связок $\sim, \wedge, \vee$ достаточно для представления логического выражения (двух переменных ***p*** и ***q***) с любой таблицей истинности.

В случае таблиц истинности с тремя переменными имеем аналогичную ситуацию. Для каждой строки следующей таблицы приведено высказывание, истинное только для этой строки.

Случай	p	q	r	
1	T	T	T	$p \wedge q \wedge r$
2	T	T	F	$p \wedge q \wedge \sim r$
3	T	F	T	$p \wedge \sim q \wedge r$
4	T	F	F	$p \wedge \sim q \wedge \sim r$
5	F	T	T	$\sim p \wedge q \wedge r$
6	F	T	F	$\sim p \wedge q \wedge \sim r$
7	F	F	T	$\sim p \wedge \sim q \wedge r$
8	F	F	F	$\sim p \wedge \sim q \wedge \sim r$

Если требуется построить высказывание, соответствующее конкретной таблице истинности, необходимо выбрать выражения, соответствующие случаям (строкам), где высказывание истинно, и соединить их связкой или$\vee$. Например, пусть дана таблица истинности

p	q	r	
T	T	T	T
T	T	F	F
T	F	T	F
T	F	F	F
F	T	T	T
F	T	F	F
F	F	T	T
F	F	F	T

Тогда соответствующее ей высказывание может иметь вид

$$(p \wedge q \wedge r) \vee (\sim p \wedge q \wedge r) \vee (\sim p \wedge \sim q \wedge r) \vee (\sim p \wedge \sim q \wedge \sim r)$$

Заметим, что в п.1.2 мы строили таблицу истинности для высказывания $(p \Rightarrow q) \wedge (q \Rightarrow r)$, которая была такой же, как в нашем примере. Т.о. для логического выражения $(p \Rightarrow q) \wedge (q \Rightarrow r)$ мы построили эквивалентное высказывание без использования значка (логической связки) импликации $\Rightarrow$.

Рассмотрим вопрос о минимальном количестве логических связок, необходимых для представления любого высказывания, образованного с помощью определенных нами выше логических связок. Известно, что $p \Leftrightarrow q$ можно выразить как $(p \Rightarrow q) \wedge (q \Rightarrow p)$, т.е. $p \Leftrightarrow q \equiv (p \Rightarrow q) \wedge (q \Rightarrow p)$. Так что использовать операцию $\Leftrightarrow$ удобно, но не необходимо. К тому

$$p \veebar q \equiv (p \wedge \sim q) \vee (\sim p \wedge q).$$

$$p \Rightarrow q \equiv \sim p \vee q$$

Поэтому нет необходимости использовать $\veebar$ и $\Rightarrow$, если применяется $\sim$ и $\vee$.

Выше мы показали, что трех логических связок $\sim, \wedge, \vee$ достаточно для представления любого логического выражения (с любой таблицей истинности). Но поскольку

$$p \wedge q \equiv \sim(\sim p \vee \sim q)$$
$$p \vee q \equiv \sim(\sim p \wedge \sim q),$$

то можно обойтись парой связок $\sim, \wedge$ или парой $\sim, \vee$, причем в любом случае необходимы обе связки. Однако существуют две связки, обладающие тем свойством, что любое высказывание может быть выражено с использованием только одной из них. Эти связки: $|$ — штрих Шеффера и $\downarrow$ — стрелка Пирса. Этим связкам соответствуют таблицы истинности

Случай	p	q	$p \mid q$
1	T	T	F
2	T	F	T
3	F	T	T
4	F	F	T

Случай	p	q	$p \downarrow q$
1	T	T	F
2	T	F	F
3	F	T	F
4	F	F	T

Для того чтобы показать, что любую связку можно заменить связкой $|$, достаточно показать это для пар связок $\sim$ и $\wedge$ или $\sim$ и $\vee$, поскольку возможность выразить любую связку одной из этих пар уже показана. Эквивалентность $p \mid p \equiv \sim p$ устанавливается при помощи следующей таблицы истинности:

№	p	$p \mid p$
1	T	F
4	F	T

Точно так же таблица

№	p	q	$p \mid p$	$q \mid q$	$(p \mid p) \mid (q \mid q)$
1	T	T	F	F	T
2	T	F	F	T	T
3	F	T	T	F	T
4	F	F	T	T	F

показывает, что $(p \mid p) \mid (q \mid q) \equiv p \vee q$. Аналогично можно показать, что $(p \mid q) \mid (p \mid q) \equiv p \wedge q$.

Проверив, что $\sim$ и $\wedge$ или $\sim$ и $\vee$ можно выразить, используя только операцию $|$, мы доказали, что любую связку можно выразить, используя лишь штрих Шеффера $|$.

Проверьте самостоятельно, что

$$p \downarrow p \equiv \sim p$$
$$(p \downarrow q) \downarrow (p \downarrow q) \equiv p \vee q$$
$$(p \downarrow p) \downarrow (q \downarrow q) \equiv p \wedge q$$

Заметим, что $p \mid q$ эквивалентно $\sim(p \wedge q)$, а $p \downarrow q$ эквивалентно $\sim(p \vee q)$. Поэтому связка $|$ называется не – и, а связка $\downarrow$ называется не – или.

2. Алгебра множеств.

Слово «дискретный» означает «составленный из отдельных частей», а дискретная математика имеет дело с совокупностями объектов, называемых множествами. Элементы этих множеств, как правило, изолированы друг от друга, геометрически не связаны и различимы. Большинство рассматриваемых в курсе дискретной математики множеств конечны или счетны. При этом элементы множеств могут иметь различную природу – быть битами или байтами, высказываниями, геометрическими объектами, функциями, людьми, оттенками цвета и т.д. В данной главе мы займемся изучением языка описания элементов множеств и операций с ними.

2.1 Понятие множества

Множество – это любая определенная совокупность объектов. Объекты, из которых составлено множество, называются его элементами. Элементы множества различны и отличимы друг от друга.

Примеры.

Множество страниц в учебнике.

Множество студентов в аудитории.

Множество **N** натуральных чисел 1, 2, 3,…

Множество **Z** целых чисел: …, -2, -1, 0, 1, 2, …

Множество **R** вещественных чисел.

Множество $\mathbf{Q} = \left\{ \frac{p}{q} : p, q \in \mathbf{Z}, q \neq 0 \right\}$ – множество рациональных чисел.

Если объект *x* является элементом множества М, то говорят, что *x* принадлежит М. Этот факт записывают $x \in M$. В противном случае говорят, что *x* не принадлежит М и записывают $x \notin M$.

Пример. $3 \in \{1, 3, 5\}$, $4 \notin \{1, 3, 5\}$.

Обычно множества мы будем обозначать прописными буквами латинского алфавита, а элементы множеств – строчными. Множества, как объекты, могут быть элементами других множеств.

Чтобы задать множество, нужно указать какие элементы ему принадлежат. Это можно сделать различными способами:

- перечислением элементов: $M := \{ a_1, a_2, ..., a_k \}$. При таком задании множества обозначения элементов заключают в фигурные скобки и разделяют запятыми. Перечислением можно задавать только конечные множества.
- характеристическим предикатом: $M := \{ x \mid P(x) \}$. Характеристический предикат – это условие, выраженное в форме логического утверждения. Если оно истинно для данного элемента *x*, то он принадлежит множеству М, если ложно – то не принадлежит.
- порождающей процедурой $M := \{ x \mid x := f \}$, которая, будучи запущенной, генерирует (создает) объекты, являющиеся элементами множества М.

Например, запись S = $\{x : x$ — нечетное натуральное число$\}$ описывает множество S = {1, 3, 5, 7, . . . } с помощью предиката «x — нечетное натуральное число». Предикат – это логическое выражение с аргументом. Точнее, предикат становится логическим выражением, когда его аргумент заменяют конкретным элементом.

Примеры.

$M_9 = \{1,2,3,4,5,6,7,8,9\}$

$\{x : x$ — *футболист, играющий за Металлист*$\}$ — множество, состоящее из всех футбольных игроков харьковского футбольного клуба Металлист.

$\{x : x$ – *гражданин Украины*$\}$ описывает множество всех граждан Украины.

$C = \{1, 4, 9, 16, .. n^2\}$ описывает множество квадратов всех положительных чисел, которые меньше или равны n.

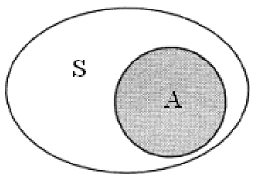
В современных языках программирования переменные объявляются как принадлежащие к определенному типу данных. Тип данных представляет собой множество объектов со списком определенных операций над ними. Задание типа переменных равносильно указанию множества, из которого переменным присваиваются значения.

Поскольку множество однозначно определяется только элементами, которые оно содержит, порядок их перечисления роли не играет. Например, {1,2,4,6} = {2,1,6,4}. Любой элемент либо принадлежит данному множеству, либо нет. Каждый элемент может входить в множество не более одного раза.

Множество, не содержащее элементов, называется пустым и обозначается $\varnothing$. Универсальное множество U (или универсум) есть множество, обладающее тем свойством, что элементы всех рассматриваемых множеств являются его подмножествами. В известном смысле $\varnothing$ и U представляют собой противоположности, поскольку пустое множество не содержит элементов, а универсальное множество содержит «все» элементы.

В теории чисел универсальное множество обычно совпадает с множеством всех целых или натуральных чисел. В математическом анализе универсальное множество может быть множеством всех действительных чисел или множеством всех точек n-мерного пространства. Следует отметить, что универсальное множество U, хотя, и названо универсальным, однозначно не определено, если точно не указана предметная область.

Говорят, что множество A является подмножеством множества S, если каждый его элемент в то же время является элементом множества S. При этом говорят, что множество A содержится в множестве S. Этот факт обозначают так: $A \subset S$. На следующем рисунке дана иллюстрация этого определения. Такого сорта картинки называются диаграммами Эйлера – исходные множества изображаются овалами (или кругами), а результат графически выделяется штриховкой.



Два множества считаются равными, если каждое из них содержится в другом

$$A = B := A \subset B \wedge B \subset A$$

Для доказательства равенства множеств нужно показать, что они состоят из одних и тех же элементов. На формальном языке для равенства множеств A = B необходимо проверить истинность двух импликаций:

$$\{x \in A \Rightarrow x \in B\} \text{ и } \{x \in B \Rightarrow x \in A\}$$

Если $A \subset B$ и $A \neq B$, то A называется собственным подмножеством B. Если может оказаться, что $A = B$, то пишут $A \subseteq B$.

По определению, каждое множество есть подмножество универсального множества $M \subset U$.

Пустое множество есть подмножество любого множества M, т.е. $\varnothing \subset M$, поскольку «каждый элемент пустого множества содержится в M». Точнее, не существует элементов пустого множества, которые не принадлежали бы M.

Пустое множество единственно, т.е. любые два пустых множества M и N совпадают. Действительно, пусть $M \neq N$. Это означает, что существует элемент, который принадлежит одному множеству и не принадлежит другому. Но тогда то множество, которому принадлежит элемент, будет непустым. Получили противоречие.

2.2 Операции над множествами

Рассмотренные ниже операции над множествами позволяют строить новые множества, используя уже существующие.

Пересечением множеств A и B называется множество, состоящее из тех и только тех элементов, которые одновременно принадлежат как A, так и B. Пересечение множеств A и B обозначается $A \cap B$. Это определение равносильно следующему: $A \cap B = \{x : x \in A \wedge x \in B\}$.

Например, если A = {1, 2, 3, 4, 5} и B = {1, 3, 5, 7, 9}, то $A \cap B = \{1, 3, 5\}$.

Если A = {x : x имеет рост выше 180см} и B = {x : x любит играть в шахматы}, то $A \cap B$ – {x : x имеет рост выше 180см и любит играть в шахматы}. Обратите внимание, что в описании пересечения множеств $A \cap B$ использована связка «и». В дальнейшем мы убедимся, что символы $\cap$ и $\wedge$ связаны между собой и имеют схожие свойства.

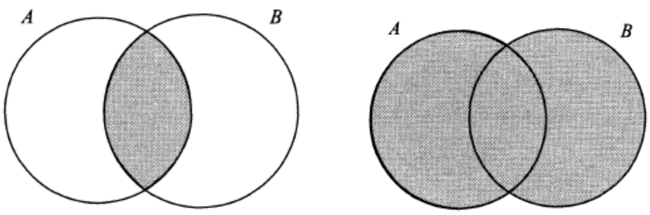
Пересечение трех множеств A_1, A_2, A_3 можно определить так $B = A_1 \cap (A_2 \cap A_3)$. Пересечение n множеств $A_1, \ldots A_n$ обозначают $\bigcap_{i=1}^{n} A_i$ и определяют $A_1 \cap (A_2 \cap (\ldots (A_{n-1} \cap A_n)))$. Иногда используют обозначение $\bigcap_{i \in I} A_i$, где индекс i принадлежит некоторому конечному множеству I.

Объединением множеств A и B называется множество, состоящее из всех тех элементов, которые принадлежат хотя бы одному из множеств A или B.

Объединение множеств A и B обозначается $A \bigcup B$. Это определение равносильно следующему: $A \bigcup B = \{x : x \in A \vee x \in B\}$

Например, если A = {1, 2, 6, 7,} и B = {2, 3, 5, 6}, то $A \bigcup B = \{1, 2, 3, 5, 6, 7\}$.

На следующем рисунке приведены диаграммы Эйлера для пересечения (слева) и объединения (справа) множеств A и B.

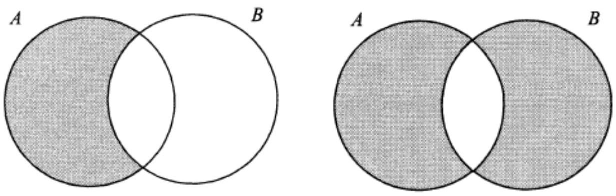


Объединение n множеств $A_1, \dots A_n$ обозначают $\bigcup_{i=1}^{n} A_i$ и определяют $A_1 \cup (A_2 \cup (\dots (A_{n-1} \cup A_n)))$. Иногда используют обозначение $\bigcup_{i \in I} A_i$, где индекс i принадлежит некоторому конечному множеству I.

Пусть A и B множества. **Разностью** множеств A \ B называется множество тех и только тех элементов A, которые не содержатся в B. Это определение равносильно следующему: $A \setminus B = \{ x : x \in A \wedge x \notin B \}$

Симметрической разностью двух множеств A и B, обозначаемой $A \Delta B$, называют множество $(A \setminus B) \bigcup (B \setminus A)$ или, по – другому, $A \Delta B = \{ x : (x \in A \wedge x \notin B) \vee (x \in B \wedge x \notin A) \}$. Оно состоит из всех тех и только тех элементов, которые либо принадлежат A и не принадлежат B , либо наоборот, принадлежат B, но не A.

На следующем рисунке приведены диаграммы Эйлера для разности A \ B множеств (слева) и для симметрической разности $A \Delta B$ (справа).



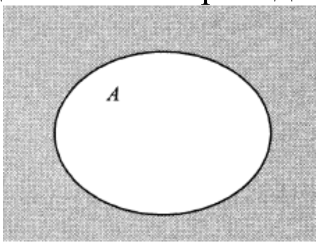
Например, если A = {1, 2, 4, 6, 7}, а B = {2, 3, 4, 5,6}, то A \ B = {1,7} , а $A \Delta B$ есть множество {1, 3, 5, 7}.

Обратите внимание на сходство симметрической разности со связкой «исключающее или» из главы 1.

Дополнение множества A, обозначаемое $\overline{A}$ – это множество элементов универсума, которые не принадлежат A. Следовательно, $\overline{A} = U \setminus A = \{ x : x \in U \wedge x \notin A \}$.

При определении дополнения мы предполагаем, что все элементы множеств обладают некоторой природой, т.е. принадлежат некоторому всеобъемлющему множеству, которое называется универсум. Операция дополнения всегда подразумевает некоторый универсум $U : \overline{A} = U \setminus A$. Например, если U — множество положительных целых чисел, а A = {2, 4, 6, 8, ...} – множество всех четных положительных чисел, то $\overline{A} = \{1, 3, 5, 7, \ldots\}$ – множество всех нечетных положительных чисел.

Диаграмма Эйлера для дополнения приведена на следующем рисунке.



Пример. Пусть A={1, 2,3}, B={3,4,5}. Тогда

$$A \bigcup B = \{1, 2, 3, 4, 5\},\ A \bigcap B = \{3\},\ A \setminus B = \{1, 2\},\ A \Delta B = \{1, 2, 4, 5\}$$

Декартово произведение множеств A и B, обозначаемое A x B, есть множество пар элементов $(a,b): a \in A \wedge b \in B$. Объект (a,b) называется упорядоченной парой с первой компонентой *a* и второй компонентой *b*. Порядок элементов в паре существенен!

Если A пусто или B пусто, то, по определению, A x B пусто.

В том случае, когда каждое из множеств A_1, A_2,..., A_n совпадает с множеством A, мы пишем A^n для обозначения декартового (прямого) произведения n экземпляров множества A. A^n называется n – ой степенью множества A и $A^n : \underbrace{A \times A \times \ldots \times A}_{n\ раз}$. Соответственно, $A^1 := A$, $A^2 = A \times A$ и вообще $A^n := A \times A^{n-1}$.

В качестве следующего примера рассмотрим прямое произведение множества вещественных чисел на само себя. Множество **R** x **R** или $\mathbf{R}^2$ состоит из всех упорядоченных пар вещественных чисел (x, y). Их можно представлять себе как координаты точек на плоскости. Метод координат ввел Рене Декарт (1596 – 1650), отсюда название «декартово произведение». Множество $\mathbf{R}^2$ называется декартовой плоскостью.

Пусть B={0, 1}. Тогда множество B^n состоит из последовательностей нулей и единиц длины n. Они называются *строкой бит* или *битовой строкой* длины n.

Пример. Пусть A={x, y} и B={1,2,3}. Тогда

$$A \times B = \{(x,1), (x,2), (x,3), (y,1), (y,2), (y,3)\}$$

$$B \times A = \{(1,x), (2,x), (3,x), (1,y), (2,y), (3,y)\}$$

$$B \times B = \{(1,1), (1,2), (1,3), (2,1), (2,2), (2,3), (3,1), (3,2), (3,3),\}$$

Заметим, что множества $A \times B$ и $B \times A$ различны!

□

Как и в случае предыдущих операций на множествах мы можем нарисовать диаграмму Эйлера, иллюстрирующую прямое произведение. Для предыдущего примера диаграмма Эйлера для прямого произведения имеет следующий вид

$$\begin{array}{cc} & \boldsymbol{B} \\ & \begin{pmatrix} \bullet 1 & \bullet 2 & \bullet 3 \end{pmatrix} \\ \boldsymbol{A}\begin{pmatrix} \bullet x \\ \bullet y \end{pmatrix} & \begin{pmatrix} \bullet(x,1) & \bullet(x,2) & \bullet(x,3) \\ \bullet(y,1) & \bullet(y,2) & \bullet(y,3) \end{pmatrix} \boldsymbol{A}\times\boldsymbol{B} \end{array}$$

В заключение нашего обсуждения множеств мы покажем, как строка бит применяется для моделирования операций на конечных множествах. Пусть $S=\{s_1, s_2,\ldots,s_n\}$ есть некоторое множество, элементы которого мы пометили числовыми индексами для удобства ссылок. Если $A\subset S$, мы поставим ему в соответствие `n`-битную строку $(b_1,b_2,\ldots,b_n)$, где $b_i=1$, если $s_i\in A$ и $b_i=0$ в противном случае. Такая строка бит называется характеристическим вектором подмножества `A`. Теперь мы можем имитировать операции на множествах логическими операциями, применяемыми к соответствующим характеристическим векторам, условившись считать 1 за `Истину`, а 0 за `Ложь`. При этом будем считать, что имеет место соответствие логических и «битовых» операций, т.е справедливы соотношения $1\wedge 1=1, 1\wedge 0=0,\ldots$ $1\vee 1=1, 1\vee 0=1,\ldots$, $\sim 1=0, \sim 0=1$.

Пример. Пусть `S={1,2,3,4,5}`, `A={1,3,5}` и `B={3,4}`. Характеристическим вектором множества `A` является *a*=`(1,0,1,0,1)`, а характеристический вектор множества `B` равен *b*=`(0,0,1,1,0)`. Тогда

$$a\vee b=(1,0,1,0,1)\vee(0,0,1,1,0)=(1,0,1,1,1)$$
$$a\wedge b=(1,0,1,0,1)\wedge(0,0,1,1,0)=(0,0,1,0,0)$$
$$\sim a\ =\sim(1,0,1,0,1)=(0,1,0,1,0)$$

Полученные векторы позволяют нам «прочитать» элементы требуемых множеств: $A\bigcup B=\{1,3,4,5\}$, $A\bigcap B=\{\ 3\ \}$ и $\overline{A}=\{2,4\}$.

2.3 Алгебра множеств

Из определения операций, о которых шла речь в предыдущем параграфе, можно вывести многочисленные свойства. Доказательства будут основываться на соответствии между операциями на множествах и логическими операциями.

Теорема. Для произвольных множеств `A` и `B` справедливо равенство $A\setminus B=A\bigcap\overline{B}$.

Доказательство. Для доказательства равенства двух множеств нужно показать, что каждое из множеств является подмножеством другого. Это можно осуществить, выбирая произвольный элемент одного множества и доказывая, что он принадлежит другому множеству. Имеем

$$a\in A\setminus B \Leftrightarrow (a\in A)\wedge(a\notin B)\Leftrightarrow \ \textit{определение разности } A\setminus B$$
$$\Leftrightarrow (a\in A)\wedge(a\in\overline{B})\Leftrightarrow \ \textit{определение дополнения}$$
$$\Leftrightarrow a\in(A\bigcap\overline{B}) \qquad \textit{определение пересечения}$$

Проверка обратного включения выполняется прочтением предыдущих выражений в обратном порядке. □

Теорема. Для произвольных множеств A и B справедливы равенства

$$a)\ \overline{A\cap B}=\overline{A}\cup\overline{B}$$

$$b)\ \overline{A\cup B}=\overline{A}\cap\overline{B}$$

Докажем a). Имеем

$$a\in\overline{A\cap B}\Leftrightarrow a\notin A\cap B\Leftrightarrow \qquad \textit{определение дополнения}$$

$$\Leftrightarrow \sim(a\in A\cap B)\ \Leftrightarrow \qquad \textit{определение } \notin$$

$$\Leftrightarrow \sim((a\in A)\wedge(a\in B))\Leftrightarrow \qquad \textit{определение пересечения}$$

$$\Leftrightarrow \sim(a\in A)\vee\sim(a\in B)\Leftrightarrow \textit{закон логики де Моргана}$$

$$\Leftrightarrow (a\notin A)\vee(a\notin B)\Leftrightarrow \textit{определение } \notin$$

$$\Leftrightarrow \left(a\in\overline{A}\right)\vee\left(a\in\overline{B}\right)\Leftrightarrow \textit{определение дополнения}$$

$$\Leftrightarrow a\in\left(\overline{A}\cup\overline{B}\right) \qquad \textit{определение объединения}$$

Проверка обратного включения выполняется прочтением предыдущих выражений в обратном порядке. Доказательство равенства b) выполняется аналогично. □

Обратите внимание, что при доказательстве предыдущей теоремы (законы де Моргана в теории множеств) мы использовали законы де Моргана в логике.

Теорема. Для произвольных множеств A и B справедливы равенства

$$a)\ A\cap(B\cup C)=(A\cap B)\cup(A\cap C);$$

$$б)\ A\cup(B\cap C)=(A\cup B)\cap(A\cup C).$$

Доказательство выполняется аналогично предыдущей теореме. Цепочка эквивалентных выражений для случая a) имеет вид

$$a\in A\cap(B\cup C)\Leftrightarrow$$

$$\Leftrightarrow (a\in A)\wedge(a\in(B\cup C))\Leftrightarrow \qquad \text{определение пересечения}$$

$$\Leftrightarrow (a\in A)\wedge((a\in B)\vee(a\in C))\Leftrightarrow \qquad \text{определение объединения}$$

$$\Leftrightarrow ((a\in A)\wedge(a\in B))\vee \qquad \text{логический}$$

$$\vee((a\in A)\wedge(a\in C))\Leftrightarrow \qquad \text{закон дистрибутивности}$$

$$\Leftrightarrow (a\in(A\cap B))\vee(a\in(A\cap C))\Leftrightarrow \qquad \text{определение пересечения}$$

$$\Leftrightarrow a\in((A\cap B)\cup(A\cap C)) \qquad \text{определение объединения}$$

Проверка обратного включения выполняется прочтением предыдущих выражений в обратном порядке. Здесь при доказательстве закона дистрибутивности в теории множеств мы использовали логический закон дистрибутивности. Доказательство равенства *б*) выполняется аналогично.

□

Теоремы, приведенные выше, показывают, что равенства, доказанные в теории множеств, имеют свой аналог в логике. Законы для множеств, сходные с законами логики, перечислены ниже. Каждое из равенств может быть доказано с помощью логических выражений, аналогичных тем, что использованы выше.

Закон ассоциативности

$$A\bigcup(B\bigcup C)=(A\bigcup B)\bigcup C$$
$$A\bigcap(B\bigcap C)=(A\bigcap B)\bigcap C$$

Закон коммутативности

$$A\bigcup B=B\bigcup A$$
$$A\bigcap B=B\bigcap A$$

Закон дистрибутивности

$$A\bigcap(B\bigcup C)=(A\bigcap B)\bigcup(A\bigcap C)$$
$$A\bigcup(B\bigcap C)=(A\bigcup B)\bigcap(A\bigcup C)$$

Законы тождества

$$A\bigcup\varnothing=A \qquad A\bigcap U=A$$
$$A\bigcup U=U \qquad A\bigcap\varnothing=\varnothing$$

Законы идемпотентности

$$A\bigcup A=A \qquad A\bigcap A=A$$

Законы дополнения

$$A\bigcup\overline{A}=U \qquad A\bigcap\overline{A}=\varnothing$$
$$\overline{U}=\varnothing \qquad \overline{\varnothing}=U \qquad \overline{\overline{A}}=A$$

Законы де Моргана

$$a)\ \overline{A\bigcap B}=\overline{A}\bigcup\overline{B}$$
$$b)\ \overline{A\bigcup B}=\overline{A}\bigcap\overline{B}$$

Законы поглощения

$$(A\bigcap B)\bigcup A=A \qquad (A\bigcup B)\bigcap A=A$$

Выражение для разности

$$A\setminus B=A\bigcap\overline{B}$$

В справедливости некоторых из перечисленных равенств мы убедились выше. Доказать другие можно аналогично, проведя формальные рассуждения. Для пояснения доказательств удобно нарисовать диаграммы Эйлера для левой и правой части равенств и убедиться, что они совпадают.

В качестве еще одного примера покажем, что $A\bigcup A=A$. Возьмем произвольный элемент x принадлежащий множеству из левой части равенства. Тогда $x\in A\bigcup A\Leftrightarrow(x\in A)\vee(x\in A)$. Но $p\vee p\equiv p$ для любого высказывания *p*. Поэтому $(x\in A)\vee(x\in A)\equiv(x\in A)$ и $x\in A$. Следовательно, $A\bigcup A\subset A$. Обратно, пусть есть высказывание *p*: $x\in A$. Тогда $p\equiv p\vee p$, т.е. $x\in A\vee x\in A$. Но тогда, по определению операции объединения, $x\in A\bigcup A$. Т.о. $A\subset A\bigcup A$. Поскольку $A\bigcup A\subset A$ и $A\subset A\bigcup A$, то $A=A\bigcup A$.

2.4 Количество элементов конечных множеств

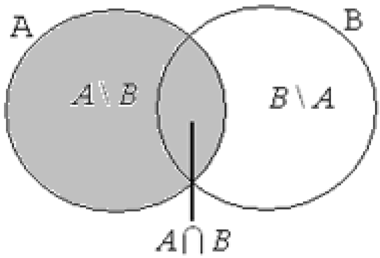
Мощностью конечного множества М называется количество его элементов и обозначается символом | М |. Например, $|\varnothing|=0$ (мощность пустого множества равна 0), $|\{a,b,c\}|=3$, $|\{\varnothing\}|=1$ (мощность множества, состоящего из одного пустого множества, равна 1). Если $|A| = |B|$, то множества A и B называются равномощными.

Правило суммы. Если ***A*** и ***B*** непересекающиеся конечные множества, содержащие m и n элементов соответственно, то объединение $A \cup B$ содержит m + n элементов. Т.е., если $A \cap B = \varnothing$, то $|A \cup B| = |A| + |B|$.

Теорема (*формула включений и исключений*) .

$$|A \cup B| = |A| + |B| - |A \cap B|$$

Доказательство. Из определений операций над множествами следует, что множество $A \cup B$ состоит из подмножеств A и $B \setminus A$, которые не имеют общих элементов, т.е. $A \cup B = A \cup (B \setminus A)$

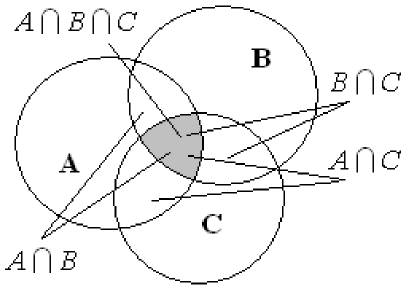


Кроме того $B = (B \setminus A) \cup (A \cap B)$ и при этом множества, стоящие в скобках, не пересекаются. Поэтому $|B| = |B \setminus A| + |A \cap B|$ и $|B \setminus A| = |B| - |A \cap B|$. Тогда

$$|A \cup B| = |A \cup (B \setminus A)| = |A| + |B \setminus A| = |A| + |B| - |A \cap B|$$

□

Формула для подсчета мощности объединения может быть обобщена на произвольное количество множеств. Для трех множеств она может быть получена из диаграммы Эйлера



и имеет вид

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |B \cap C| - |A \cap C| + |A \cap B \cap C|$$

Выполним формальную проверку формулы. Обозначим $A \cup B = D$. Тогда

$$|A \cup B \cup C| = |D \cup C| = |D| + |C| - |D \cap C| = |A \cup B| + |C| - |(A \cup B) \cap C| =$$
$$= (|A| + |B| - |A \cap B|) + |C| - |(A \cap C) \cup (B \cap C)|$$

Но

$$|(A \cap C) \cup (B \cap C)| = |A \cap C| + |B \cap C| - |(A \cap C) \cap (B \cap C)| =$$
$$= |A \cap C| + |B \cap C| - |A \cap B \cap C|$$

Тогда

$$|A \cup B \cup C| = |A| + |B| - |A \cap B| + |C| - (|A \cap C| + |B \cap C| - |A \cap B \cap C|) =$$

$$=|A|+|B|+|C|-|A\cap B|-|A\cap C|-|B\cap C|+|A\cap B\cap C|$$

□

Теорема. $|A\times B|=|A|\cdot|B|$ – мощность декартового произведения двух конечных множеств равна произведению их мощностей.

Доказательство. Первый компонент упорядоченной пары можно выбрать $|A|$ способами, второй – $|B|$ способами. На каждый способ выбора первого компонента второй можно выбрать $|B|$ способами. Таким образом, всего имеется $|A|\cdot|B|$ различных упорядоченных пар.

□

Следствие. $|A^n|=|A|^n$. Формула $|A\times A|=|A^2|=|A|^2$ для n=2 является следствием предыдущей теоремы. Учитывая, что $A^n=A\times A^{n-1}$ формула легко доказывается для любого n.

Если одно из множеств декартового произведения (или оба) бесконечны, то и произведение будет иметь бесконечное число упорядоченных пар. Если A пусто или B пусто, то, по определению, A x B пусто и его мощность равна нулю.

Множество всех подмножеств множества M, или булеан множества M, обозначаемый P(M), есть множество, состоящее из всех подмножеств множества M. Например, булеан множества M = {1,2,3} есть множество

$$\mathcal{P}(\mathrm{M})=\{\varnothing,\{1\},\{2\},\{3\},\{1,2\},\{2,3\},\{1,3\},\{1,2,3\}\}$$

Когда M содержит 3 элемента, P(M) состоит из $2^3 = 8$ элементов или, что то же самое, M включает $2^3 = 8$ подмножеств. В общем случае, если M содержит n элементов, множество P(M) включает 2^n элементов, т.к. M имеет 2^n подмножеств. Этот факт будет доказан ниже. По этой причине P(M) часто обозначают через 2^M. Поэтому $2^M:=\{A\,|\,A\subset M\}$

Теорема. Для конечного множества M: $|2^M|=2^{|M|}$.

Доказательство проведем по индукции.

Если $M=\varnothing$, то $2^M=\{\varnothing\}$, т.е. $2^{\varnothing}=\{\varnothing\}$ (множество всех подмножеств пустого множества состоит из одного пустого множества). Следовательно, $|2^{\varnothing}|=|\{\varnothing\}|=1=2^0=2^{|\varnothing|}$ и для пустого множества утверждение теоремы выполняется.

Предположим, что для множеств с количеством элементов $|M|<k$ утверждение теоремы уже проверены. Докажем его для множества с количеством элементов *k*, т.е. для $M=\{a_1,a_2,\ldots,a_k\}$ и $|M|=k$. Положим $M_1:=\{X\subset 2^M\,|\,a_k\in X\}$ – множество всех подмножеств множества M, которые содержат элемент a_k, и $M_2:=\{X\subset 2^M\,|\,a_k\notin X\}$ – множество всех подмножеств множества M, которые не содержат этот элемент. Имеем $2^M=M_1\cup M_2$ и $M_1\cap M_2=\varnothing$. Поэтому $|2^M|=|M_1|+|M_2|$.

Все подмножества, входящие в M_1, содержат фиксированный элемент a_k и его присутствие в подмножестве не влияет на их количество. Поэтому $|M_1| = 2^{k-1}$. Но также $|M_2| = 2^{k-1}$. Следовательно,

$$|2^M| = |M_1| + |M_2| = 2^{k-1} + 2^{k-1} = 2 \cdot 2^{k-1} = 2^k = 2^{|M|}$$

3. Алгебра отношений.

3.1 Понятие отношения

Рассмотрим множество целых чисел **Z**. Между элементами этого множества существуют различные отношения. Так, если $a, b \in \mathbf{Z}$, то a либо делит b, либо не делит. В первом случае говорят, что a находится в отношении делимости к b. Если числа a и b взаимно простые, то говорят, что a и b находятся в отношении взаимной простоты. Еще один пример отношения на множестве **Z** дает отношение порядка: число a находится в отношении порядка к числу b, если $a \le b$. При этом в первом и третьем примерах порядок элементов a и b играет существенную роль. Действительно, из того, что a делит b, вовсе не следует, что b делит a. Другими словами, нельзя путать делитель и делимое.

В качестве еще одного примера рассмотрим множество S студентов факультета и множество K читаемых им курсов. В прямом произведении S x K можно выделить большое подмножество упорядоченных пар (s,k), обладающих свойством: студент s слушает курс k «дискретная математика». Построенное подмножество отражает отношение «...слушает...», возникающее между множествами студентов и читаемых курсов.

Пусть A и B — произвольные непустые множества. Задание отношения между элементами множеств $a \in A$ и $b \in B$ равносильно описанию множества пар (a, b) элементов $a \in A$ и $b \in B$, связанных этим отношением. При этом важен порядок элементов пары. Напомним, что множество всех упорядоченных пар элементов (a, b) таких, что $a \in A$ и $b \in B$ мы называли декартовым произведением $A \times B$ множеств A и B. Всякому отношению R между элементами множеств A и B соответствует некоторое вполне определенное подмножество $R \subset A \times B$, состоящее из пар (a, b), таких, что a находится в отношении R к b. Это подмножество содержит всю информацию об отношении R. Если множество B совпадает с множеством A, то говорят об отношении, заданном на множестве A.

Определение. Пусть A и B — произвольные непустые множества $(A \neq \varnothing, B \neq \varnothing)$. Всякое подмножество $R \subset A \times B$ ($R \subset A^2$) называется бинарным отношением, заданным на множестве $A \times B$ (на множестве A^2). Говорят, что элемент a из A находится в отношении R к элементу b из B (из A), и пишут a R b, если пара $(a,b) \in R$.

Указанные выше три отношения на множестве **Z** можно теперь определить так:

$R_1 = \{(a,b) \in \mathbf{Z}^2 : a \text{ делит } b\}$,

$R_2 = \{(a,b) \in \mathbf{Z}^2 : a \text{ и } b \text{ взаимно просты}\}$,

$R_3 = \{(a,b) \in \mathbf{Z}^2 : a \le b\}$.

Если A = {1,2,3}, а B={r,s}, то A x B = {(1,r),(1,s),(2,r),(2,s),(3,r),(3,s)}. Тогда R={(1,r),(1,s),(3,s)} есть отношение на множествах A и B, и $R \subset A \times B$. При этом можно, например, писать 3 R s, поскольку $(3,s) \in R$. Множество A x B содержит шесть элементов, поэтому существует 2^6 = 64 подмножества множества A x B. Это значит, что существует 64 различных отношения на A x B (включая пустое).

Вот еще несколько примеров отношений:

1. Если A — множество действительных чисел, то $\{(x,y) \in A \times A : x^2 + y^2 = 4\}$ есть бинарное отношение на A.
2. Пусть A — множество товаров в магазине, а B — множество действительных чисел. Тогда $\{(x,y) \in A \times B : y - \textit{цена}\ x\}$ — отношение на множестве A x B.
3. Пусть A - множество женщин, а B — множество мужчин, тогда {(x,y) : y является мужем x} есть отношение на множестве A x B.
4. Если A - множество людей, то $\{(x,y) \in A \times A : y \text{ является родственником } x\}$ есть бинарное отношение на A^2.

Область определения D_R отношения R на $A \times B$ есть множество всех $x \in A$ таких, что для некоторых $y \in B$ имеем $(x,y) \in R$. Другими словами, область определения $D_R \subset A$ отношения R есть множество всех первых координат упорядоченных пар из R . **Множество значений** V_R отношения R на A и B есть множество всех $y \in B$ таких, что $(x,y) \in R$ для некоторого $x \in A$. Другими словами, множество значений $V_R \subset B$ отношения R есть множество всех вторых координат упорядоченных пар из R.

В примерах отношений, приведенных выше, в частности, в 1) область определения и множество значений совпадают с множеством $\{ x : -2 \le x \le 2 \}$. В 2) область определения есть множество товаров A, а множество значений есть множество всех действительных чисел, каждое из которых совпадает с ценой некоторого товара в магазине. В 3) область определения есть множество всех замужних женщин, а множество значений — множество всех женатых мужчин. В 4) область определения и множество значений есть множество всех людей, имеющих родственников.

Замечание. Вместо записи $(x,y) \in R$, где R — бинарное отношение, часто употребляют обозначение x R y. Например, высказывание «x – сестра y» определяет отношение на множестве всех людей. Аналогично высказывание «x – делитель y» дает ясное словесное описание заданного отношения.

Бинарное **отношение тождества**, заданное на множестве A, состоит из всех пар вида (*a*,*a*), где $a \in A$, и обозначается I (или I_A, если надо явно указать, на каком множестве задано отношение).

С каждым отношением R на A x B связано отношение R^{-1} на B x A. Пусть $R \subseteq A \times B$ есть отношение на A x B. Тогда отношение R^{-1} на B x A определяется следующим образом:

$$R^{-1} = \{(b,a):(a,b) \in R\}$$

Другими словами, $(b,a) \in R^{-1}$ тогда и только тогда, когда $(a,b) \in R$. Отношение R^{-1} называется **обратным отношением** к данному отношению.

Пусть R={(1,r),(1,s),(3,s)}, тогда $R^{-1} = \{(r,1),(s,1),(s,3)\}$. Пусть R – отношение {(x,y):x является мужем y}, тогда обратное отношение R^{-1}={(y,x):y жена x}. Если R – отношение $\{(x,y): x^2 + y^2 = 4\}$, тогда $R^{-1} = R$.

Легко видеть, что $(R^{-1})^{-1} = R$. Действительно, если $(a,b) \in R$, то $(b,a) \in R^{-1}$. Но тогда $(a,b) \in (R^{-1})^{-1}$, т.е. $R \subset (R^{-1})^{-1}$. Обратно, если $(a,b) \in (R^{-1})^{-1}$, то это значит, что $(b,a) \in R^{-1}$ и, следовательно, $(a,b) \in R$. Т.е. $(R^{-1})^{-1} \subset R$. Следовательно $(R^{-1})^{-1} = R$.

Дополнение $\overline{R}$ отношения $R \subset A \times B$ определяется как $\overline{R} = \{(x,y):(x,y) \notin R\}$ или $\overline{R} = \{(x,y):(x,y) \in A \times B \setminus R\}$.

Аналогично бинарным отношениям можно определить отношение на декартовом произведении n множеств $X_1 \times X_2 \times \ldots \times X_n$ или на n – ой степени одного множества $X^n = \underbrace{X \times X \times \ldots \times X}_{n\ раз}$. Подмножество $R \subset X_1 \times X_2 \times \ldots \times X_n$ (или $R \subset X^n$) называется n-местным или n - арным отношением. Говорят, что элементы $x_1, x_2, \ldots, x_n$ (в указанном порядке) связаны отношением R или находятся в отношении R, если $(x_1, x_2, \ldots, x_n) \in R$. При n=2 мы говорим – бинарное отношение (би – означает два, т.е. между двумя элементами). При n=1 отношение называется унарным. Унарное отношение на множестве X – это просто некоторое подмножество R множества X. Унарное отношение можно трактовать как свойство: элемент $x \in X$ обладает свойством *R*, если $x \in R$.

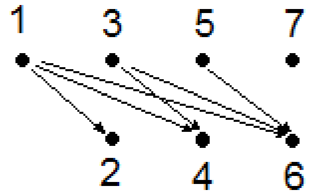
Познакомимся со способами задания отношений.

Пусть A и B — два конечных множества и R — бинарное отношение между ними. Изобразим элементы этих множеств точками на плоскости. Для каждой упорядоченной пары отношения R нарисуем стрелку, соединяющую точки, представляющие компоненты пары. Такой объект называется ориентированным графом или орграфом. Точки, изображающие элементы множеств, принято называть вершинами графа.

Пример 1. Рассмотрим отношение $V = \{(x, y : x < y)\}$ между множествами `A={1,3,5,7}` и `B={2,4,6}`. Легко видеть, что оно состоит из элементов

$$V = \{(1,2), (1,4), (1,6), (3,4), (3,6), (5,6)\}$$

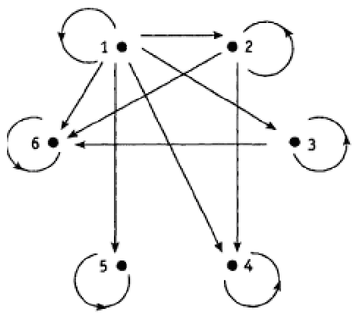
Соответствующий ориентированный граф показан на следующем рисунке



∎

Для иллюстрации отношения на отдельном множестве `A` (т.е. на `A x A`) мы чертим орграф, чьи вершины соответствуют одному лишь множеству `A`, а стрелки соединяют элементы этого множества.

Пример 2. Пусть на множестве `A={1,2,3,4,5,6}` задано отношение $R = \{(x, y) : x - \text{делитель } y\}$. Перечислим все упорядоченные пары ему принадлежащие: `(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,2),(2,4),(2,6),(3,3),(3,6),(4,4),(5,5),(6,6).` Ориентированный граф имеет 6 вершин и имеет следующий вид



∎

Другой способ задания бинарного отношения на конечных множествах основан на использовании матриц. Предположим, что мы хотим определить бинарное отношение `R` между множествами `A` и `B`. Необходимо обозначить элементы множеств и выписать их в каком-нибудь порядке. Сделаем это так:

$$A = \{a_1, a_2, \ldots, a_n\}, \qquad B = \{b_1, b_2, \ldots, b_m\}$$

Для определения отношения `R` заполним таблицу `M` с `n` строками и `m` столбцами. Строки «перенумеруем» элементами множества `A`, а столбцы — элементами множества `B`, в соответствии с порядком, в котором мы выписали элементы. Ячейку таблицы, стоящую на пересечении *i*-той строки и *j*-того столбца будем обозначать через `M(i,j)`, а заполнять ее будем следующим образом:

$$M(i, j) = T\ ,\ (a_i, b_j) \in R$$

$$M(i, j) = F\ ,\ (a_i, b_j) \notin R$$

Здесь `T=True=ИСТИНА` и `F=False=ЛОЖЬ`. Подобные таблицы называются $n \times m$ матрицами (логическими). В этих терминах, отношение `V` из примера 1 с помощью матрицы задается следующим образом:

$$\begin{array}{c c} & \begin{array}{ccc} 2 & 4 & 6 \end{array} \\ \begin{array}{c} 1 \\ 3 \\ 5 \\ 7 \end{array} & \begin{pmatrix} T & T & T \\ F & T & T \\ F & F & T \\ F & F & F \end{pmatrix} \end{array}$$

Чтобы лучше понять такой способ задания отношений, мы явно пометили столбцы и строки матрицы элементами множеств A и B. В общем случае это делать не обязательно.

Наоборот, имея логическую матрицу, можно «восстановить» отношение. Пусть, например, на множестве `A={a,b,c,d}` задана матрица

$$\begin{pmatrix} F & T & T & F \\ F & F & T & T \\ F & T & F & F \\ T & T & F & T \end{pmatrix}$$

порядок строк и столбцов в которой соответствует порядку выписанных элементов множества `A`. Тогда отношение `R` содержит упорядоченные пары:

`(a,b),(a,c),(b,c),(b,d),(c, b),(d,a),(d,b, {d,d).`

Отношение еще можно изобразить как подмножество в таблице, представляющей декартовое произведение множеств A и B. Для этого в таблице декартового произведения $A \times B$, надо выделить (например, цветом) те клетки, которые содержат элементы (пары), принадлежащие отношению. Для примера 1 такое графическое представление отношения V может иметь вид

	2	4	6
1	(1,2)	(1,4)	(1,6)
3	(3,1)	(3,4)	(3,6)
5	(5,2)	(5,4)	(5,6)
7	(7,2)	(7,4)	(7,6)

Пусть множество S состоит из трех студентов: Демьяненко, Крохмаль, Ткаченко. Они сдают в сессию 4 предмета: дискретную математику (ДМ), высшую математику (ВМ), программирование (ПР) и физику (ФИЗ). Предметы образуют множество T. Множество V допустимых положительных оценок состоит из A, B, C, D, E. Любое подмножество декартового произведение $S \times T \times V$ является отношением (трехместным) и состоит из элементов (фамилия, предмет, оценка). Пусть, например, отношение состоит из следующих элементов: (Демьяненко, ВМ, A), (Демьяненко, ДМ, B), (Крохмаль, ПР, C), (Ткаченко, ДМ, E), (Крохмаль, ФИЗ, A). Это отношение удобно представлять в виде следующей таблицы

Демьяненко	ВМ	A
Демьяненко	ДМ	B
Крохмаль	ПР	C
Ткаченко	ДМ	E
Крохмаль	ФИЗ	A

Таким образом, отношение как подмножество декартового произведения множеств $A_1 \times A_2 \times \ldots \times A_n$ может быть представлено в виде таблицы. Наоборот, если есть таблица, удовлетворяющая некоторым естественным условиям, то она может рассматриваться как представление некоторого отношения. Чтобы таблица представляла отношение, она должна представлять некоторое подмножество в декартовом произведении. Значит в ней не должно быть повторяющихся строк (каждый элемент множества уникальный) и ее строки не должны быть упорядочены (элементы множества неупорядочены).

Таким образом, бинарное отношение между конечными множествами может быть задано одним из следующих способов:

- словами (с помощью подходящих предикатов);
- как множество упорядоченных пар;
- как орграф;
- как матрица;

Любое n – арное отношение (в том числе бинарное) можно представлять в виде таблицы.

Операции над отношениями. Все отношения, заданные на $A_1, A_2, \ldots, A_n$, являются подмножествами множества $A_1 \times A_2 \times \ldots \times A_n$, которое для них можно считать универсумом. Поэтому для отношений определены операции объединения $R_1 \bigcup R_2$, пересечения $R_1 \bigcap R_2$, разности $R_1 \setminus R_2$ и дополнения $\overline{R} = A_1 \times A_2 \times \ldots \times A_n \setminus R$.

Имея два заданных отношения, можно строить новые отношения путем их комбинации. Пусть $R \subseteq A \times B$ - отношение на $A \times B$ и $S \subseteq B \times C$ - отношение на $B \times C$. **Композицией** (произведением, суперпозицией) отношений S и R, где $R \subseteq A \times B, S \subseteq B \times C,$ называется отношение $T \subseteq A \times C$, которое определяется следующим образом

$$T = S \circ R = \left\{ (a,c) : a \in A \wedge c \in C \;\; и \;\; существует \;\; b \in B : (a,b) \in R \wedge (b,c) \in S \right\}$$

Пример. Пусть A = {1,2,3}, B = {x, y}, а C={s,t,c,z} и пусть отношения R на A x B и S на B x C заданы в виде:

R={(1,x),(1,y),(3,x)

S={(x,s),(x,t),(y,c),(y,z)}

Тогда

$$S \circ R = \{(1,s),(1,t),(1,c),(1,z),(3,s),(3,t)\}$$

∎

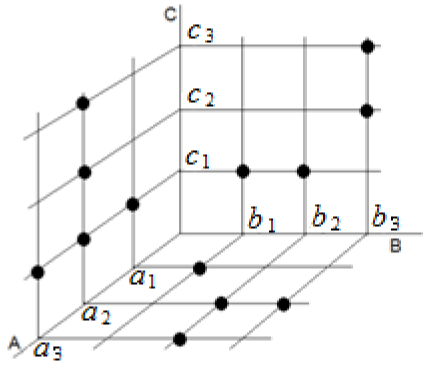
Пример. Пусть $A = \{ a_1, a_2, a_3 \}$, $B = \{ b_1, b_2, b_3 \}$, а $C = \{ c_1, c_2, c_3 \}$ и пусть отношения R на A x B и S на B x C заданы в виде:

$$R = \{ (a_1, b_1), (a_2, b_2), (a_2, b_3), (a_3, b_2) \}$$

$$S = \{ (b_1, c_1), (b_2, c_1), (b_3, c_2), (b_3, c_3) \}$$

Тогда $S \circ R = \{(a_1, c_1), (a_2, c_1), (a_2, c_2), (a_2, c_3), (a_3, c_1)\}$.

Изобразим элементы множеств A, B, C точками на осях в 3-х мерной декартовой системе координат. Отношения R и S можно представить точками на плоскостях AB и BC соответственно (см. следующий рисунок).



Тогда композиция отношений $S \circ R$ будет представляться точками на плоскости АС.

∎

Пример. Пусть R и S — бинарные отношения на множестве положительных целых чисел, заданные в виде S={(x,x+2): x — положительное целое число} и R={(x,x^2): x — целое положительное число}. Тогда

S ∘ R={(x,x^2+2): x — положительное целое число}

и

R ∘ S={(x,(x + 2)^2): x — положительное целое число}.

∎

Теорема. Композиция отношений ассоциативна. Т.е., если A, B и C - множества и $R \subseteq A \times B, S \subseteq B \times C, \;\; T \subseteq C \times D$, то $T \circ (S \circ R) = (T \circ S) \circ R$.

Доказательство. Покажем сначала, что $T \circ (S \circ R) \subseteq (T \circ S) \circ R$. Пусть $(a,d) \in T \circ (S \circ R)$, тогда существует такое $c \in C$, что $(a,c) \in S \circ R$ и $(c,d) \in T$. Поскольку $(a,c) \in S \circ R$, существует такое b, что $(a,b) \in R$ и $(b,c) \in S$. Поскольку $(b,c) \in S$ и $(c,d) \in T$ имеем $(b,d) \in T \circ S$. Поскольку $(b,d) \in T \circ S$ и $(a,b) \in R$ имеем $(a,d) \in (T \circ S) \circ R$. Таким образом, $T \circ (S \circ R) \subseteq (T \circ S) \circ R$. Обратная часть доказательства, показывающая, что $(T \circ S) \circ R \subseteq T \circ (S \circ R)$, выполняется аналогично.

∎

Композицию бинарных отношений можно вычислить с помощью матриц, их определяющих. Имея матрицы двух отношений, мы построим матрицу их композиции. Она называется логическим или булевым произведением матриц. Рассмотрим три множества:

$$A = \{a_1, a_2, \ldots, a_n\}, \qquad B = \{b_1, b_2, \ldots, b_m\}, \qquad C = \{c_1, c_2, \ldots, c_p\}.$$

Предположим, что R — отношение между A и B , а S — отношение между B и C. Напомним, что матрица M отношения R определяется условием:

$$M(i,j) = T \,, \;\; (a_i, b_j) \in R$$

$$M(i,j) = F \,, \;\; (a_i, b_j) \notin R$$

Аналогично заполняется матрица N отношения S

$$N(i,j) = T \,, \;\; (b_i, c_j) \in S$$

$$N(i,j) = F \,, \;\; (b_i, c_j) \notin S$$

Композиция отношений S и R, где $R \subseteq A \times B, S \subseteq B \times C$, представляется множеством пар (a, c): $a \in A \wedge c \in C$ таких, что существует

$b \in B : (a,b) \in R \wedge (b,c) \in S$. Возьмем a_i – i - й элемент множества A и c_j – j – й элемент множества C. Пара (a_i,c_j) принадлежит композиции отношений (т.е. в матрице P, представляющей суперпозицию отношений, в i – ой строке, j – м столбце стоит T), если найдется элемент $b_k \in B$, такой, что пара $(a_i, b_k) \in R$ и $(b_k, c_j) \in S$. Это значит, что $M(i,k) = T$ и $N(k, j) = T$. Т.е. мы должны просмотреть всю i – ю строку матрицы M и весь j – й столбец матрицы N и найти в них на одинаковом по порядку месте пару значений T. Если такая пара есть, то $(a_i, c_j) \in S \circ R$ и $P(i, j) = T$. Если же в i – й строке матрицы M нет значений T, соответствующих T на таком же месте в j – м столбце матрицы N, то P(i,j)=F.

Таким образом, логическая матрица P композиции S ∘ R заполняется по следующему правилу:

$$P(i, j) = (M(i,1) \wedge N(1, j)) \vee (M(i,2) \wedge N(2, j)) \vee \ldots \vee (M(i,n) \wedge N(n, j))$$

или

$$P(i, j) = \bigvee_{k=1}^{n} (M(i,k) \wedge N(k, j)).$$

Для матрицы P пишут $P = M \times N = M \cdot N$ и она называется булевым произведением матриц.

Пример. Пусть A ={a,b}, B ={1,2,3} и C={x,y}. На этих множествах заданы отношения $R \subset A \times B$ и $S \subset B \times C$.

R ={(a,1),(a,2),(a,3),{(b,2)} и S={(1,y),(2,x),(3,x)}

Найдем отношение $S \circ R$. Отношение R между A ={a,b} и B ={1,2,3} задается матрицей

$$M = \begin{bmatrix} T & T & T \\ F & T & F \end{bmatrix}$$

Аналогично, S — отношение между B и C, задается матрицей

$$N = \begin{bmatrix} F & T \\ T & F \\ T & F \end{bmatrix}$$

Логическая матрица P отношения S ∘ R будет содержать две строки (два элемента множества A) и два столбца (два элемента множества C).

$$P = \begin{pmatrix} T & T & T \\ F & T & F \end{pmatrix} \begin{pmatrix} F & T \\ T & F \\ T & F \end{pmatrix} = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}$$

Для определения элемента P_{11} просматриваем первую строку и первый столбец матриц, определяя в них на одинаковом месте символ T. У нас есть, например, элемент $2 \in B$ для которого $(a,2) \in R$ и $(2, x) \in S$. Следовательно, $(a, x) \in S \circ R$ и $P_{11} = T$. По – другому

$$P_{11} = (T \wedge F) \vee (T \wedge T) \vee (T \wedge T) = F \vee T \vee T = T$$

Аналогично

$$P_{12} = (T \wedge T) \vee (T \wedge F) \vee (T \wedge F) = T \vee F \vee F = T$$
$$P_{21} = (F \wedge F) \vee (T \wedge T) \vee (F \wedge T) = F \vee T \vee F = T$$
$$P_{22} = (F \wedge T) \vee (T \wedge F) \vee (F \wedge F) = F \vee F \vee F = F$$

Таким образом,

$$P = \begin{pmatrix} T & T \\ T & F \end{pmatrix}.$$

Восстанавливая отношение по матрице, получаем $S \circ R = \{(a,x),(a,y),(b,x)\}$.

∎

Теорема. Если R_1, R_2, R_3 – бинарные отношения, заданные на множестве A (т.е. $R_1 \subseteq A^2$, $R_2 \subseteq A^2$, $R_3 \subseteq A^2$,), тогда:

1. $(R_1 \bigcup R_2) \circ R_3 = R_1 \circ R_3 \bigcup R_2 \circ R_3$;
2. $R_1 \subseteq R_2 \Rightarrow R_1 \circ R_3 \subseteq R_2 \circ R_3$;
3. $(R_1 \circ R_2)^{-1} = R_2^{-1} \circ R_1^{-1}$;
4. $(R_1 \bigcap R_2)^{-1} = R_1^{-1} \bigcap R_1^{-1}$.

Доказательство.

1. Пусть $(a,b) \in (R_1 \bigcup R_2) \circ R_3$. Тогда существует элемент $c \in A$ такой, что $(a,c) \in R_1 \bigcup R_2$ и $(c,b) \in R_3$. Следовательно $(a,c) \in R_1$ или $(a,c) \in R_2$. Тогда имеем $(a,b) \in R_1 \circ R_3$ или $(a,b) \in R_2 \circ R_3$, т.е. $(a,b) \in R_1 \circ R_3 \bigcup R_2 \circ R_3$. Обратное включение доказывается прочтением предыдущих выкладок в обратную сторону.

2. Пусть $R_1 \subseteq R_2$. Тогда $R_1 \bigcup R_2 = R_2$. Учитывая утверждение 1, имеем

$$R_2 \circ R_3 = (R_1 \bigcup R_2) \circ R_3 = (R_1 \circ R_3) \bigcup (R_2 \circ R_3).$$

Отсюда, $R_1 \circ R_3 \subseteq R_2 \circ R_3$.

3. Пусть $(a,b) \in (R_1 \circ R_2)^{-1}$. Тогда $(b,a) \in R_1 \circ R_2$ и существует $c \in A$ такое, что $(b,c) \in R_1$ и $(c,a) \in R_2$. Но тогда $(c,b) \in R_1^{-1}$ и $(a,c) \in R_2^{-1}$. Следовательно, $(a,b) \in R_2^{-1} \circ R_1^{-1}$. Т.е. $(R_1 \circ R_2)^{-1} \subseteq R_2^{-1} \circ R_1^{-1}$. Обратное включение доказывается аналогично.

4. Пусть $(a,b) \in (R_1 \bigcap R_2)^{-1}$. Тогда $(b,a) \in R_1 \bigcap R_2$. Поэтому $(b,a) \in R_1$ и $(b,a) \in R_2$. Но тогда $(a,b) \in R_1^{-1}$ и $(a,b) \in R_2^{-1}$, т.е. $(a,b) \in R_1^{-1} \bigcap R_2^{-1}$ и, следовательно, $(R_1 \bigcap R_2)^{-1} \subseteq R_1^{-1} \bigcap R_2^{-1}$. Включение в обратную сторону доказывается аналогично. Т.о. $(R_1 \bigcap R_2)^{-1} = R_1^{-1} \bigcap R_2^{-1}$.

∎

Замечание. Кроме уже описанных операций над отношениями можно рассматривать и другие операции. Все множество операций и их свойства образуют алгебру отношений или реляционную алгебру. Термин «реляционный» произошел от английского слова `realation` (отношение). Введению в реляционную алгебру посвящен последний параграф этой главы.

3.2 Свойства отношений

Рассмотрим бинарные отношения, заданные на одном множестве, т.е. $R \subset A^2$, и определим некоторые их свойства. Говорят, что отношение R на множестве A:

рефлексивно, если для всех $x \in A$: x R x ;

антирефлексивно, если из x R y следует $x \neq y$;

симметрично, если для любых $x, y \in A$ из x R y следует y R x; в частности это означает, что $R \subseteq R^{-1}$; по-другому, отношение симметрично, если вместе с каждой парой $(x, y) \in R$ отношение содержит пару $(y, x) \in R$.

антисимметрично (кососимметрично**)**, если для любых $x, y \in A$ из x R y и y R x следует x=y , т.е. $R \cap R^{-1} \subseteq I$;

транзитивно, если для любой тройки элементов x, y, z из A из x R y и y R z следует x R z;

полно, если для любых $x, y \in A$ из условия $x \neq y$ следует $x R y \vee y R x$.

Пример 1. Пусть A={1,2,3,4,5,6} и пусть отношение $R_1 \subseteq A \times A$ есть множество R_1={(1,1),(2,2),(3,3),(4,4),(5,5),(6,6),(1,2),(1,4),(2,1),(2,4),(3,5),(5,3),(4,1),(4,2)}. Матрица отношения имеет вид

$$\begin{pmatrix} T & T & F & T & F & F \\ T & T & F & T & F & F \\ F & F & T & F & T & F \\ T & T & F & T & F & F \\ F & F & T & F & T & F \\ F & F & F & F & F & T \end{pmatrix}$$

Отношение R_1 рефлексивно, т.к. для каждого $a \in A$ $(a, a) \in R_1$. Кроме того, видно, что вместе с каждой парой (a,b) в R_1 содержится пара (b,a). А это означает, что отношение R_1 симметрично. Используя метод прямого перебора, представленного в следующей таблице, можно убедиться, что отношение R_1 транзитивно.

$(a,b) \in R_1$	$(b,c) \in R_1$	(a,c)	$(a,c) \in R_1$?
(1,2)	(2,1)	(1,1)	T
(1,2)	(2,2)	(1,2)	T
(1,2)	(2,4)	(1,4)	T
(1,4)	(4,1)	(1,1)	T
(1,4)	(4,2)	(1,2)	T
(2,1)	(1,1)	(2,1)	T
...	...	...	...

Проанализировав каждый возможный случай, когда $(a,b) \in R_1$ и $(b,c) \in R_1$ получаем, что $(a,c) \in R_1$. Отношение R_1 не является антисимметричным, поскольку $(1,2) \in R_1$ и $(2,1) \in R_1$, но $1 \neq 2$.

Примеры.

а) Отношение $x \leq y$ рефлексивно. Рефлексивное отношение на множестве действительных чисел изображается на координатной плоскости множеством точек, содержащим прямую y=x. Оно также транзитивно, поскольку из $x \leq y$ и $y \leq z$ следует, что $x \leq z$. Это отношение также антисимметрично, поскольку из $x \leq y$ и $y \leq x$ следует $x = y$.

б) Отношение $x + y > 0$ симметрично, но не транзитивно. Действительно, $1+5>0 \wedge 5+(-4)>0$, но $1+(-4)<0$.

в) Отношение $x + 2y > 0$ не симметрично (симметричное отношение на множестве действительных чисел изображается на координатной плоскости множеством точек, симметричным относительно прямой y=x). В нашем примере из того, что $x + 2y > 0$ не следует, что $y + 2x > 0$.

г) Отношение нестрогого включения $X \subseteq Y$ рефлексивно (т.к. $X \subseteq X$), антисимметрично (из $X \subseteq Y$ и $Y \subseteq X$ следует $X = Y$) и транзитивно (из $X \subseteq Y$ и $Y \subseteq Z$ следует $X \subseteq Z$). Отношение строго включения $X \subset Y$, X≠Y транзитивно.

д) Отношение «прямая x параллельна прямой y» рефлексивно и симметрично. Действительно, из элементарной геометрии известно, что две прямые, которые лежат в одной плоскости, параллельны, когда они совпадают, либо не имеют ни одной общей точки. Поскольку прямая x совпадает сама с собой, то пара (x,x) принадлежит данному отношению. Если «прямая x параллельна прямой y», то «прямая y параллельна прямой x». Следовательно, отношение симметрично.

е) Отношение «студент x — ровесник студента y» рефлексивно, поскольку каждый студент сам себе ровесник. Оно также симметрично, т.к., если студент x ровесник y, то y ровесник x.

ж) Отношение «прямая x перпендикулярна прямой y в плоскости XY» симметрично, поскольку, если прямая x перпендикулярна прямой y, то и прямая y перпендикулярна прямой x. Это отношение не рефлексивно.

з) Отношение «студент x является соседом по парте студента y» симметрично. Действительно, каждый студент может убедиться, что когда студент y приходится ему соседом, то студенту y приходится соседом он сам. Это отношение не рефлексивно (студен не является соседом самому себе).

и) Отношение «город x связан с городом y шоссейной дорогой» транзитивно, поскольку, если между городами x и y есть шоссейная дорога и между городами y,z также есть шоссейная дорога, то ясно, что между городами x, z тоже есть шоссейная дорога, проходящая через город y.

к) Отношения «треугольник x подобен треугольнику y», «действительное число x больше действительного числа y» очевидно транзитивны.

Пример. Что можно сказать о свойствах (рефлексивности, симметричности, кососимметричности и транзитивности) следующих отношений:

(а) «x делит y» на множестве натуральных чисел;

(б) $x \neq y$ на множестве целых чисел.

Решение.

а) Поскольку x всегда делит сам себя, то это отношение рефлексивно. Оно не симметрично, поскольку, например, 2 является делителем 6, но не наоборот: 6 не делит 2. Проверим, что отношение делимости транзитивно. Предположим, что x делит y, а y в свою очередь делит z. Тогда из первого предположения вытекает, что $y = m \cdot x$ для некоторого натурального числа m, а из второго — $z = n \cdot y$ где n — натуральное число. Следовательно, $z = n \cdot y = (n \cdot m)x$, т.е. x делит z. Значит, данное отношение транзитивно. Наконец, наше отношение кососимметрично, поскольку из предположений: x делит y и y делит x немедленно вытекает, что y = x.

б) Так как высказывание $x \neq x$ ложно, то это отношение не рефлексивно. Оно симметрично, поскольку $x \neq y$ тогда и только тогда, когда $y \neq x$. Это отношение не обладает свойством транзитивности, так как, например, $2 \neq 3$ и $3 \neq 2$, но, тем не менее, 2 = 2. Наше отношение не кососимметрично, поскольку из условий $x \neq y$ и $y \neq x$ нельзя заключить, что x = y.

■

Перечислим свойства ориентированных графов, задающих отношения на множестве A с тем или иным свойством. У ориентированного графа, изображающего рефлексивное отношение, каждая вершина снабжена петлей, т.е. стрелкой, начинающейся и заканчивающейся в одной и той же вершине. Орграф симметричного отношения вместе с каждой стрелкой из вершины x в вершину y имеет стрелку, направленную в обратную сторону: из y в *x*. Если отношение кососимметрично, то при наличии стрелки из вершины x в несовпадающую с ней вершину y, стрелка из y в x будет обязательно отсутствовать. Орграф транзитивного отношения устроен так, что вместе со стрелками из вершины x в y и из y в z у него будет стрелка и из x в z.

Перечислим свойства матриц, задающих отношения. Прежде всего заметим, что матрица отношения на отдельном множестве A будет квадратной, т. е. количество ее строк будет равно количеству столбцов. Матрица M, задающая рефлексивное отношение, отличается от других тем, что каждый ее элемент, стоящий на главной диагонали M(i,i), равен T; матрица M симметричного отношения будет симметричной, т.е. M(i,j)=M(j,i); в матрице кососимметричного отношения выполнено условие:

$$\left(M(i,j) = T \wedge i \neq j\right) \Rightarrow M(j,i) = F$$

Для сокращения записей при доказательстве следующей теоремы будем использовать два специальных знака (квантора): $\forall$ - читается «для любого» и $\exists$ - читается «существует». Напомним также, что квантор $\Leftrightarrow$ читается как «тогда и только тогда».

Теорема. Пусть $R \subset A^2$ - отношение на A. Тогда:

1. R рефлексивно $\Leftrightarrow I \subset R$;
2. R симметрично $\Leftrightarrow R = R^{-1}$;

3. R транзитивно $\Leftrightarrow R \circ R \subset R$;

Доказательство.

1. Пусть отношение R рефлексивно. Тогда $\forall a \in A$, $(a,a) \in R \Rightarrow I \subset R$. Обратно, пусть $I \subset R$. Тогда $\forall a \in A$, $(a,a) \in I \subset R \Rightarrow \forall a \in A, a\, R\, a$ и отношение R рефлексивно.

2. Пусть R симметрично. Тогда $\forall a,b \in A$ таких, что $(a,b) \in R \Rightarrow (b,a) \in R$. Но $(b,a) \in R^{-1}$, т.е. $R \subset R^{-1}$. Аналогично, если $(a,b) \in R$, то $(b,a) \in R^{-1}$ и, в силу симметричности R, $(b,a) \in R$, т.е. $R^{-1} \subset R$. Тогда $R = R^{-1}$.

Пусть $R = R^{-1}$. Тогда $R \subset R^{-1}$ и, если $(a,b) \in R$, то $(a,b) \in R^{-1}$. Поэтому $(b,a) \in R$. Поскольку $R = R^{-1}$, то это значит, что вместе с любой парой $(a,b) \in R$ пара $(b,a) \in R$, т.е. отношение R симметрично.

3. Пусть $(a,c) \in R \circ R$. Тогда существует $b \in A$ такое, что $(a,b) \in R \wedge (b,c) \in R$. Если R транзитивно, то из $(a,b) \in R \wedge (b,c) \in R$ следует $(a,c) \in R$. Это значит, что $R \circ R \subset R$.

Пусть теперь $R \circ R \subset R$. Это значит, что для любой пары $(a,b) \in R \circ R$ имеем $(a,b) \in R$. Тем самым, из условия $a\, R\, c \wedge c\, R\, b$, следует $(a,b) \in R$, т.е. имеет место транзитивность.

∎

Если отношение R на множестве A не обладает тем или иным свойством, то его (отношение) можно попытаться продолжить до отношения R*, которое будет иметь нужное свойство. Под «продолжением» мы понимаем присоединение некоторых упорядоченных пар к подмножеству $R \subset A^2$ так, что новое полученное множество R* уже будет обладать требуемым свойством. Ясно, что исходное множество R будет подмножеством R*. В том случае, если вновь построенное множество R* будет минимальным среди всех расширений R с выделенным свойством, то говорят, что R* является замыканием R относительно данного свойства.

Более строго, R* называется **замыканием** отношения R относительно свойства P, если

1. R* обладает свойством P;
2. $R \subset R^*$;
3. R* является подмножеством любого другого отношения, содержащего R, и обладающего свойством P.

Пример. Пусть A={1,2,3}, а отношение R на A задано упорядоченными парами R={(1,1),(1,2),(1,3),(3,1),(2,3)}. Оно не рефлексивно, не симметрично и не транзитивно. Построим соответствующие замыкания.

Решение. Замыкание относительно рефлексивности должно содержать все пары вида (*x, x*). Поэтому, искомое замыкание имеет вид:

R_1^*={(1,1),(1,2),(1,3),(3,1),(2,3);(2,2),(3,3)},

где добавленные пары отделены от исходных точкой с запятой.

	1	2	3
1	(1,1)	(1,2)	(1,3)
2	(2,1)	(2,2)	(2,3)
3	(3,1)	(3,2)	(3,3)

отношение R

	1	2	3
1	(1,1)	(1,2)	(1,3)
2	(2,1)	(2,2)	(2,3)
3	(3,1)	(3,2)	(3,3)

отношение R_1^*

Замыкание относительно симметричности должно содержать все пары, симметричные исходным. Значит,

R_2^*= {(1,1),(1,2),(1,3),(3,1),(2,3);(2,1),(3,2)}.

	1	2	3
1	(1,1)	(1,2)	(1,3)
2	(2,1)	(2,2)	(2,3)
3	(3,1)	(3,2)	(3,3)

отношение R_2^*

Чтобы найти замыкание относительно транзитивности, необходимо выполнить несколько шагов. Так как R содержит пары (3, 1) и (1, 2), замыкание обязано включать в себя и пару (3, 2). Аналогично, пары (2, 3) и (3, 1) добавляют пару (2, 1), а пары (3, 1) и (1, 3) — пару (3, 3). Добавим сначала эти пары:

$$\{(1,1),(1,2),(1,3),(3,1),(2,3);(3,2),(2,1),(3,3)\}\subset R^*$$

Теперь у нас возникло сочетание (2, 1) и (1, 2). Стало быть, замыкание R* должно содержать пару (2, 2). Теперь можно увидеть, что все необходимые пары мы добавили. Следовательно,

R_3^*={(1,1),(1,2),(1,3),(3,1),(2,3);(3,2),(2,1),(3,3),(2,2)}.

В данном примере мы получили, что $R_3^* = A \times A$.

∎

Теорема. Пусть R – отношение на множестве A. Тогда

1. $R \bigcup I$ есть рефлексивным замыканием R;
2. $R \bigcup R^{-1}$ есть симметричное замыкание R.

Доказательство предоставляем читателю.

3.3 Отношение эквивалентности

Важную роль в математике играет один специальный класс отношений – отношения эквивалентности. Отношение R на A есть отношение эквивалентности, если оно рефлексивно, симметрично и транзитивно.

Точнее, отношение R на множестве A называется **отношением эквивалентности**, если оно обладает следующими свойствами: 1) $a\,R\,a$ для любого $a \in A$ (рефлексивность); 2) если $a\ R\ b$, то $b\ R\ a$ (симметричность); 3) если $a\ R\ b$ и $b\ R\ c$, то $a\ R\ c$ (транзитивность). Это отношение часто обозначают символом ~, т.е. пишут $a \sim b$ и читают «a эквивалентно b».

Пример 1. Пусть A = {1,2,3,4,5,6} и задано отношение $R \subset A \times A$:

R={(1,1),(2,2),(3,3),(4,4),(5,5),(6,6),(1,2),(1,4),
(2,1),(2,4),(3,5),(5,3),(4,1),(4,2)}.

В примере 1 предыдущего параграфа мы установили, что это отношение рефлексивно, транзитивно и симметрично, поэтому R есть отношение эквивалентности на множестве A.

Пример 2. Пусть A – множество целых чисел. Определим отношение $R \subset A \times A$ посредством $R = \{(a,b): a-b = 5 \cdot k \text{ для } k \in \mathbf{Z}\}$, т.е. два целых числа эквивалентны, если их разность кратна 5. Например, $(7,2) \in R$, поскольку $7-2=5=5 \cdot 1$, и $(-11,4) \in R$, так как $-11-4=-15=5 \cdot (-3)$.

Отношение R рефлексивно. Действительно, если a – целое число (т.е. $a \in \boldsymbol{Z}$), то $a-a=0=5 \cdot 0$, так что $(a,a) \in R$.

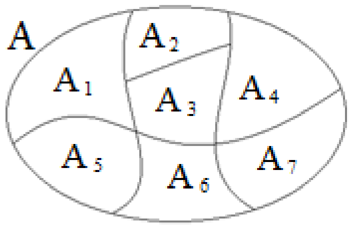
Отношение R симметрично. Действительно, если $(a,b) \in R$, то существует такое целое число m, что $a-b=5 \cdot m$. Тогда $b-a=-(a-b)=-(5 \cdot m)=5 \cdot (-m)$, а это значит, что $(b,a) \in R$.

Отношение R транзитивно. Предположим, что a, b и c – целые числа и $(a,b) \in R$, $(b,c) \in R$. Тогда имеем $a-b=5 \cdot k$ для некоторого целого k, и $b-c=5 \cdot m$ для некоторого целого m. Суммируем эти равенства $(a-b)+(b-c)=5 \cdot k+5 \cdot m$, т.е. $a-c=5 \cdot (k+m)$ для целого числа k+m. Значит $(a,c) \in R$, т.е. введенное отношение транзитивно.

Поскольку R рефлексивно, симметрично и транзитивно, оно является отношением эквивалентности.

∎

Если множество A представлено тем или иным способом как объединение своих попарно непересекающихся подмножеств, то говорят о его разбиении, а входящие в это разбиение подмножества называют классами разбиения.



Например, множество точек плоскости можно разбить на прямые, параллельные оси X; множество целых чисел разбивается на классы, состоящие из чисел, дающих один и тот же остаток при делении на фиксированное натуральное число n. В первом случае имеем бесконечное семейство классов, во втором случае классов разбиения будет ровно n.

Более точно. Пусть A и Ω – множества и пусть $\langle A \rangle = \{A_\alpha : \alpha \in \Omega, \Omega \neq \varnothing\}$ есть множество непустых подмножеств множества A. Множество $\langle A \rangle$ называется **разбиением** A, если выполнены два условия:

a) $A_\alpha \cap A_\beta = \varnothing$ для всех $\alpha \neq \beta$;

b) $A = \bigcup_{\alpha \in \Omega} A_\alpha$, т.е. a принадлежит A тогда и только тогда, когда $a \in A_\alpha$ для некоторого $\alpha \in \Omega$.

Отношение эквивалентности R на множестве A разбивает его (множество A) на подмножества, элементы которых эквивалентны друг другу и не эквивалентны элементам других подмножеств.

Например, пусть множество A — это набор разноцветных шаров, а отношение R задается условием: $(a,b) \in R$ тогда и только тогда, когда a и b имеют одинаковый цвет. Поскольку R — отношение эквивалентности (проверьте это), каждый класс эквивалентности будет состоять из шаров, имеющих одинаковый цвет.

Докажем, что каждое отношение эквивалентности на множестве A задает некоторое вполне определенное разбиение A. Пусть K_a — класс элементов из A, эквивалентных фиксированному элементу a, т. е. $K_a = \{ c \in A : c \sim a \}$. В силу свойства рефлексивности $a \in K_a$. Покажем, что любые два класса K_a и K_b, имеющие общий элемент, совпадают. Пусть элемент c принадлежит одновременно K_a и K_b, т. е. $c \sim a$ и $c \sim b$. В силу симметричности $a \sim c$, и в то же время $c \sim b$. Используя транзитивность, заключаем, что $a \sim b$ (*). Пусть теперь x — произвольный элемент класса K_a. Тогда $x \sim a$, и в силу (*) и свойства транзитивности $x \sim b$, т. е. $x \in K_b$. Этим доказано включение $K_a \subset K_b$. Точно так же доказывается, что $K_b \subset K_a$. В итоге $K_b = K_a$.

Итак, мы видим, что всякий элемент из A лежит в некотором классе и различные классы не пересекаются. Тем самым получается разбиение множества A на классы эквивалентных элементов.

∎

Рассмотрим теперь произвольное разбиение множества A на классы. Два элемента a и b из A будем считать эквивалентными ($a \sim$ b), если они лежат в одном классе. Легко проверяется, что ~ есть отношение эквивалентности на множестве A.

Таким образом, отношение эквивалентности дает универсальный способ разбиения множества на классы, а любое разбиение позволяет построить отношение эквивалентности, соответствующее этому разбиению.

Множество $\{ x : x\,R\,a \} = \{ x : (x,a) \in R \}$, называют классом эквивалентности, содержащим a, и часто обозначают $[a]$. Символ $[A]_R$ обозначает множество всех классов эквивалентности множества A по отношению R.

Пример. В примере 1 этого параграфа было показано, что отношение R определенное на множестве A={1,2,3,4,5,6}:

R={1,1),(2,2),(3,3),(4,4),(5,5),(6,6),(1,2),(1,4),
(2,1),(2,4),(3,5),(5,3),(4,1),(4,2)}

есть отношение эквивалентности. Классы эквивалентности для каждого элемента A имеют вид:

$$[1]=\{1,2,4\}$$
$$[2]=\{2,1,4\}$$
$$[3]=\{3,5\}$$
$$[4]=\{4,1,2\}$$
$$[5]=\{5,3\}$$
$$[6]=\{6\}$$

Итак, имеется только три различных класса эквивалентности: [1] = [2] =[4] = {1,2,4}, [3]=[5]={3,5}, [6]={6}. Таким образом,
$$[A]_R=\{[1],[3],[6]\}=\{\{1,2,4\},\{3,5\},\{6\}\}.$$
∎

Этот пример показывает, что, если $b\in[a]$, то $[a]$ = $[b]$. На основании этого свойства говорят, что любой элемент класса эквивалентности представляет класс.

Каждый класс эквивалентности содержит, по крайней мере, один элемент, поэтому, в силу рефлексивности отношения, множество всех элементов, эквивалентных элементу a, должно содержать a. С другой стороны, никакой элемент не может принадлежать двум различным классам эквивалентности.

Пример. Рассмотрим отношение эквивалентности R из примера 2. Для множества A всех целых чисел отношение $R\subset A\times A$ было определено посредством $R=\{(a,b): a-b=5\cdot k \text{ для } k\in\mathbf{Z}\}$. Поскольку
$$[a]=\{x:(x,a)\in R\}=\{x: x\,R\,a\}=$$
$$=\{x: x-a=5\cdot k \text{ для } k\in\mathbf{Z}\}=$$
$$=\{x: x=a+5\cdot k \text{ для } k\in\mathbf{Z}\}$$
получаем, что множества
$$[0]=\{\ldots,-15,-10,-5,0,5,10,15,20,25,\ldots\}=[-5]=[0]=[5]=[10]=\ldots$$
$$[1]=\{\ldots,-14,-9,-4,1,6,11,16,\ldots\}=[-9]=[-4]=[1]=[6]=\ldots$$
$$[2]=\{\ldots,-13,-8,-3,2,7,12,17,\ldots\}=[-8]=[-3]=[2]=[7]=\ldots$$
$$[3]=\{\ldots,-12,-7,-2,3,8,13,18,\ldots\}=[-7]=[-2]=[3]=[8]=\ldots$$
$$[4]=\{\ldots,-11,-6,-1,4,9,14,19,\ldots\}=[-6]=[-1]=[4]=[9]=\ldots$$
представляют собой различные классы эквивалентности по отношению R. Таким образом,
$$[A]_R=\{[0],[1],[2],[3],[4]\}.$$
Элементы класса [0] «похожи» в том смысле, что каждый из них кратен пяти. Элементы любого другого класса эквивалентности также «похожи» в том смысле, что имеют один и тот же остаток при делении на пять.

3.4 Реляционная алгебра

Как говорилось ранее в этой главе, каждую таблицу можно трактовать как отношение. А любое отношение, как подмножество декартового произведения $A_1 \times A_2 \times \ldots \times A_n$, может быть изображено в виде таблицы.

A_1	A_2	...	A_n
a_1^1	a_2^1	$\cdots$	a_n^1
a_1^2	a_2^2	$\cdots$	a_n^2
...	...	...	...

С одной стороны таблица – это способ перечислить все элементы некоторого отношения $R \subset A_1 \times A_2 \times \ldots \times A_n$. С другой, если есть таблица, то она может рассматриваться как способ представления некоторого отношения. Для этого она должна представлять некоторое подмножество в декартовом произведении. Поскольку каждый элемент множества уникальный, то в таблице не должно быть повторяющихся строк. Кроме того, строки таблицы не должны быть упорядочены (элементы множества неупорядочены). В реляционной алгебре на таблицы накладывается еще одно условие – порядок столбцов не должен нести в себе никакой информации.

Элементы декартового произведения множеств $A_1 \times A_2 \times \ldots \times A_n$ называются кортежами, т.е. строка таблицы – это кортеж. Имена множеств A_i называют атрибутами (свойствами). Само множество A_i называют доменом $i-$го атрибута (т.е. доменом $i-$го атрибута называют множество всех допустимых значений $a_i^j \in A_i$). Часто имена атрибутов совпадают с именами соответствующих доменов. Количество атрибутов называется степенью (или арностью) отношения, а количество кортежей в отношении (строк в таблице) называется мощностью отношения. Степень отношения R мы будем обозначать $\|R\|$, а мощность – $|R|$. Все множество строк таблицы называют телом отношения.

Например, экзаменационная ведомость – это таблица со столбцами ФИО (фамилия, имя, отчество), НЗ (номер зачетки), ОЦЕНКА. Названия столбцов являются атрибутами, а их набор называется схемой отношения. Доменом атрибута ОЦЕНКА является множество всех возможных оценок, т.е. $\{A, B, C, D, E, F, FX\}$. Заполненная ведомость называется текущим значением отношения.

Отношения/таблицы являются множествами и над ними можно выполнять теоретико-множественные операции. В результате выполнения операций над множествами – отношениями получаются новые множества – отношения. Когда отношения (по-английски `relation`) выступают в роли операндов операций над множествами, результатом которых являются новые отношения, то говорят о реляционных операциях.

Любая операция, результатом которой является отношение, является реляционной операцией. Реляционная алгебра представляет собой набор реляционных операций над отношениями и законы их комбинирования.

Операции над одним отношением называются унарными, над двумя отношениями – бинарными, а n-арную реляционную операцию f можно считать функцией, возвращающей отношение и имеющей n отношений в качестве аргументов $R = f(R_1, R_2, ..., R_n)$.

Основные операции реляционной алгебры обычно разбивают на две группы. Теоретико-множественные операции: *объединение*, *пересечение*, *вычитание*, *декартово произведение*. Специальные реляционные операции: *выборка*, *проекция*, *соединение*, *деление*. Для этих операций реляционной алгебры используются специальные обозначения, которые могут отличаться от обозначений обычных операций над множествами.

На операции реляционной алгебры накладываются некоторые естественные ограничения. Например, «объединение» двух отношений/таблиц предполагает, что отношения являются подмножествами одного и того же декартового произведения $A_1 \times A_2 \times ... \times A_n$, т.е. соответствующие таблицы имеют одни и те же атрибуты (заголовки, схемы). Говорят, что отношения совместимы по типу, если они имеют идентичные атрибуты или, что тоже, являются подмножествами декартового произведения одних и тех же множеств $A_1, A_2, ..., A_n$, возможно перечисленных в ином порядке. Т.о. операция «объединение» выполняется над совместимыми по типу отношениями.

Объединением двух совместимых по типу отношений A и B называется отношение с теми же атрибутами, что у A и B, и телом, состоящим из кортежей, принадлежащих или A, или B, или обоим одновременно. В частности, если какой-либо кортеж содержится в нескольких таблицах, в результирующей таблице он появится только один раз. Операцию объединения обозначают следующим образом

A UNION B (эквивалент обозначения $A \cup B$)

Пример 1. Пусть даны два отношения A и B – ведомости сдачи экзаменов. Отношение A – ведомость сдачи экзамена по ДМ.

Фамилия	Оценка
Ткачук	B
Крохмаль	E
Бондаренко	A

Отношение B – ведомость пересдачи экзамена по ДМ.

Фамилия	Оценка
Сидоров	A
Бездетко	C
Крохмаль	E

Объединением отношений **A UNION B** будет отношение C:

Фамилия	Оценка
Ткачук	B
Крохмаль	E
Бондаренко	A
Сидоров	A
Бездетко	C

Очевидно, что для операции объединения выполняются обычные свойства: $A \bigcup B = B \bigcup A$, $A \bigcup (B \bigcup C) = (A \bigcup B) \bigcup C$, $A \bigcup \varnothing = A$. Для степени объединения отношений выполняются равенства: $\| A \| = \| B \| = \| A \bigcup B \|$. Для мощности выполняются неравенства: $| A \bigcup B | \le | A | + | B |$, $| A \bigcup B | \ge | A |$, $| A \bigcup B | \ge | B |$.

Пересечением двух совместимых по типу отношений A и B называется отношение с теми же атрибутами, что у отношений A и B и телом, состоящим из кортежей, принадлежащих одновременно обоим отношениям A и B. Синтаксис операции пересечения:

A INTERSECT B (эквивалент обозначения $A \bigcap B$)

Пример 2. Для тех же отношений A и B, что и в предыдущем примере пересечение **`A INTERSECT B`** имеет вид:

Фамилия	Оценка
Крохмаль	E

∎

Для операции пересечения выполняются обычные свойства: $A \bigcap B = B \bigcap A$, $A \bigcap (B \bigcap C) = (A \bigcap B) \bigcap C$, $A \bigcap \varnothing = \varnothing$. Для степени пересечения отношений выполняются равенства: $\| A \| = \| B \| = \| A \bigcap B \|$. Для мощности выполняются неравенства: $| A \bigcap B | \le | A |$, $| A \bigcap B | \le | B |$.

Вычитанием (разностью) двух совместимых по типу отношений A и B называется отношение с теми же атрибутами, что у отношений A и B, и телом, состоящим из кортежей, принадлежащих отношению A и не принадлежащих отношению B. Синтаксис операции вычитания:

A MINUS B (эквивалент обозначения $A \setminus B$)

Пример 3. Для тех же отношений A и B, что и в примере 1 результат вычитания **`A MINUS B`** имеет вид:

Фамилия	Оценка
Ткачук	B
Бондаренко	A

∎

Для операции вычитания справедливо свойство $A \setminus \varnothing = \varnothing$. Для степени разности отношений выполняются равенства $\| A \| = \| B \| = \| A \setminus B \|$. Для мощности справедливо неравенство $| A \setminus B | \le | A |$.

Декартовым произведением двух отношений $R \subset A_1 \times A_2 \times \ldots \times A_n$ и $S \subset B_1 \times B_2 \times \ldots \times B_m$ называется отношение $Q \subset A_1 \times \ldots \times A_n \times B_1 \times \ldots \times B_m$, получаемое как декартовое произведение множеств $R \times S$. Отношение Q состоит из кортежей вида $(a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_m)$ таких, что $(a_1, a_2, \ldots, a_n) \in R$ и $(b_1, b_2, \ldots, b_m) \in S$. Синтаксис операции декартового произведения:

R TIMES S (эквивалент обозначения $R \times S$)

Перемножать можно любые два отношения, совместимость по типу при этом не требуется.

Пример 4. Пусть даны два отношения R и S с информацией о поставщиках и деталях:

Отношение R (Поставщики)
Иванов
Петров
Сидоров

Отношение S (Изделия)
Блок
Плита
Балка

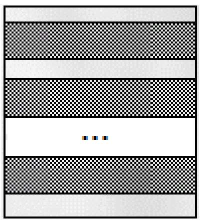
Декартово произведение отношений R и S будет иметь вид:

Поставщики	Изделия
Иванов	Блок
Иванов	Плита
Иванов	Балка
Петров	Блок
Петров	Плита
Петров	Балка
Сидоров	Блок
Сидоров	Плита
Сидоров	Балка

■

Фактически из двух таблиц составляется новая, в которой каждый кортеж первой таблицы сцепляется с каждым кортежем второй. Множество атрибутов результирующей таблицы получается объединением множеств атрибутов обеих таблиц. Схема результирующей таблицы получается объединением схем таблиц-аргументов так, что ее степень равна сумме степеней аргументов $\| R \times S \| = \| R \| + \| S \|$. Мощность декартового произведения равна произведению мощностей сомножителей $| R \times S | = | R | \cdot | S |$.

Выборкой (селекцией) на отношении A с условием α называется отношение с теми же атрибутами, что у отношения A, и телом, состоящим из кортежей, значения атрибутов которых при подстановке в условие α дают значение ИСТИНА. α представляет собой логическое выражение с аргументами, которыми могут быть атрибуты отношения A и/или скалярные выражения. В аргументы условия α подставляются значения полей кортежа из A (значения из ячеек таблицы A) и, если результатом вычисления условия α является ИСТИНА, то кортеж выбирается.



Т.о. тело отношения – результата является подмножеством тела отношения A, а схема таблицы не изменяется. Синтаксис операции выборки:

`A WHERE Условие` (или **`A[условие]`**)

Очевидно, что $\| \mathrm{A}[\text{условие}] \| = \| \mathrm{A} \|$ и $| \mathrm{A}[\text{условие}] | \le | \mathrm{A} |$.

Пример. Для отношения С из примера 1 операция выборки

`C WHERE Оценка=A`

вернет отношение/таблицу

Фамилия	Оценка
Бондаренко	А
Сидоров	А

Пример 5. Пусть отношение *R* задано таблицей:

A1	**A2**	**A3**	**A4**
x	2000	400	*a*
y	3000	600	*a*
p	1500	100	*b*
p	700	200	*b*
p	2500	400	*b*

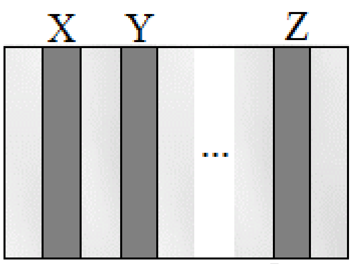
Операция выборки

`R WHERE A2<A2`

вернет пустую таблицу (отношение).

∎

Проекцией отношения $\mathrm{R} \subset \mathrm{A}_1 \times \ldots \times \mathrm{X} \times \ldots \times \mathrm{Y} \times \ldots \times \mathrm{A}_k \times \ldots \times Z \times \ldots \times \mathrm{A}_\mathrm{n}$ по атрибутам $\mathrm{X}, \mathrm{Y}, \ldots, Z$, где каждый из атрибутов принадлежит отношению R, называется отношение с атрибутами $\mathrm{X}, \mathrm{Y}, \ldots, Z$ и телом, содержащим множество кортежей исходного отношения с возможным удалением повторяющихся кортежей (при сокращении набора атрибутов некоторые кортежи могут совпасть и появляющиеся «дублирования» исключаются).



Синтаксис операции проекции: $\mathrm{R}[\mathrm{X}, \mathrm{Y}, \ldots, \mathrm{Z}]$. Очевидно, что $\| \mathrm{R}[\mathrm{X}, \mathrm{Y}, \ldots, \mathrm{Z}] \| < \| \mathrm{R} \|$ и $| \mathrm{R}[\mathrm{X}, \mathrm{Y}, \ldots, \mathrm{Z}] | \le | \mathrm{R} |$.

Пример 6. Пусть отношение/таблица R имеет вид:

Фамилия	Оклад	Подоходный налог	Налог на бездетность	Надбавка	Итого
Иванов	2000	400	0	500	2100
Петров	3000	600	0	700	3100
Сидоров	4000	800	100	400	3500

Тогда операция **`R[Фамилия, Итого]`** вернет отношение/таблицу

Фамилия	Итого
Иванов	2100
Петров	3100
Сидоров	3500

Пример 7. Пусть R задано таблицей из примера 5. Тогда

$$R[A_4]=\begin{bmatrix}a\\a\\b\\b\\b\end{bmatrix}=\begin{bmatrix}a\\b\end{bmatrix}\text{ – после чистки;}\quad R[A_1,A_4]=\begin{bmatrix}x&a\\y&a\\p&b\\p&b\\p&b\end{bmatrix}=\begin{bmatrix}x&a\\y&a\\p&b\end{bmatrix}.$$

Операция *соединения* отношений имеет несколько форм. Соединением (общим) отношений A и B по условию α называется отношение

(A TIMES B) WHERE α,

где α представляет собой логическое выражение, в которое могут входить атрибуты отношений A и B и/или скалярные выражения. Таким образом, операция соединения есть результат последовательного применения операций декартового произведения и выборки. По-другому, соединение это выборка над декартовым произведением. Если в отношениях A и B имеются атрибуты с одинаковыми именами, то перед выполнением соединения такие атрибуты необходимо переименовать.

В зависимости от условия α рассматривают частные случаи соединений: естественные соединения, тэта – соединения и экви – соединения. Наиболее важным из этих частных случаев является операция естественного соединения. Операция *естественного соединения* объединяет две таблицы в большую, выписывая в одну строку информацию, соответствующую общим атрибутам. Если R отношение в декартовом произведении $\mathrm{R}\subset \mathrm{A}_1\times\ldots\times \mathrm{A}_m\times \mathrm{B}_1\times\ldots\times \mathrm{B}_n$, S отношение в декартовом произведении $\mathrm{S}\subset \mathrm{A}_1\times\ldots\times \mathrm{A}_m\times \mathrm{C}_1\times\ldots\times \mathrm{C}_p$, и R, S имеют общие атрибуты $\mathrm{A}_1,\mathrm{A}_2,\ldots,\mathrm{A}_m$, то естественное соединение R и S – это подмножество в декартовом произведении $\mathrm{A}_1\times\ldots\times \mathrm{A}_m\times \mathrm{B}_1\times\ldots\times \mathrm{B}_n\times \mathrm{C}_1\times\ldots\times \mathrm{C}_p$, состоящее из элементов вида $(a_1,\ldots,a_m,b_1,\ldots,b_n,c_1,\ldots,c_p)$, где $(a_1,\ldots,a_m,b_1,\ldots,b_n)\in R$ и $(a_1,\ldots,a_m,c_1,\ldots,c_p)\in S$.

Обозначается естественное соединение так:

R JOIN S.

В синтаксисе естественного соединения не указываются, по каким атрибутам производится соединение. Оно производится по всем одинаковым атрибутам.

Пример 8. Пусть отношение A – ведомость сдачи экзамена по ДМ (дискретная математика).

Фамилия	**ДМ**
Ткачук	B
Крохмаль	E
Бондаренко	A

Пусть отношение B – ведомость сдачи экзамена по ВМ (высшая математика).

Фамилия	**ВМ**
Ткачук	A
Бондаренко	C
Крохмаль	E

Естественным соединением отношений **A JOIN B** (по атрибуту `Фамилия`) будет отношение С:

Фамилия	**ДМ**	**ВМ**
Ткачук	B	A
Крохмаль	E	E
Бондаренко	A	C

∎

Легко проверить, что естественное соединение (как, впрочем, и соединение общего вида) обладает свойством ассоциативности, т.е.

(R JOIN S)JOIN Q = R JOIN (S JOIN Q)

Поэтому это можно записывать, опуская скобки: **R JOIN S JOIN Q.**

Пусть отношение А содержит атрибут X, отношение В содержит атрибут Y, а θ – один из операторов сравнения $(=, \neq, <, \leq, >, \geq)$. Тогда θ – соединением отношения A по атрибуту X с отношением B по атрибуту Y называют соединение отношений вида

(A TINES B) WHERE X θ Y.

Иногда для операции θ – соединения используют обозначение $A[X\theta Y]B$.

Пример. Пусть отношение *R* задано таблицей из примера 5, а S отношение задано таблицей:

A5	**A6**
700	*a*
400	*c*
200	*d*
500	*b*

Тогда отношение **(R TINES S) WHERE A3 > A5** имеет вид:

A1	**A2**	**A3**	**A4**	**A5**	**A6**
x	2000	400	*a*	200	*d*
y	3000	600	*a*	400	*c*
y	3000	600	*a*	200	*d*
y	3000	600	*a*	500	*b*
p	2500	400	*b*	200	*d*

∎

Частным случаем θ – соединения является экви – соединение:

(A TINES B) WHERE X=Y
или
$A[X=Y]B$.
Т.е. это соединение с проверкой на равенство. Недостатком этого соединения является то, что если соединение происходит по атрибутам с одинаковыми наименованиями (а так чаще всего бывает), то в результирующем отношении появляется два атрибута с одинаковыми значениями. Избавиться от этого недостатка можно, взяв проекцию по всем атрибутам, кроме одного из повторяющихся. Именно так действует естественное соединение, которое мы описали выше.
Пример. Пусть отношения R и S заданы таблицами из предыдущего примера. Тогда **(R TINES S) WHERE A3 = A5** имеет вид:

A1	**A2**	**A3**	**A4**	**A5**	**A6**
x	2000	400	*a*	400	*c*
p	700	200	*b*	200	*d*
p	2500	400	*b*	400	*c*

∎

Операция *соединения* отношений, наряду с операциями выборки и проекции, является одной из наиболее важных реляционных операций. Мы не будем описывать все ее варианты. С ними вы можете познакомиться самостоятельно в литературе по теории реляционных баз данных.

Пусть даны отношения $R \subset A_1 \times \ldots \times A_n \times B_1 \times \ldots \times B_m$ и $S \subset B_1 \times \ldots \times B_m$, имеющие общие атрибуты $B_1,\ldots,B_m$. *Делением* отношений R на S называется отношение Q с атрибутами $A_1,\ldots,A_n$ и телом, содержащим множество кортежей $(a_1,\ldots,a_n)$ таких, что для всех кортежей $(b_1,\ldots,b_m) \in S$ в отношении R найдется кортеж $(a_1,\ldots,a_n,b_1,\ldots,b_m)$. Это значит, что $Q \times S \subset R$. Синтаксис операции деления:
R DIVIDEBY S (эквивалент обозначения R/S)
Можно сказать, что отношение Q – это наибольшее (по множеству кортежей) отношение, произведение которого на S содержится в R.

Пример. В качестве делимого возьмем таблицу R, содержащую название поставщика и изделие, которое он поставляет.

Поставщик	Изделие
ООО “Железяка”	Плита
ТОВ “Сталь”	Плита
ЧП “Бондаренко”	Плита
ЧП “Бондаренко”	Блок
ТОВ “Сталь”	Блок
ЧП “Бондаренко”	Балка

В качестве делителя возьмем таблицу S наименований всех изделий

Изделие
Плита
Блок
Балка

Деление R DIVIDEBY S даст таблицу Q поставщиков, которые поставляют все типы изделий.

Изделие
ЧП “Бондаренко”

∎

Как мы уже говорили, традиционно определяют восемь реляционных операций, объединенных в две группы. Теоретико-множественные операции: объединение, пересечение, вычитание, декартово произведение. Специальные реляционные операции: выборка, проекция, соединение, деление. Для выполнения некоторых реляционных операций требуется, чтобы отношения были совместимы по типу. Операции объединения, вычитания, декартова произведения, выборки и проекции называются примитивными – их нельзя выразить через другие реляционные операции. Оставшиеся реляционные операции соединения, пересечения и деления можно выразить через другие, т.е. эти операции не являются примитивными.

Реляционная алгебра представляет собой набор операций (и их свойств), использующих отношения в качестве аргументов, и возвращающих отношения в качестве результата. Результаты одних реляционных выражений можно использовать в других выражениях. При этом в выражениях реляционной алгебры можно использовать скобки. Некоторые из свойств реляционных операций мы привели ранее при их описании. Здесь мы приведем без доказательства еще несколько формул реляционной алгебры.

Теорема 1. Пусть R и Q два совместимых отношения, X один из атрибутов отношения R и Y один из атрибутов отношения Q. Тогда справедливы следующие утверждения:

1. Выборка из объединения отношений равна объединению выборок, т.е.

$$(\mathrm{R} \cup \mathrm{Q})[\mathrm{X}\,\theta\,\mathrm{Y}] = \mathrm{R}[\mathrm{X}\,\theta\,\mathrm{Y}] \cup \mathrm{Q}[\mathrm{X}\,\theta\,\mathrm{Y}].$$

Здесь через $R[\alpha]$ обозначена операция выборки на отношении R с условием α и θ представляет один из операторов сравнения.

2. Выборка из пересечения отношений равна пересечению выборок, т.е.

$$(R \cap Q)[\mathrm{X}\,\theta\,\mathrm{Y}] = R[\mathrm{X}\,\theta\,\mathrm{Y}] \cap Q[\mathrm{X}\,\theta\,\mathrm{Y}].$$

3. Выборка из разности отношений равна разности выборок, т.е.

$$(R \setminus Q)[\mathrm{X}\,\theta\,\mathrm{Y}] = R[\mathrm{X}\,\theta\,\mathrm{Y}] \setminus Q[\mathrm{X}\,\theta\,\mathrm{Y}].$$

∎

Теорема 2. Пусть R и Q два совместимых отношения и $\Omega = \{X, Y, \ldots, Z\}$ некоторые атрибуты этих отношений. Тогда справедливы следующие утверждения:

1. Проекция объединения отношений равна объединению проекций, т.е.

$$(\mathrm{R} \cup \mathrm{Q})[\Omega] = \mathrm{R}[\Omega] \cup \mathrm{Q}[\Omega].$$

Здесь через $R[\Omega]$ мы обозначаем проекцию отношения R по атрибутам Ω.

2. Проекция пересечения отношений равна пересечению проекций, т.е.

$$(\mathrm{R} \cap \mathrm{Q})[\Omega] = \mathrm{R}[\Omega] \cap \mathrm{Q}[\Omega].$$

3. Проекция разности отношений равна разности проекций, т.е.

$$(R \setminus Q)[\Omega] = R[\Omega] \setminus Q[\Omega].$$

∎

Теорема 3. Для произвольных совместимых по типу отношений R и Q выполняются равенства:

1. $(R[X\theta Y]Q)[G\theta H] = (R[G\theta H])[X\theta Y]Q$,

где через $R[X\theta Y]Q$ мы обозначаем θ – соединение отношения R по атрибуту X с отношением Q по атрибуту Y. На полученном отношении выполняется выборка с условием G θ H. При этом предполагается, что G и H – атрибуты отношения R, не являющиеся атрибутами отношения Q. Кратко это свойство можно сформулировать так: выборка на соединении отношений R и Q равняется соединению выборки из R с отношением Q.

2. $(R[X_1, X_2, ..., X_k])\lfloor X_i \theta X_j \rfloor = (R\lfloor X_i \theta X_j \rfloor)[X_1, X_2, ..., X_k]$,

где $i, j \in 1, ..., k$. Здесь через $R[X_1, X_2, ..., X_k]$ обозначена проекция отношения R по атрибутам $X_1, X_2, ..., X_k$, а через $R\lfloor X_i \theta X_j \rfloor$ обозначена выборка на отношении R по условию $X_i \theta X_j$. Кратко это свойство формулируется так: выборка на проекции равняется проекции на выборке.

3. Операции выборки перестановочны, т.е.

$$(R[X\theta Y])[G\theta H] = (R[G\theta H])[X\theta Y],$$

где $R[X\theta Y]$ представляет выборку на отношении R по условию $X\theta Y$, а X, Y, G и H атрибуты отношения R.

∎

Заметим, что здесь перечислены далеко не все свойства операций реляционной алгебры.

Реляционной базой данных называется набор отношений. При использовании базы данных описанные операции применяются для формирования новых таблиц. Имеется специальный язык SQL (Structured Query Language – язык структурированных запросов), который можно использовать для создания новых таблиц из имеющихся с использованием перечисленных операций. Практическая значимость законов реляционной алгебры состоит в том, что при компьютерной реализации выражений, стоящих в левых и правых частях реляционных формул, требуется различное количество операций и у разработчиков системы управления базами данных есть возможность выбрать наиболее эффективное решение.

4. Отображения и функции.

4.1 Понятие отображения

Пусть X и Y — произвольные непустые множества. Будем говорить, что определено **отображение** множества X в множество Y, если каждому элементу *x* множества X поставлен в соответствие некоторый вполне определенный элемент *y* множества Y. Этот элемент *y* называется образом элемента *x* при данном отображении.

Отображение часто обозначают одной буквой (например, *f*), и тогда запись $f: X \to Y$ заменяет фразу «f — отображение множества X в множество Y». Образ элемента *x* при этом отображении обозначается символом f(x). Подчеркнем, что **каждый** $x \in X$ имеет **единственный** образ f(x).

Отображение $f: X \to Y$ можно представлять себе как некое действие, которое переводит элементы x в их образы $y = f(x) \in Y$.

Примеры.

1. Если каждому действительному числу *x* поставить в соответствие число x^2, то тем самым будет определено отображение $f: \mathbf{R} \to \mathbf{R}$ (и даже в $\mathbf{R} \to [0,+\infty)$).
2. Пусть $\mathbf{X} = \mathbf{Y} = \mathbf{R}$ – множество действительных чисел. Формула $y = 2x$ задает другое отображение $\mathbf{R} \to \mathbf{R}$.
3. Пусть X – множество всех треугольников на плоскости, а Y – множество действительных чисел. Сопоставляя треугольнику его площадь, получаем отображение первого множества во второе.
4. Пусть X={1,2,3}, Y={2,3,4}. Множество пар

G={(1,2),(2,2),(3,3)}

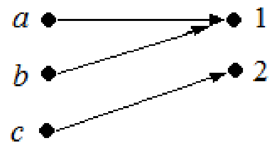
задает отображение f, при котором f(1)=f(2)=2, f(3)=3.

Понятия «функция» и «отображение» иногда отождествляют, пользуясь ими как синонимами. Все же чаще под функцией понимают отображение одного числового множества в другое.

Функции представляют из себя специальный тип бинарных отношений. По – другому определение функции можно дать следующим образом. Функцией из множества A в множество B называется бинарное отношение, при котором каждый элемент множества A связан с единственным элементом множества B. Другими словами, для каждого $a \in A$ существует ровно одна пара из отношения вида (*a*, *b*). В графических терминах функция описывается таким графом, у которого из каждой вершины, изображающей элементы множества A, выходит ровно одна стрелка.

Пример. Пусть A={-2,-1,0,1,2}, а B={0,1,2,3,4,5}. Определим отношение $f \subseteq A \times B$ f={(-2,5),(-1,2),(0,1),(1,2),(2,5)}. Отношение f — функция из A в B, так как $f \subseteq A \times B$, и каждый из элементов A присутствует в качестве первой компоненты упорядоченной пары из f ровно один раз.

Пример. На рисунке изображен граф, представляющий функцию из множества {a,b,c} в {1,2}, состоящую из пар (a, 1), (b, 1) и (c, 2).



Пример. Определим, какие из следующих отношений между множествами A={a,b,c} и B={1,2,3} являются функциями из множества A в B.

(а) f={(a,1),(a,2),(b,3),(c,2)};

(б) g={(a,1),(b,2),(c,1)};

(в) h ={(a,1),(c,2)}.

Решение.

(*а*) Отношение f — не функция, поскольку элементу a соответствуют два разных элемента 1 и 2 множества B .

(*б*) Отношение g является функцией.

(*в*) Последнее отношение функцией не является, поскольку элементу b не соответствует ни одного элемента.

Пример. Определим, какие из следующих бинарных отношений между множествами A в B являются функциями.

(а) «x — брат или сестра y» на множестве всех людей;

(б) отношение на множестве **R**, заданное парами: $\{(x,y):x=y^2\}$

Решение.

(а) Это не функция, поскольку есть люди с несколькими братьями и сестрами, а также бывают семьи с единственным ребенком.

(б) Последнее отношение – не функция, так как, например, обе упорядоченные пары: $\left(2,\sqrt{2}\right)$ и $\left(2,-\sqrt{2}\right)$ — ему принадлежат. Кроме того, в нем отсутствуют пары (x, y) с отрицательными x.

∎

Пусть $X=\{x_1,x_2,\ldots,x_n\}$ — конечное, а Y— произвольное множества. В этом случае отображение $f:X\to Y$ удобно задавать таблицей

$$\begin{pmatrix} x_1 & x_2 & \ldots & x_n \\ f(x_1) & f(x_2) & \ldots & f(x_n) \end{pmatrix}$$

где ниже элементов $x_i \in X$ указаны их образы $f(x_i)\in Y$. Например, если X={1,2,3,4}, то отображение $f:X\to Y$ с таблицей

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 1 \end{pmatrix}$$

переводит 1 в 3, 2 в 1, 3 в 4, 4 в 1. А таблица

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix}$$

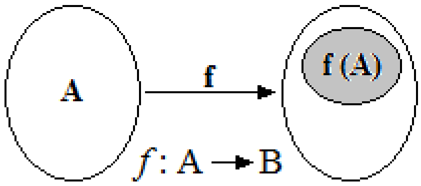
задает отображение множества четвертей координатной плоскости в себя при повороте плоскости на 90^0 против часовой стрелки вокруг начала координат (вместо самих четвертей в этой таблице мы используем их номера).

Два отображения $f: X \to Y$ и $g: U \to V$ считаются равными, если X=U, Y=V и для каждого элемента $x \in X = U$ выполняется равенство f(x)=g(x).

Пусть $f: X \to Y$. Множество $f(X) = \{ f(x) : x \in X\}$ образов всех элементов из X называется образом множества X при отображении f. Аналогично определяется образ любого подмножества A множества X:

$$f(A) = \{ f(x) : x \in A\}.$$

Функцию, определенную на множестве A со значениями в множестве B удобно представить в виде диаграммы Эйлера.



Пусть y — фиксированный элемент из Y. Подмножество

$$f^{-1}(y) = \{ x \in X : f(x) = y\}$$

множества X, состоящее из всех тех $x \in X$, для которых y является образом при отображении f, называется полным прообразом элемента y при отображении f. Каждый элемент из $f^{-1}(y)$ называют прообразом элемента y при отображении f. Может случиться, что $f^{-1}(y) = \varnothing$ для некоторого $y \in Y$.

Если B — произвольное подмножество Y, то его полный прообраз определяется следующим образом

$$f^{-1}(B) = \{ x \in X : f(x) \in B\}$$

Пример. Рассмотрим отображение $f: \mathbf{R} \to \mathbf{R}$, где $f(x) = x^2$ для любого $x \in \mathbf{R}$. Тогда $f(\mathbf{R}) = \{x \in \mathbf{R} : x \geq 0\}$, $f^{-1}(1) = \{1, -1\}$, $f^{-1}(0) = \{0\}$, $f^{-1}(-1) = \varnothing$.

Пример. Пусть A={-2,-1,0,1,2}, B={0,1,2,3,4,5} и функция $f: \mathrm{A} \to \mathrm{B}$ определена соотношением $f(x) = x^2 + 1$. Для $E = \{1, 2\} \subset A$ имеем $f(E) = \{2{,}5\}$ и множество $\{2{,}5\}$ является образом E при отображении f. Если $\mathrm{F} = \{0{,}2{,}3{,}4{,}5\} \subset B$, то множество

$$f^{-1}(F) = \{ y : \exists x \in A : f(x) = y \} = \{ -1, 1, -2, 2 \}.$$

является прообразом F. Заметим, что элементы 0, 3 и 4 не вносят никаких элементов в $f^{-1}(F)$, поскольку они не принадлежат области значений функции f. Прообраз может быть пустым. Так, например, в случае W={0,3} прообраз $f^{-1}\{W)$ пуст, поскольку не существует такого $x \in A$, для которого $f(x) = 0$ или $f(x) = 3$.

∎

Отображение $f: \mathrm{X} \to \mathrm{Y}$ называется инъективным или инъекцией, если для любых $x_1, x_2 \in X$ из $x_1 \neq x_2 \Rightarrow f(x_1) \neq f(x_2)$.

Другими словами, f инъективно, если разные элементы множества X имеют разные образы при отображении f.

Отображение $f: \mathrm{X} \to \mathrm{Y}$ называется сюръективным или сюръекцией, если f(X)=Y, т. е. если каждый элемент множества Y имеет хотя бы один прообраз.

Отображение, которое одновременно инъективно и сюръективно, называется биективным или биекцией. Часто биекцию называют взаимно однозначным отображением. Очевидно, $f: \mathrm{X} \to \mathrm{Y}$ биективно тогда и только тогда, когда каждый элемент множества Y имеет точно один прообраз.

Пример. Пусть X={1,2,3}, Y={1,2}. Тогда отображение $f: X \to Y$, где

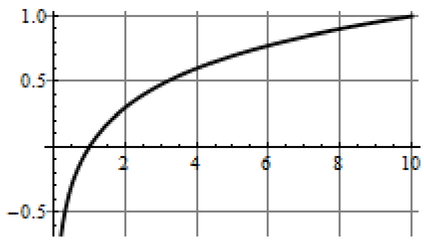
$$f = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 1 \end{pmatrix}$$

сюръективно, но не инъективно, а отображение $g: Y \to X$, где

$$f = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$$

инъективно, но не сюръективно.

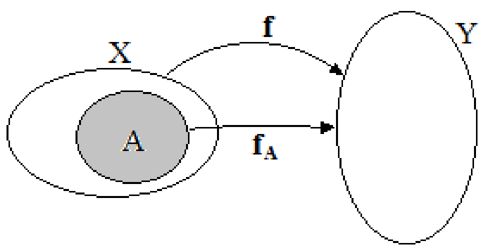
Пример. Рассмотрим функцию $y = \lg x$. Ее область определения есть множество всех положительных действительных чисел. Поэтому она задает отображение $f: (0, +\infty) \to \mathbf{R}$.



Легко понять, что f — биективное отображение.

∎

Пусть задано отображение $f: X \to Y$ и A некоторое подмножество X. Тогда можно рассмотреть отображение f_A заданное на A: $A \to Y$, определяемое равенством $f_A(x) = f(x)\ \forall x \in A$. Отображение f_A называется сужением отображения f.



Часто встречаются отображения множеств в себя, т. е. отображения вида $f: X \to X$. Такие отображения называют преобразованиями множества X. Биективные преобразования $f: X \to X$ называют подстановками множества X.

Преобразование $I_X: X \to X$, такое, что $I_X(x) = x$ для любого $x \in X$, называется тождественным преобразованием множества X. Таким образом, I_X

оставляет на месте каждый элемент из X. Часто вместо I_X пишут просто I, если из контекста ясно, какое множество X имеется в виду.

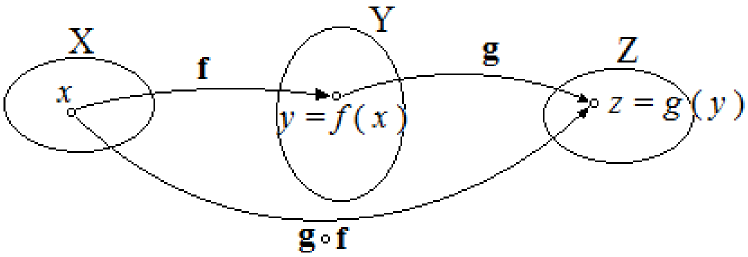
Множество X называется областью определения отображения f, а множество Y – областью значений. Множество упорядоченных пар

$$\Gamma_f = \{(x, y) : x \in X, y \in Y, y = f(x)\}$$

называют графиком отображения f. Непосредственно из определения вытекает, что график отображения f является подмножеством декартова произведения X x Y: $\Gamma_f \subset X \times Y$.

4.2. Композиция отображений и обратное отображение

Пусть имеются два отображения вида $f : X \to Y$ и $g : Y \to Z$. Выберем произвольный элемент $x \in X$ и применим к нему отображение f. Под действием f элемент x перейдет в элемент $y = f(x)$ множества Y. Если теперь к элементу y применить отображение g, то он перейдет в элемент $z = f(y)$ множества Z. В результате каждому $x \in X$ ставится в соответствие вполне определенный элемент $z = g(f(x))$ множества Z. Таким образом, последовательное применение отображений f и g приводит к отображению множества X в множество Z, которое называется произведением (или композицией) отображений g и f. Так как произведение отображений g и f переводит элемент $x \in X$ в элемент g(f(x)), то это произведение естественно обозначать символом g ∘ f или просто g f.



По определению $(g \circ f)(x) = g(f(x))$ для любого $x \in X$.

Отметим, что композиция отображений $f : X \to Y$ и $g : U \to Z$ определено лишь в случае, когда U=Y. Может оказаться, что $g \circ f$ определено, а $f \circ g$ не определено. Но если f и g — преобразования множества X, то определены оба произведения $g \circ f$, $f \circ g$ и они также являются преобразованиями X.

Пример. Пусть X={1,2,3}. Рассмотрим следующие два преобразования множества X:

$$f = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 1 \end{pmatrix} \text{ и } g = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}.$$

Тогда

$$(g \circ f)(1) = g(f(1)) = g(3) = 1,$$
$$(g \circ f)(2) = g(f(2)) = g(1) = 2,$$
$$(g \circ f)(3) = g(f(3)) = g(1) = 2.$$

Итак $g \circ f = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 2 \end{pmatrix}$. Аналогично получаем, что $f \circ g = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 1 & 3 \end{pmatrix}$. Таким образом, $g \circ f \neq f \circ g$ и, следовательно, композиция отображений, вообще говоря, зависит от порядка сомножителей.

Пример. Пусть $f(x) = \sqrt{x}$ и $g(x) = x + 5$ - функции, заданные на множестве действительных чисел. Имеем $f(g(x)) = f(x+5) = \sqrt{x+5}$, т.е. $(f \circ g)(x) = \sqrt{x+5}$ Аналогично $(g \circ f)(x) = g(f(x)) = g(\sqrt{x}) = \sqrt{x} + 5$.

Пример. Пусть f и g – преобразования множества **R**, соответствующие функциям $y = \sin x$ и $y = x^2$, т.е. $f(x) = \sin x$, $g(x) = x^2$ $\forall x \in \mathbf{R}$. Композиции $g \circ f$ и $f \circ g$ определены и $(g \circ f)(x) = g(f(x)) = g(\sin x) = (\sin x)^2$, $(f \circ g)(x) = f(g(x)) = f(x^2) = \sin(x^2)$. Очевидно $g \circ f \neq f \circ g$.

∎

Отметим, что операция композиции преобразований множества **R** называется в математическом анализе суперпозицией функций, так что $g \circ f$ — сложная функция, а под произведением функций понимается операция, являющаяся обычным умножение действительных чисел, возвращаемых функциями.

Теорема. Пусть f, g, h — отображения такие, что одно из произведений

$$(h \circ g) \circ f,\ h \circ (g \circ f)$$

определено. Тогда определено и другое произведение, причем

$$(h \circ g) \circ f = h \circ (g \circ f).$$

Т.о. композиция функций ассоциативна.

Доказательство. Пусть, например, определено $h \circ (g \circ f)$ (для какого – то элемента $x \in X$). Если $f : X \to Y$, то g должно быть отображением вида $g : Y \to Z$. Но тогда $g \circ f : X \to Z$ и, значит, $h : Z \to U$.

Проверим, что $(h \circ g) \circ f$ тоже определено. Действительно, $(h \circ g) : Y \to Z \to U = Y \to U$. Следовательно, определено и $(h \circ g) \circ f$ причем $(h \circ g) \circ f : X \to Y \to U = X \to U$.

Покажем теперь, что отображения $(h \circ g) \circ f$ и $h \circ (g \circ f)$ равны, т. е. для каждого $x \in X$

$$((h \circ g) \circ f)(x) = (h \circ (g \circ f))(x)$$

Имеем

$$((h \circ g) \circ f)(x) = (h \circ g)(f(x)) = h(g(f(x))),$$
$$(h \circ (g \circ f))(x) = h((g \circ f)(x)) = h(g(f(x))).$$

∎

Опираясь на понятие произведения двух отображений, можно определить композицию трех, четырех и более отображений.

Пусть, например, $f_1, f_2, \ldots, f_k$ – преобразования множества X. Их произведение определим индуктивно $f_k \circ (f_{k-1} \circ \ldots (f_3 \circ (f_2 \circ f_1)))$. Пользуясь ассоциативностью композиции отображений, нетрудно доказать, что справедливо равенство

$$f_k \circ \ldots \circ f_{i+1} \circ f_i \circ \ldots \circ f_1 = \left(f_k \circ \ldots \circ f_{i+1}\right) \circ \left(f_i \circ \ldots \circ f_1\right)$$

для любого $1 \le i < k$

Приведем еще несколько важных свойств композиции отображений.

Теорема. Пусть $f : X \to Y$, $g : Y \to Z$, $h = g \circ f$. Тогда:

1) если *f* и *g* инъективны, то и *h* инъективно;
2) если *f* и *g* сюръективны, то и *h* сюръективно;
3) если *f* и *g* биективны, то и *h* биективно.

Доказательство. Пусть f и g – инъекции. При любых $x_1, x_2 \in X$, $x_1 \ne x_2$ будет $f(x_1), f(x_2) \in Y$ и $f(x_1) \ne f(x_2)$ в силу инъективности *f*. Тогда $g(f(x_1)) \ne g(f(x_2))$ из – за инъективности g. Поскольку $h(x_1) = g(f(x_1))$ и $h(x_2) = g(f(x_2))$, то $h(x_1) \ne h(x_2)$, т.е. $h = g \circ f$ инъективно.

Пусть теперь f и g сюръективны. Возьмем произвольный элемент $z \in Z$. Поскольку $g : Y \to Z$ сюръективно, то существует хотя бы один элемент $y \in Y$, такой, что $g(y) = z$. Аналогично, из сюръективности $f : X \to Y$ следует существование такого элемента $x \in X$, что $f(x) = y$. Но тогда $h(x) = g(f(x)) = g(y) = z$ и произвольный элемент $z \in Z$ имеет прообраз в X при отображении *h*, т.е. $h(X) = Z$.

Очевидно, третье свойство есть прямое следствие первых двух.

∎

Теорема. Пусть $f : X \to Y$. Для тождественного отображения имеет место

$$f \circ I_X = f \text{ и } I_Y \circ f = f.$$

Доказательство. Проверим, например, справедливость первого равенства. Для любого $x \in X$

$$(f \circ I_X)(x) = f(I_X(x)) = f(x)$$

т. е. $f \circ I_X = f$. Второе равенство проверяется аналогично.

Следствие. В случае, когда f — преобразование множества X, то

$$f \circ I_X = I_X \circ f = f$$

Пусть $f : X \to Y$ — биективное отображение. Тогда для любого элемента $y \in Y$ существует единственный элемент $x \in X$, такой, что f(x)=y. Отображение $f^{-1} : Y \to X$, которое ставит в соответствие каждому $y \in Y$ его прообраз $x \in X$ при отображении f, называется **обратным** к f. Таким образом, если f переводит *x* в *y*, то f^{-1} переводит y в x. Инъективность и сюръективность отображения f^{-1} очевидны и, следовательно, для любого биективного отображения обратное к нему тоже биективно. При этом $\left(f^{-1}\right)^{-1} = f$, т. е. обратное отображение к f^{-1} совпадает с f.

Пример. Пусть X={1,2,3,4} и

$$f = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix}$$

есть подстановка множества X. Тогда

$$f^{-1} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 2 & 3 \end{pmatrix}$$

Действительно, так как f переводит 4 в 1, то f^{-1} переводит 1 в 4; f переводит 1 в 2 и, значит, f^{-1} переводит 2 в 1 и т. д.

Пример. Если f – поворот плоскости вокруг точки O на угол α, то f^{-1} – поворот плоскости вокруг той же точки на угол $-\alpha$.

Пример. Требуется найти обратную функцию для $y = 3x + 6$.

Функция – это частный случай отношения и ее можно задать указанием множества пар чисел (x, y) связанных заданным отношением, т.е.

$$f = \{(x, y) : y = 3x + 6\}.$$

Тогда

$$f^{-1} = \{(y, x) : y = 3x + 6\}.$$

Поменяв имена переменных, можем записать

$$f^{-1} = \{(x, y) : x = 3y + 6\}.$$

Решая уравнение относительно *y*, получаем

$$f^{-1} = \{(x, y) : y = (x - 6)/3\}.$$

∎

Очевидно, определены композиции $f^{-1} \circ f$ и $f \circ f^{-1}$, причем

$$f^{-1} \circ f = I_X, \quad f \circ f^{-1} = I_Y$$

Действительно, $f^{-1} \circ f : X \to X$ $\forall x \in X$ и $\left(f^{-1} \circ f\right)(x) = f^{-1}\left(f(x)\right) = x = I_X(x)$.

Аналогично проверяется второе из равенств.

Теорема. Если $f : X \to Y$ и $g : Y \to Z$ – биекции, то $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.

Доказательство. Поскольку $g \circ f : X \to Z$, то $(g \circ f)^{-1} : Z \to X$. Также $g^{-1} : Z \to Y$ и $f^{-1} : Y \to X$. Поэтому $f^{-1} \circ g^{-1} : Z \to Y \to X = Z \to X$.

Пусть $x \in X$. Тогда $f(x) = y$, $g(y) = z$ и, следовательно, $(g \circ f)(x) = g(f(x)) = z \in Z$. Также $\left(f^{-1} \circ g^{-1}\right)(z) = f^{-1}\left(g^{-1}(z)\right) = f^{-1}(y) = x$. Итак, если $g \circ f$ переводит *x* в *z*, то $f^{-1} \circ g^{-1}$ переводит *z* в *x*, т.е. $f^{-1} \circ g^{-1}$ – обратное отображение к $g \circ f$.

4.3 Принцип Дирихле.

Пусть $f : A \to B$ — функция, причем как A, так и B — конечные множества. Предположим, что A состоит из n элементов $a_1, a_2, \ldots, a_n$. Принцип Дирихле гласит, что если $|A| > |B|$, то по крайней мере одно значение f встретится более одного раза. Проще говоря, найдется пара элементов $a_i \neq a_j$, для которых $f(a_i) = f(a_j)$.

Допуская некоторую вольность, принцип Дирихле можно переформулировать в легко запоминающейся форме: нельзя рассадить 10 зайцев в 9 клеток так, чтобы в каждой клетке сидел один заяц.

Чтобы убедиться в истинности принципа, предположим, что для любой пары разных индексов $i \neq j$ мы имеем $f(a_i) \neq f(a_j)$. Тогда множество B содержит по крайней мере n различных элементов: $f(a_1), f(a_2), \ldots, f(a_n)$. И уж во всяком случае $|B| \geq n$, что противоречит предположению $n = |A| > |B|$. Следовательно, есть хотя бы два разных элемента $a_i, a_j \in A$, для которых $f(a_i) = f(a_j)$.

Пример 1. В группе 15 студентов. Покажем что, по крайней мере, у двоих из них день рождения в одном и том же месяце.

Решение. Множество студентов группы обозначим буквой A, а множество всех 12 месяцев обозначим через B. Рассмотрим функцию $f: A \to B$, сопоставляющую каждому студенту месяц его рождения. Так как $|A| = 15$, а $|B| = 12$, то $|A| > |B|$. По принципу Дирихле функция f должна иметь повторяющиеся значения, т. е. найдутся два студента с одним и тем же месяцем рождения.

∎

Для данной задачи трудно понять, зачем нам применять формальное рассуждение. Действительно. Дано 15 человек и 12 месяцев. Поэтому совершенно очевидно, что хотя бы двое из них родились в один и тот же месяц. Но в более сложных задачах применение принципа является полезным.

Пример 2. Какое наименьшее число фамилий должно быть записано в телефонном справочнике, чтобы с гарантией можно было утверждать, что хотя бы две фамилии начинаются с одной и той же буквы и заканчиваются одинаковыми буквами?

Решение. Пусть A — множество фамилий в справочнике, а B — множество пар букв, выписанных из стандартного алфавита, насчитывающего 33 буквы. Обозначим через $f: A \to B$ функцию, которая каждой фамилии справочника ставит в соответствие пару букв: первую и последнюю буквы фамилии. Например, $f(\text{Кузнецов}) \to (\text{к}, \text{в})$. Множество B содержит $33 \cdot 33 = 1089$ пар букв. Принцип Дирихле гарантирует нам, что если $|A| > |B| = 1089$, то найдется по крайней мере две фамилии, начинающиеся и оканчивающиеся на одинаковые буквы. Поэтому телефонный справочник должен содержать не менее 1090 фамилий.

Пример 3. Покажите, что какие бы пять цифр из 1, 2, 3, 4, 5, 6, 7 и 8 мы ни выбрали, найдутся хотя бы две из них, сумма которых равна 9.

Решение. Обозначим через A множество выбранных пяти цифр (не важно каких конкретно), а через B множество пар цифр , дающих в сумме 9 (в нем содержатся все восемь цифр):

B={{1,8},{2,7},{3,6},{4,5}}.

Рассмотрим функцию $f:A\to B$, сопоставляющую каждой цифре из пятерки пару из множества B, которая в ней содержится. Например, f(3)={3,6}. По принципу Дирихле хотя бы две цифры из множества A попадут в одну и ту же пару. Или, по – другому, две из пяти цифр дадут в сумме 9.

4.4 Специальные виды отображений и функций.

Пусть X — конечное множество, состоящее из n элементов. Эти элементы можно перенумеровать с помощью первых n натуральных чисел. Так как природа элементов множества X для нас не важна, то будем считать, что X={1,2,...,n}. Всякое преобразование f множества X будем записывать так:

$$f=\begin{pmatrix} 1 & 2 & \dots & n \\ f(1) & f(2) & \dots & f(n) \end{pmatrix} \tag{1}$$

Если f – **подстановка**, т. е. биективное преобразование множества в себя, то в строке f(1),f(2),...,f(n) выписаны все числа 1,2,...,n без повторений, только в другом порядке. Строки такого вида называются **перестановками** из n чисел. Таким образом, перестановка из n чисел – это расположение чисел 1,2,...,n в некотором определенном порядке. Две перестановки из n чисел различаются порядком элементов, но не элементами (как множества они совпадают). Например, (1, 2, 3, 4); (3, 1, 2, 4); (4, 2, 1, 3) − различные перестановки из четырех чисел.

Итак, если f − подстановка множества X, то нижняя строка (1) есть некоторая перестановке из n чисел. Обратно, если $x_1, x_2, \dots, x_n$ – произвольная перестановка из n чисел, то преобразование

$$f=\begin{pmatrix} 1 & 2 & \dots & n \\ x_1 & x_2 & \dots & x_n \end{pmatrix}$$

множества X является подстановкой. При этом различным перестановкам соответствуют различные подстановки.

Напомним, что нам известна одна специальная подстановка − тождественная функция I, определенная соотношением $I(x)=x$ для всех $x\in X$. Она может быть представлена в виде

$$f=\begin{pmatrix} 1 & 2 & \dots & n \\ 1 & 2 & \dots & n \end{pmatrix}.$$

Пусть $\mathbf{Z}_0=\mathbf{N}\bigcup\{0\}$ является множеством неотрицательных целых чисел (множество натуральных чисел **N** объединено с множеством, состоящим из одного числа – нуля). **Факториалом** называют функцию $f:\mathbf{Z}_0 \to \mathbf{N}$, обозначаемую через $f(n)=n!$ и определяемую следующими соотношениями:

```
0! = 1
1! = 1 =1·0!
2! = 1·2 =2=2·1!
3! = 1·2·3 =6 =3·2!
4! = 1·2·3·4 = 24 = 4-3!
..........
n! = 1·2·3·4·...·n = n·(n-1)!
..........
```

Теорема. Количество различных перестановок из n чисел равно n!

Доказательство проведем индукцией по n. При n=1 утверждение теоремы очевидно. Будем считать, что n>1 и число различных перестановок из n−1 чисел равно (n−1)!. Разобьем множество всех перестановок из n чисел на классы, состоящие из перестановок с одинаковым последним числом. Таких классов будет ровно n. Для фиксированного последнего элемента перестановок из первых n−1 элементов, по предположению, равно (n−1)! Но тогда число всех перестановок из n чисел равно $n\cdot(n-1)!=n!$.

∎

Следствие. Число всех подстановок множества X из n элементов равно n!

Пример. Выпишем все перестановки из трех чисел:

1, 2, 3; 1, 3, 2; 2, 1, 3; 2, 3, 1; 3, 1, 2; 3, 2, 1.

Так как число различных перестановок из трех чисел равно 3! = 6, то других перестановок нет.

∎

Конечной последовательностью называют функцию из множества {1,2,3,4,...,n} в некоторое множество S. Бесконечной последовательностью назовем функцию из множества натуральных чисел {1,2,3,4,...} в некоторое множество S. Любая конечная или бесконечная последовательность может быть названа просто последовательностью.

В большинстве случаев S представляет собой множество чисел. Последовательность обычно изображают списком или перечислением элементов. Если A – функция, то значение A(1) можно обозначить через A_1, значение A(2)– через A_2 и так далее. Например, последовательности 1,4,9,16,... соответствует запись $A_i=i^2$. Здесь A(1)=1,A(2)=4,A(3)=9, A(4)=16 и т.д.

Пример. Заданы первые пять членов последовательности: 0,3,8,15,24, ... Требуется представить ее как функцию. Легко видеть, что $A(n)=n^2-1$.

∎

Для суммирования членов последовательностей $x_r+x_{r+1}+x_{r+2}+\ldots+x_{r+k}$ используется специальное обозначение $\sum\limits_{i=r}^{r+k}x_i$. Например,

$$\sum_{i=1}^{5} i = 1+2+3+4+5;$$

$$\sum_{i=2}^{6} i^2 = 2^2+3^2+4^2+5^2+6^2;$$

$$\sum_{i=1}^{3}(i+2)^3 = (1+2)^3+(2+2)^3+(3+2)^3.$$

Два вида последовательностей – арифметическая прогрессия и геометрическая прогрессия – представляют особый интерес.

Каждый член **арифметической** прогрессии может быть получен из предыдущего прибавлением константы c:

$$a_1 = a$$

$$a_2 = a + c$$

$$a_3 = a_2 + c = a + 2\cdot c$$

. . . .

$$a_i = a_{i-1} + c = a + (i-1)\cdot c$$

. . . .

Таким образом, арифметическая прогрессия, как функция, записывается в виде $A(i) = a + (i-1)\cdot c$, где a – первый член прогрессии. Так последовательность 3, 5, 7, 9, 11, ... представляет собой арифметическую прогрессию с a=3 и c=2.

Выведем формулу для суммы первых n членов арифметической прогрессии $S_n = \sum_{i=1}^{n}(a+(i-1)\cdot c)$. Для этого запишем сумму первых n членов арифметической прогрессии в прямом и обратном порядке

$$S_n = a_1 + a_2 + \ldots + a_{n-1} + a_n$$

$$S_n = a_n + a_{n-1} + \ldots + a_2 + a_1,$$

и сложим

$$2S_n = (a_1+a_n)+(a_2+a_{n-1})+\ldots+(a_{n-1}+a_2)+(a_n+a_1) =$$

$$= \left(a + \underbrace{a+(n-1)\cdot c}_{a_n}\right) + \left(\underbrace{(a+c)}_{a_2} + \underbrace{a+(n-2)\cdot c}_{a_{n-1}}\right) + \ldots +$$

$$+ \left(\underbrace{a+(n-2)\cdot c}_{a_{n-1}} + \underbrace{(a+c)}_{a_2}\right) + \left(\underbrace{a+(n-1)\cdot c}_{a_n} + a\right) =$$

$$= (2a+(n-1)\cdot c)+(2a+(n-1)\cdot c)+\ldots+(2a+(n-1)\cdot c)+(2a+(n-1)\cdot c) =$$

$$= n\cdot(2a+(n-1)\cdot c)$$

Тогда

$$S_n = \frac{1}{2}n\cdot(2a+(n-1)\cdot c) = \frac{1}{2}n\cdot\left(a + \underbrace{a+(n-1)\cdot c}_{a_n}\right) = \frac{n\,(a_1+a_n)}{2}. \qquad (2)$$

∎

В **геометрической** прогрессии каждый член может быть получен из предыдущего путем умножения на константу q.

$a_1 = a$;

$a_2 = a \cdot q$;

$a_3 = a_2 \cdot q = a \cdot q^2$;

. . . .

$a_i = a_{i-1} \cdot q = a \cdot q^{i-1}$;

. . . .

Таким образом, геометрическая прогрессия, как функция, представляется в виде $A(i) = a \cdot q^{i-1}$, где a − первый член прогрессии. Например, последовательность 4, 12, 36, 108, 324, ... является геометрической прогрессией, в которой $a = 4$ и $q = 3$. Последовательность 32, 16, 8, 4, 2, ... − геометрическая прогрессия, где $a = 32$ и $q = 1/2$.

Формула суммы n первых членов геометрической прогрессии имеет вид

$$S_n = \sum_{i=1}^{n} a \cdot q^{i-1} = a\frac{1-q^n}{1-q} \quad (q \neq 1). \tag{3}$$

Доказательство проведем по индукции. При n=1 формула верна, так как $S_1 = a\frac{1-q^1}{1-q} = a$. Предположим, что формула верна при n=k, т.е. $S_k = a\frac{1-q^k}{1-q}$. Тогда

$$S_{k+1} = S_k + a_{k+1} = S_k + a \cdot q^k = a\frac{1-q^k}{1-q} + a \cdot q^k = a\frac{1-q^k + (1-q)q^k}{1-q} =$$

$$= a\frac{1-q^k + q^k - q^{k+1}}{1-q} = a\frac{1-q^{k+1}}{1-q}.$$

Как видим, если формула верна для n=k, то она верна и для n=k+1. Следовательно формула верна при каждом натуральном значении n.

∎

Приведем определения некоторых часто встречающихся функций.

Характеристические функции. Пусть X – некоторое множество, и A – его подмножество. Определим отображение χ_A множества X в двухэлементное множество $\{0,1\}$ следующим образом: $\chi_A(x)=1$, если $x \in A$, и $\chi_A(x)=0$, если $x \notin A$. Функция χ_A называется характеристической функцией подмножества A.

Имеется простая связь между операциями над подмножествами множества X и операциями над их характеристическими функциями:

$\chi_{A\cap B}(x) = \chi_A(x) \cdot \chi_B(x)$;

$\chi_{A\cup B}(x) = \chi_A(x) + \chi_B(x) - \chi_A(x) \cdot \chi_B(x)$;

$\chi_{A\setminus B}(x) = \chi_A(x) \cdot (1 - \chi_B(x))$;

Очевидно, $A \subset B$ тогда и только тогда, когда $\chi_A(x) \leq \chi_B(x)$ для всех x.

Функция $f:\mathbf{R}\to\mathbf{Z}$, где $\mathbf{R}$ – множество действительных чисел, а $\mathbf{Z}$ – множество целых чисел, называется **целой частью числа** и обозначается $f(x)=[\,x\,]$, если она каждому $x\in\mathbf{R}$ ставит в соответствие наибольшее целое число, меньшее или равное x (рис. 1*а*). Например, [2.99]=2, [4]=4, [–4]= –4, [–4.1]= –5.

Пример. Пусть A={-1,0,1,2} и функция $f:A\to\mathbf{Z}$ определяется формулой: $f(x)=\left[\dfrac{x^2+1}{3}\right]$. Тогда ее множество значений {0,1}. ∎

Через $\mathrm{Mod}(x,n)$ обозначается функция вычисление *остатка деления* числа x на число n (рис. 1*б*). Для нее справедливо соотношение $\mathrm{Mod}(x,n)=x-n\cdot\left[\dfrac{x}{n}\right]$.

Функция вычисления *абсолютного значения* числа x (рис. 1*в*) определяется как

$$|\,x\,|=\begin{cases} x,\ x\ge 0\\ -x,\ x<0\end{cases}.$$

Функция *знака числа x* (знаковая функция, рис. 1*г*) определяется как

$$\operatorname{sign} x=\begin{cases} 1,\ x>0\\ 0,\ x=0\\ -1,\ x<0\end{cases}.$$

Имеет место соотношение $x\cdot\operatorname{sign} x=|\,x\,|$.

Функция *Хевисайда* $\mathrm{H}(x)$(рис. 1*д*) определяется как

$$\mathrm{H}(x)=\begin{cases} 1,\ x\ge 0\\ 0,\ x<0\end{cases}. \tag{4}$$

При этом значение функции Хевисайда в нуле в разных источниках определяется по – разному. Мы будем придерживаться того, что $\mathrm{H}(0)=1$.

Приведенные функции являются кусочно – непрерывными, т.е. на разных «кусках» вещественной оси определяются разными непрерывными функциями. На следующем рисунке представлены графики этих функций.

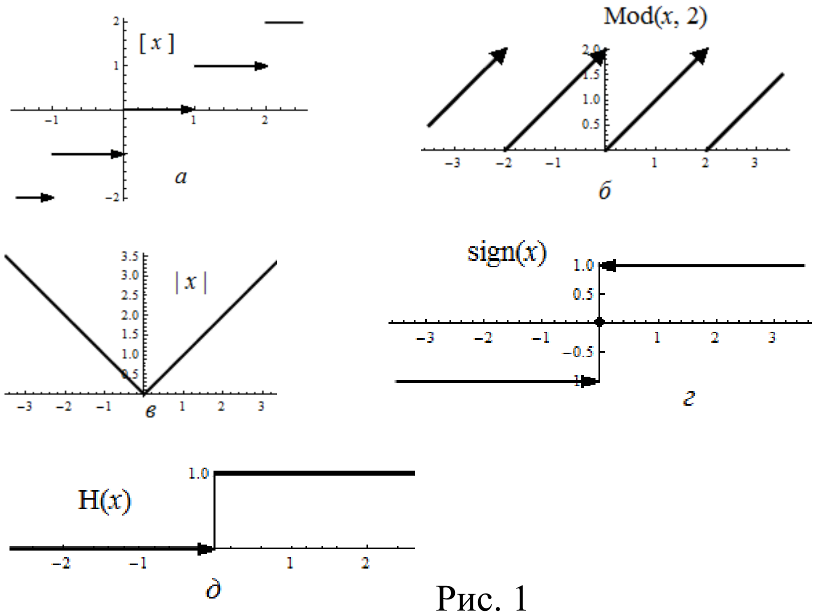


Рис. 1

Здесь стрелки на концах отрезков означают, что точки, на которые они указывают, не принадлежат графику функции.

Приведенные кусочные функции важны тем, что многие другие кусочно – непрерывные функции могут быть представлены с использованием этих базовых функций.

Теорема. Кусочно-постоянная функция

$$f(x)=\begin{cases} a_1, & x \le x_1 \\ a_2, & x_1 < x \le x_2 \\ \cdots & \\ a_{n-1}, & x_{n-2} < x \le x_{n-1} \\ a_n, & x > x_{n-1} \end{cases}, \tag{5}$$

эскиз графика которой показан на следующем рисунке,

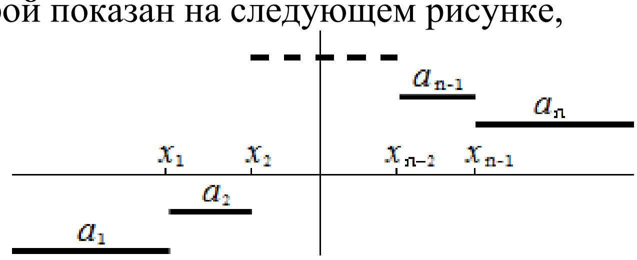


Рис. 2

может быть представлена в виде

$$f(x)=a_1+\sum_{i=1}^{n-1}(a_{i+1}-a_i)\mathrm{H}(x-x_i); \tag{6}$$

Доказательство. Для доказательства достаточно проверить, что правая часть выражения (6) на любом полуинтервале $x_{i-1} < x \le x_i$ совпадает с a_i.

При $x \le x_1$ имеем $\mathrm{H}(x-x_i)=0$ для любого $i=1,2,\ldots,n-1$ и, следовательно, правая часть (6) равна a_1.

При $x_{p-1} < x \le x_p$ $(2 \le p \le n-1)$ имеем $\mathrm{H}(x-x_i)=1$ для $i=1,2,\ldots,p-1$ и $\mathrm{H}(x-x_i)=0$ для $i=p,\ldots,n-1$. Тогда (6) дает

$$f(x)=a_1+(a_2-a_1)+\ldots+(a_p-a_{p-1})+0+\ldots+0=a_p.$$

При $x > x_{n-1}$ $\mathrm{H}(x-x_i)=1$ для всех $i=1,2,\ldots,n-1$ и (6) дает

$$f(x)=a_1+(a_2-a_1)+\ldots+(a_{n-1}-a_{n-2})+(a_n-a_{n-1})=a_n.$$

Таким образом, функция, заданная выражением (6), на всех участках совпадает с функцией, заданной выражением (5).

∎

В выражении (5) величины a_i можно рассматривать как некоторые аналитические функции $f_i(x)$, т.е. функции, заданные некоторыми аналитическими выражениями на своих участках $x_{i-1} < x \le x_i$. В этом случае справедлива следующая

Теорема. Кусочно – непрерывную функцию $f(x)$

$$f(x)=\begin{cases} f_1(x), & x\le x_1\\ f_2(x), & x_1<x\le x_2\\ \cdots & \\ f_{n-1}(x), & x_{n-2}<x\le x_{n-1}\\ f_n(x), & x>x_{n-1}\end{cases}, \tag{7}$$

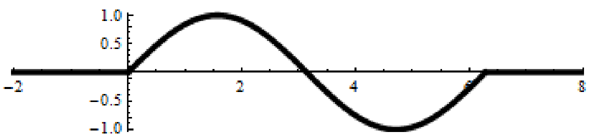
где все $f_i(x)$ аналитические функции, можно представить в виде

$$f(x)=f_1(x)+\sum_{i=1}^{n-1}\left(f_{i+1}(x)-f_i(x)\right)H(x-x_i). \tag{8}$$

Проверка формулы (8) выполняется дословным повторением доказательства предыдущей теоремы.

Пример. Дана функция

$$f(x)=\begin{cases} 0\;, x<0\\ \sin x,\, 0\le x\le 2\pi\\ 0\;, x>2\pi\end{cases}$$



В представлении (7) имеем $x_1=0, x_2=2\pi$ и $f_1(x)=0,\ f_2(x)=\sin x,\ f_3(x)=0$. В соответствии с (8), получаем

$$f(x)=f_1(x)+\left(f_2(x)-f_1(x)\right)\mathrm{H}(x-x_1)+\left(f_3(x)-f_2(x)\right)\mathrm{H}(x-x_2)=$$
$$=0+(\sin x-0)\mathrm{H}(x-0)+(0-\sin x)\mathrm{H}(x-2\pi)=\left(H(x)-H(x-2\pi)\right)\cdot\sin x.$$

∎

Уравнение Бернштейна ломаной (*непрерывной кусочно-линейной функции*).

Дан набор узлов $\{(x_k,y_k)\}_{k=0}^{n}$, $x_0<x_1<\ldots<x_n$, определяющих ломаную, проходящую через эти точки. При этом отрезок, соединяющий точки $(x_0,y_0),(x_1,y_1)$ продолжается влево и определяет для всех $x<x_1$ луч ломаной, проходящий через эти две точки. Отрезок, соединяющий пару точек $(x_{n-1},y_{n-1}),(x_n,y_n)$, определяет для всех $x>x_{n-1}$ другой луч ломаной (т.е. точки (x_0,y_0) и (x_n,y_n) не являются вершинами ломаной, а используются для задания ее крайних лучей).

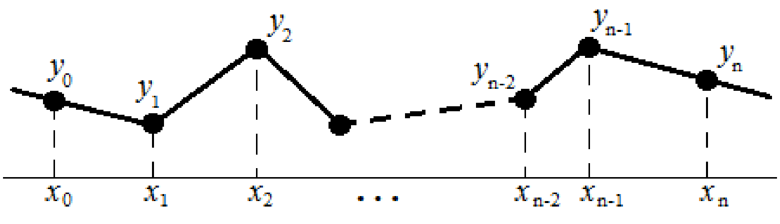


Рис. 3

Теорема (формула Бернштейна). Уравнение ломаной имеет вид

$$y=\frac{1}{2}\left(y_0+\frac{y_1-y_0}{x_1-x_0}(x-x_0)+y_{n-1}+\frac{y_n-y_{n-1}}{x_n-x_{n-1}}(x-x_{n-1})\right)+$$
$$+\frac{1}{2}\sum_{k=1}^{n-1}\left(\frac{y_{k+1}-y_k}{x_{k+1}-x_k}-\frac{y_k-y_{k-1}}{x_k-x_{k-1}}\right)\left|x-x_k\right|. \tag{9}$$

Д о к а з а т е л ь с т в о. Ломаная является непрерывной кривой, участки которой являются отрезками прямых линий. На рис. 4 приведено изображение ломаной в окрестности узла (x_k, y_k). Обозначим тангенс угла наклона $k-$го отрезка ломаной через $a_k = \dfrac{y_k - y_{k-1}}{x_k - x_{k-1}}$ и запишем уравнения прямых, на которых расположены примыкающие к $k-$му узлу отрезки ломаной.

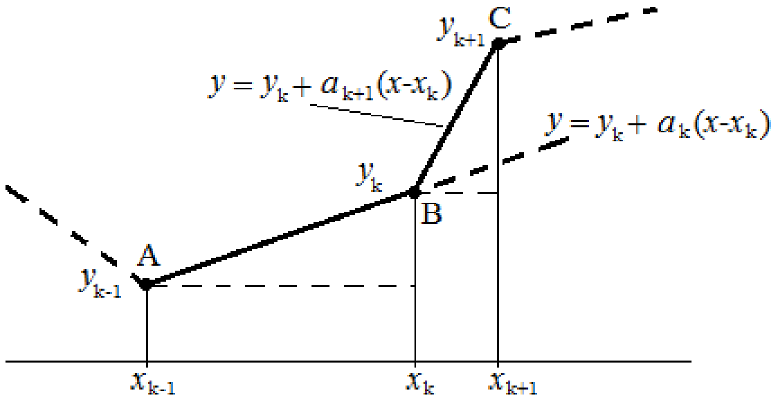


Рис. 4

Уравнение прямой АВ имеет вид $y = f_k(x) = y_k + a_k(x - x_k)$, а прямой ВС – $y = f_{k+1}(x) = y_k + a_{k+1}(x - x_k)$. Подставив в (8) уравнения прямых $f_k(x)$, получим

$$f(x) = y_1 + a_1(x - x_1) + \sum_{k=1}^{n-1}\big((y_k + a_{k+1}(x - x_k)) - (y_k + a_k(x - x_k))\big)\mathrm{H}(x - x_k) =$$

$$= y_1 + a_1(x - x_1) + \sum_{k=1}^{n-1}(a_{k+1} - a_k)(x - x_k)\mathrm{H}(x - x_k).$$

Но $(x - x_k)\cdot \mathrm{H}(x - x_k) = \dfrac{1}{2}\left(x - x_k + |\,x - x_k|\,\right)$, поскольку для $x \le x_k$ обе функции равны нулю, а для $x > x_k$ обе равны $x - x_k$. Тогда

$$f(x) = y_1 + a_1(x - x_1) + \frac{1}{2}\sum_{k=1}^{n-1}(a_{k+1} - a_k)\left(x - x_k + |\,x - x_k|\,\right) =$$

$$= y_1 + a_1(x - x_1) + \frac{1}{2}\sum_{k=1}^{n-1}(a_{k+1} - a_k)(x - x_k) + \frac{1}{2}\sum_{k=1}^{n-1}(a_{k+1} - a_k)|\,x - x_k|. \qquad (10)$$

Упростим первую сумму выражения (10). Имеем

$$\sum_{k=1}^{n-1}(a_{k+1} - a_k)(x - x_k) = a_2(x - x_1) - a_1(x - x_1) + a_3(x - x_2) - a_2(x - x_2) + \ldots +$$

$$+ a_{n-1}(x - x_{n-2}) - a_{n-2}(x - x_{n-2}) + a_n(x - x_{n-1}) - a_{n-1}(x - x_{n-1}) =$$

$$= -a_1(x - x_1) + a_2(x_2 - x_1) + a_3(x_3 - x_2) + \ldots + a_{n-1}(x_{n-1} - x_{n-2}) + a_n(x - x_{n-1}) =$$

$$= -a_1(x - x_1) + \frac{y_2 - y_1}{x_2 - x_1}(x_2 - x_1) + \ldots + \frac{y_{n-1} - y_{n-2}}{x_{n-1} - x_{n-2}}(x_{n-1} - x_{n-2}) + a_n(x - x_{n-1}) =$$

$$= -a_1(x - x_1) + (y_2 - y_1) + \ldots + (y_{n-1} - y_{n-2}) + a_n(x - x_{n-1})$$

$$= -a_1(x - x_1) - y_1 + y_{n-1} + a_n(x - x_{n-1}).$$

Тогда (10) принимает вид

$$f(x)=y_1+a_1(x-x_1)+\frac{1}{2}\left(-a_1(x-x_1)-y_1+y_{n-1}+a_n(x-x_{n-1})\right)+$$
$$+\frac{1}{2}\sum_{k=1}^{n-1}(a_{k+1}-a_k)|\,x-x_k\,|=$$
$$=\frac{1}{2}\left(y_1+a_1(x-x_1)+y_{n-1}+a_n(x-x_{n-1})\right)+\frac{1}{2}\sum_{k=1}^{n-1}(a_{k+1}-a_k)|\,x-x_k\,|. \qquad (11)$$

Точка (x_0, y_0) лежит на прямой $y=y_1+a_1(x-x_1)$, поэтому $y_0=y_1+a_1(x_0-x_1)$. Тогда

$$y_1+a_1(x-x_1)=y_0-a_1(x_0-x_1)+a_1(x-x_1)=y_0+a_1(x-x_0).$$

В результате (11) принимает вид

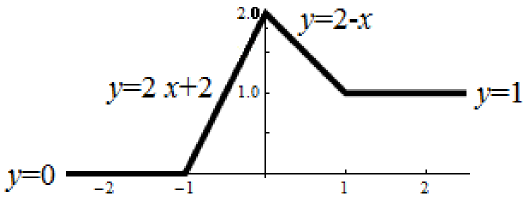
$$f(x)=\frac{1}{2}\left(y_0+a_1(x-x_0)+y_{n-1}+a_n(x-x_{n-1})\right)+\frac{1}{2}\sum_{k=1}^{n-1}(a_{k+1}-a_k)|\,x-x_k\,|. \qquad (12)$$

Заменяя в (12) все a_k их представлением $a_k=\dfrac{y_k-y_{k-1}}{x_k-x_{k-1}}$, приходим к (9).

∎

Формула (9) говорит, что для построения уравнения ломаной надо взять полусумму уравнений ее крайних лучей и прибавить полусумму произведений вида $(a_{k+1}-a_k)|\,x-x_k\,|$, где $(a_{k+1}-a_k)$ представляет разность тангенсов углов наклона отрезков, расположенных справа и слева от точки излома x_k.

Пример. Написать уравнение ломаной, график которой показан на следующем рисунке



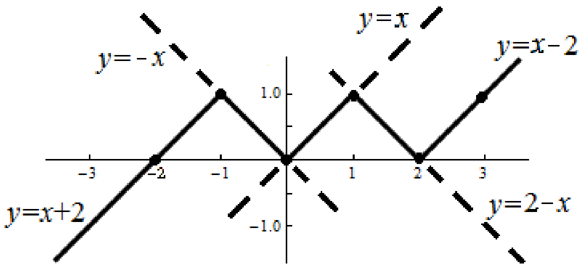
Р е ш е н и е. Уравнения отрезков прямых приведены на рисунке. Слагаемые в сумме (12) равны

$$(a_2-a_1)|\,x-x_1|=(2-0)|\,x-(-1)|=2|\,x+1\,|;$$
$$(a_3-a_2)|\,x-x_2|=(-1-2)|\,x-0\,|=-3|\,x\,|;$$
$$(a_4-a_3)|\,x-x_3|=(0-(-1))|\,x-1\,|=|\,x-1\,|.$$

Тогда в соответствии с (12) получаем

$$y=\frac{1}{2}(0+1)+\frac{1}{2}\left(2|\,x+1\,|-3|\,x\,|+|\,x-1\,|\right)=\frac{1}{2}+|\,x+1\,|-\frac{3}{2}|\,x\,|+\frac{1}{2}|\,x-1\,|.$$

Пример. Ломаная, проходит через точки $(-2,0),(-1,1),(0,0),(1,1),(2,0),(3,1)$. При этом первая и последняя точки в этом списке не являются узлами, а используются для определения крайних лучей. Требуется написать уравнение ломаной.



Решение. Из графика видно, что тангенсы наклона всех прямых равны ± 1. Уравнения крайних лучей ломаной имеют вид $y = x + 2$ и $y = x - 2$. Слагаемые из суммы (12) (суммирование ведется по всем точкам излома) равны:

$$(a_2 - a_1)|x - x_1| = ((-1) - 1)|x - (-1)| = -2|x + 1|;$$
$$(a_3 - a_2)|x - x_2| = (1 - (-1))|x - 0| = 2|x|;$$
$$(a_4 - a_3)|x - x_3| = ((-1) - 1)|x - 1| = -2|x - 1|;$$
$$(a_5 - a_4)|x - x_4| = (1 - (-1))|x - 2| = 2|x - 2|.$$

Тогда в соответствии с (12) (или (9)) получаем

$$y = \frac{1}{2}(x + 2 + x - 2) + \frac{1}{2}(-2|x + 1| + 2|x| - 2|x - 1| + 2|x - 2|) =$$
$$= x - |x + 1| + |x| - |x - 1| + |x - 2|.$$

∎

Теорема. Периодическое продолжение функции $\varphi(x)$, заданной на полуинтервале $[a, b)$, на всю вещественную ось может быть получено с помощью формулы

$$f(x) = \varphi(a + Mod(x - a, b - a)), \tag{13}$$

где $Mod(x, n)$ – функция вычисления остатка от деления x на n.

Доказательство. Действительно, при $x \in [a, b)$ $Mod(x - a, b - a) = x - a$ и

$$f(x) = \varphi(a + Mod(x - a, b - a)) = \varphi(a + (x - a)) = \varphi(x).$$

Проверим, что $f(x + T) = f(x)$, где $T = b - a$. Учитывая периодичность функции $Mod(x, T)$ с периодом Т, имеем

$$f(x + T) = \varphi(a + Mod((x + (b - a)) - a, b - a)) =$$
$$= \varphi(a + Mod(x - a, b - a)) = f(x).$$

Таким образом, при $x \in [a, b)$ функция $f(x)$ совпадает с $\varphi(x)$ и, кроме того, $f(x + (b - a)) = f(x)$. Но это и значит, что $f(x)$ является периодическим продолжением $\varphi(x)$ с полуинтервала $[a, b)$ на всю вещественную ось.

∎

Пример. Построить периодическое продолжение функции $\varphi(x) = x(1 - x)^3$ с отрезка $[0,1]$ на всю вещественную ось.

Решение. В соответствии с формулой (13) периодическое продолжение $\varphi(x)$ с полуинтервала $[0,1)$ имеет вид:

$$f(x) = \varphi(\mathrm{Mod}(x,1)) = \mathrm{Mod}(x,1)(1 - \mathrm{Mod}(x,1))^3.$$

В силу того, что $\varphi(0)=\varphi(1)=0$, результирующая функция $f(x)$ является непрерывной. На рис. 5*а* приведен график исходной функции, а на рис. 5*б* – ее периодическое продолжение.

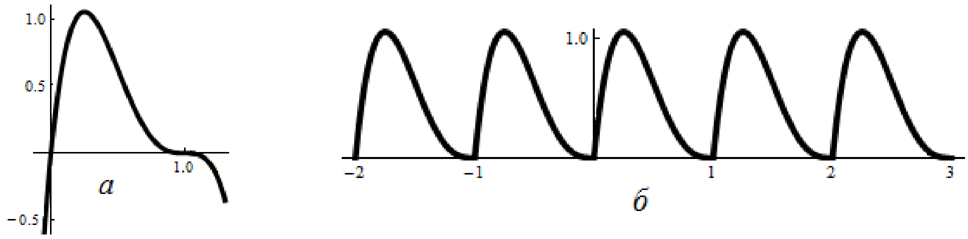


Рис. 5

Приведенные в этом параграфе функции $[x], \mathrm{Mod}(x,n), |x|, \mathrm{sign}(x), \mathrm{H}(x)$ особенно удобно использовать в системах символьной математики, но они имеются и среди библиотечных математических функций языков программирования, таких как C/C++, C#, Java, Visual Basic. Использование формул (6), (8), (9), (12), (13) позволяет создавать простые алгоритмы вычисления кусочных функций.

4.5 Дискретные функции

Значения физических величин, измеряемых приборами, обычно непрерывно меняются в зависимости от расстояния или от времени. В первом случае мы говорим о функции $f(x)$ пространственной координаты x. Во втором случае – о функциях времени $f(t)$, которые часто называют сигналами. При изучении функций природа измеряемой величины и аргумента не имеет значения.

Первоначально функция имеет непрерывно меняющийся аргумент и амплитуду. Преобразование в цифровую форму требует задания ее аргумента и значений некоторыми дискретными отсчетами. Представление аргумента конечным множеством отсчетов называется дискретизацией функции, а представление значений из конечного набора чисел называется квантованием. Если аргумент и значения функции f выбираются из фиксированных конечных наборов (дискретных величин), то говорят о цифровой функции. Для наших целей достаточно считать, что аргумент меняется дискретно, а значение может быть любым вещественным числом. В этом случае говорят, что задана дискретная функция или дискретная величина.

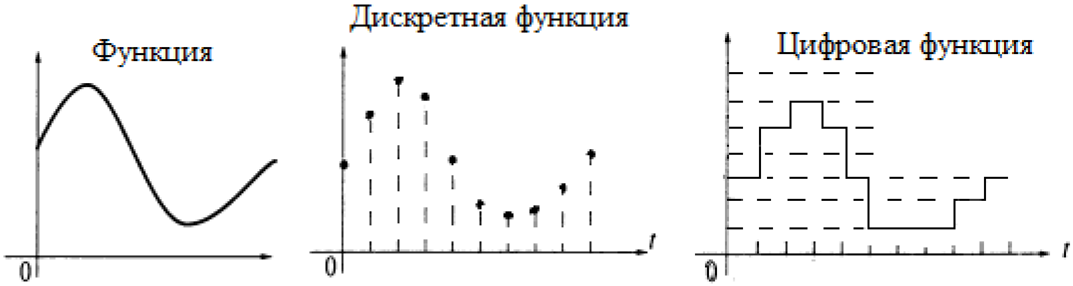


Рис. 1

Если интервал изменения аргумента непрерывной функции разделить на n одинаковых отрезков и значения функции в точках разбиения записать в виде ряда значений $f=(f_1, f_2, \ldots, f_n)$, то функцию можно представить n – мерным вектором (n – мерным вектором называется величина, представленная набором n числовых значений, расположенных в определенном порядке). Такие наборы чисел еще называют списками.

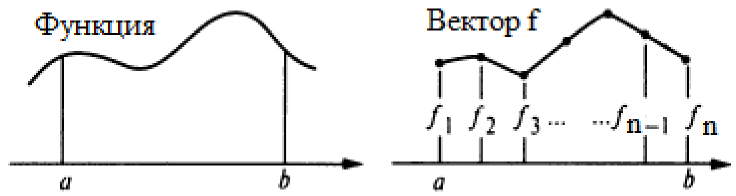


Рис. 2

Элемент f_i из этого числового набора называется компонентой дискретной функции (или координатой вектора). Иногда дискретную функцию удобно обозначать $\{f_i\}_1^n$, где снаружи фигурных скобок указываются пределы изменения индекса i.

В тех случаях, когда отсчеты функции могут продолжаться долго (не заканчиваются), тоже говорят о дискретной функции и понимают под этим бесконечную последовательность чисел $\{f_i\}_1^\infty$. Это значит, что есть способ получить значение компоненты дискретной функции для любого номера i. Так последовательность $\left\{\frac{1}{i^2}\right\}_1^\infty$ представляет дискретную функцию $\left(1, \frac{1}{4}, \frac{1}{9}, \ldots\right)$. А последовательность $(1, 3, -1, 5, 0, 0, 0, \ldots)$ фактически представляет дискретную функцию с конечным числом компонентов (все компоненты с номером, начиная с пятого, равны нулю). Наоборот, можно задать конечное количество компонент и продолжить последовательность нулями или другим способом. В результате мы получим бесконечную дискретную функцию, которую называют продолжением исходной.

Дискретной функцией общего вида называется отображение $f: \mathbf{H} \to \mathbf{R}$, где $\mathbf{H} \subseteq \mathbf{Z}$ – произвольное подмножество множества целых чисел $\mathbf{Z}$, а $\mathbf{R}$ – множество вещественных. В частности, может быть, что $\mathbf{H} = \mathbf{Z}$. Тогда мы будем иметь бесконечную в обе стороны последовательность $(\ldots, f_2, f_1, f_0, f_1, f_2, \ldots)$. В этом параграфе мы, в основном, будем иметь дело с дискретными функциями, имеющими конечное количество компонент. Их часто называют n – мерными векторами. Обращаем ваше внимание также на то, что в этом параграфе, если об этом не оговорено специально, мы будем полагать, что отсчеты дискретной функции берутся через одинаковые интервалы (времени или координаты x). Если это не так, то некоторые из формул этого параграфа должны быть изменены.

В геометрии и физике обычно имеют дело с двумерными или трехмерными векторами, которые изображаются на плоскости или в пространстве направленными отрезками. Такие отрезки обычно рисуют выходящими из начала координат (точка приложения) и идущими в точку с заданными координатами. Их часто называют геометрическими векторами в отличии от n – мерных векторов $(n > 3)$, которые изобразить отрезком невозможно.

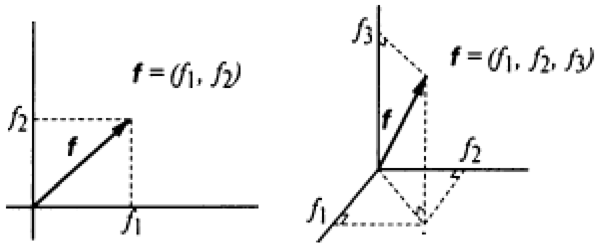


Рис. 3

Пусть заданы два геометрических вектора $\boldsymbol{a}$ и $\boldsymbol{b}$. Возьмем какую-либо точку О и отложим от нее вектор $\boldsymbol{a}$, т. е. построим такой отрезок ОА, что $\vec{OA} = \boldsymbol{a}$. Далее, от точки А отложим $\boldsymbol{b}$, т. е. построим такой отрезок АВ, что $\vec{AB} = \boldsymbol{b}$.

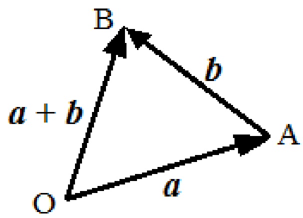


Рис. 4

Вектор, определяемый отрезком ОВ, называется суммой векторов $\boldsymbol{a}$ и $\boldsymbol{b}$ и обозначается $\boldsymbol{a}+\boldsymbol{b}$. Имея координаты векторов $\boldsymbol{a} = (a_1, a_2)$ и $\boldsymbol{b} = (b_1, b_2)$, координаты вектора $\boldsymbol{c} = (c_1, c_2)$ их суммы $\boldsymbol{c} = \boldsymbol{a}+\boldsymbol{b}$ вычисляются сложением соответствующих координат слагаемых: $\boldsymbol{c} = (a_1 + b_1, a_2 + b_2)$. Аналогично определяется сумма и разность двух n – мерных векторов $\boldsymbol{a} = (a_1, a_2, ..., a_n)$ и $\boldsymbol{b} = (b_1, b_2, ..., b_n)$:

$$\boldsymbol{c} = \boldsymbol{a} \pm \boldsymbol{b} = (a_1 \pm b_1, a_2 \pm b_2, ..., a_n \pm b_n). \quad (1)$$

Произведением геометрического вектора $\boldsymbol{a}$ на действительное число α называется вектор, который обозначается $\alpha\boldsymbol{a}$ и определяется следующими условиями: а) длина вектора $\alpha\boldsymbol{a}$ равна $|\alpha||\boldsymbol{a}|$, т. е. произведению модуля числа α и длины вектора $\boldsymbol{a}$; б) векторы $\boldsymbol{a}$ и $\alpha\boldsymbol{a}$ имеют одно и то же направление, если $\alpha > 0$, и противоположные направления, если $\alpha < 0$.

В двумерном случае координаты вектора $\alpha\boldsymbol{a}$ получаются умножением координат вектора $\boldsymbol{a} = (a_1, a_2)$ на α, т.е. $\alpha\boldsymbol{a} = (\alpha a_1, \alpha a_2)$. Аналогично определяется произведение n – мерного вектора $\boldsymbol{a} = (a_1, a_2, ..., a_n)$ на число α:

$$\alpha\boldsymbol{a} = (\alpha a_1, \alpha a_2, ..., a_n). \quad (2)$$

Для дискретных величин/функций, в отличии от векторов, также определяется понятие сложения/вычитания дискретной величины **f** с числом. Прибавление (или вычитание) числа b к дискретной функции означает прибавление (вычитание) b к каждой ее компоненте, т.е.

$$\mathbf{f} \pm b = \{f_1 \pm b,\ f_2 \pm b, ..., f_n \pm b\}. \quad (3)$$

Для двумерных и трехмерных векторов число, равное длине вектора $\boldsymbol{a}$, называется его модулем (или нормой) и обозначается $|\boldsymbol{a}|$. Применяя теорему Пифагора, можно вывести формулу вычисления модуля вектора через его координаты. Для двумерных $\boldsymbol{a}=(a_1,a_2)$ и трехмерных $\boldsymbol{a}=(a_1,a_2,a_3)$ векторов она имеет вид

$$|\boldsymbol{a}|=\sqrt{a_1^2+a_2^2} \quad \text{и} \quad |\boldsymbol{a}|=\sqrt{a_1^2+a_2^2+a_3^2}\,.$$

Аналогично, для n – мерного вектора (дискретной функции) $\boldsymbol{a}=(a_1,a_2,...,a_n)$ модуль определяется по формуле

$$|\boldsymbol{a}|=\sqrt{a_1^2+a_2^2+...+a_n^2}=\sqrt{\sum\nolimits_{i=1}^{n}a_i^2}\,. \tag{4}$$

Замечание. Когда n велико, то длина/норма вектора, определяемая формулой (4), зависит не только от координат a_i вектора $\boldsymbol{a}=(a_1,a_2,...,a_n)$, но и от их количества n. Чтобы избежать влияния размерности n иногда используют следующее определение модуля вектора: $|a|=\sqrt{\frac{1}{n}\sum\limits_{i=1}^{n}f_i^2}$.

Для того, чтобы оценить дискретную величину, наряду с нормой, используют и другие количественные характеристики. Важной характеристикой дискретной величины является ее среднее значение. Если отсчеты дискретной функции $\mathbf{f}=(f_1,f_2,...,f_n)$ взяты через одинаковые интервалы, то среднее значение вычисляется по формуле

$$\overline{f}=\frac{1}{n}\sum_{i=1}^{n}f_i\,. \tag{5}$$

Среднее значения $\overline{f}$ часто называют математическим ожиданием дискретной величины (функции) $\mathbf{f}$ и обозначают $M[\mathbf{f}]$.

Если интервалы между отсчетами неодинаковые, то для задания дискретной функции следует использовать два вектора – вектор значений аргумента $\mathbf{t}=\{t_1,t_2,...,t_n\}$ (например, моменты времени, в которые измеряются значения дискретной функции) и вектор самих значений $\mathbf{f}=\{f_1,f_2,...,f_n\}$, или таблицу

$$\begin{pmatrix} t_1 & t_2 & \dots & t_n \\ f_1 & f_2 & \dots & f_n \end{pmatrix}.$$

В этом случае среднее значение (математическое ожидание) дискретной величины вычисляется по формуле

$$M[\mathbf{f}]=\overline{f}=\frac{1}{\sum_{i=1}^{n}\Delta_i}\sum_{i=1}^{n}f_i\,\Delta_i=\frac{1}{t_n}\sum_{i=1}^{n}(t_i-t_{i-1})\cdot f_i\ , \tag{6}$$

где $\Delta_i=t_i-t_{i-1}$ и $t_0=0$.

Всюду в этом параграфе, если не оговорено противное, мы будем полагать, что все $\Delta_i=\Delta$ одинаковы.

Пример. Автомобиль движется из пункта A до пункта B первые 2 часа со скоростью 60 км/час, 1 час со скоростью 120 км/час и 3 часа со скоростью 50 км/час. Какова средняя скорость движения?

Ответ: `(60*2+120*1+50*3)/(2+1+3)=65.`

Другой количественной характеристикой, которую используют для дискретных величин является *дисперсия*. Она определяется формулой

$$D[\mathbf{f}\,]=M\left[(\mathbf{f}-M[\mathbf{f}])^2\right]=\overline{\left(\mathbf{f}-\overline{f}\right)^2}\,, \tag{7}$$

и является мерой разброса дискретной величины от ее среднего значения.

Полагая, что значения дискретной величины получены через одинаковые интервалы, формула для вычисления дисперсии принимает вид

$$D[\mathbf{f}\,]=\frac{1}{n}\sum_{i=1}^{n}\left(f_i-\overline{f}\right)^2 \tag{8}$$

Дисперсия определяет среднее значение квадрата отклонения дискретной величины **f** от ее среднего значения.

Квадратный корень из дисперсии называется *среднеквадратичным отклонением* $\sigma=\sqrt{D}$.

Легко видеть, что дисперсия любой дискретной величины неотрицательна $D[\mathbf{f}\,]\geq 0$. Если все компоненты дискретной функции одинаковы, то ее дисперсия равна нулю. Кроме того, для дисперсии справедливы следующие соотношения (проверьте самостоятельно):

- $D[a\mathbf{f}\,]=a^2 D[\mathbf{f}\,]$;
- $D[-\mathbf{f}\,]=D[\mathbf{f}\,]$;
- $D[\mathbf{f}+a\,]=D[\mathbf{f}\,]$,

где a – некоторое число.

В дискретном анализе часто приходится исследовать связи между дискретными величинами. В частности, мы можем выяснить насколько удалены векторы **`f`** и **`g`** друг от друга. В двумерном случае расстояние между векторами **`f`** и **`g`** определяется как модуль разности $|\mathbf{f}-\mathbf{g}|$. Следующий рисунок поясняет смысл этой операции

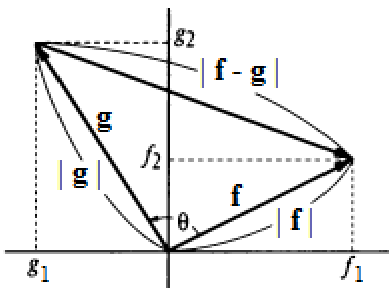


Рис. 5

Очевидно, что в двумерном случае

$$|\mathbf{f}-\mathbf{g}|=\sqrt{(f_1-g_1)^2+(f_2-g_2)^2}\,.$$

Аналогично, расстояние между `n` – мерными векторами определяется формулой

$$|\mathbf{f}-\mathbf{g}|=\sqrt{\sum_{i=1}^{n}(f_i-g_i)^2}\,. \tag{9}$$

Чем меньше значение $|\mathbf{f}-\mathbf{g}|$, тем ближе векторы $\mathbf{f}$ и $\mathbf{g}$, а значит сильнее между ними связь. При этом, если $|\mathbf{f}-\mathbf{g}|=0$, то вектора равны, поскольку из (9) сразу следует, что $f_i = g_i$ для всех i (вектора равны, если равны их координаты).

Расстояние – это один из параметров, измеряющих силу связи между векторами. Однако посмотрите на следующий рисунок

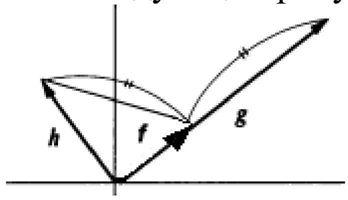


Рис. 6

На рисунке векторы **h** и **g** одинаково удалены от вектора **f**. Но векторы **f** и **g** имеют одно направление, а вектор **h** расположен под углом к вектору **f**. Увеличив вектор **f** в несколько раз, можно получить вектор **g** и невозможно получить вектор **h**. Несмотря на равноудаленность векторов **h** и **g** от **f**, связь **f** и **g** сильнее. Поэтому недостаточно выражать связь между векторами одним расстоянием. Необходимо учитывать также и угол.

Для выражения связи между двумя геометрическими векторами, образующими угол θ, используют скалярное произведение – число, которое обозначается как $\langle\mathbf{f},\mathbf{g}\rangle$ и определяется как $\langle\mathbf{f},\mathbf{g}\rangle = |\mathbf{f}|\cdot|\mathbf{g}|\cdot\cos\theta$. Тогда

$$\cos\theta = \frac{\langle\mathbf{f},\mathbf{g}\rangle}{|\mathbf{f}|\cdot|\mathbf{g}|}. \tag{10}$$

Здесь всегда $-1\le\cos\theta\le 1$. Величина $s=\cos\theta$ выражает силу связи между векторами **f** и **g** через угол θ. Если направление векторов совпадает, т.е. $\theta=0$, то s принимает максимальное значение, равное 1. С увеличением угла θ значение s уменьшается. Если $s=0$, т.е. $\langle\mathbf{f},\mathbf{g}\rangle=0$ и $\theta=\pi/2$, векторы **f** и **g** взаимно перпендикулярны. Таким образом, параметр s зависит от угла между векторами и не зависит от их длин.

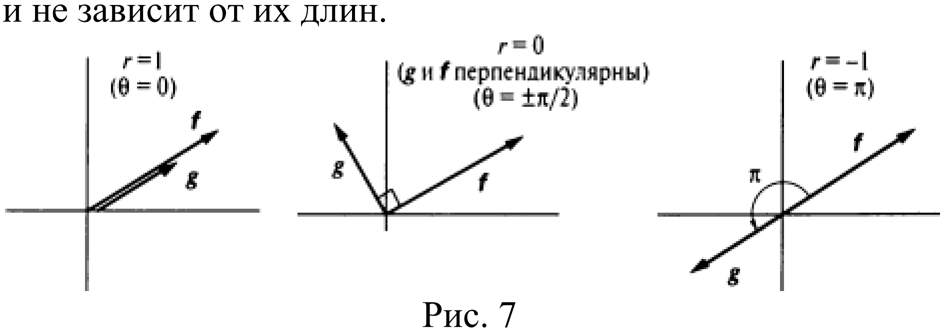


Рис. 7

Выразим скалярное произведение $\langle\mathbf{f},\mathbf{g}\rangle$ через координаты векторов $\mathbf{f}=(f_1, f_2)$ и $\mathbf{g}=(g_1,g_2)$. Для этого применим теорему косинусов для векторов (см. рис. 5). Вектора **f**, **g** и **f - g** представляют стороны треугольника. Тогда их длины связаны соотношением (теорема косинусов)

$$|\mathbf{f}-\mathbf{g}|^2 = |\mathbf{f}|^2+|\mathbf{g}|^2-2|\mathbf{f}||\mathbf{g}|\cos\theta = |\mathbf{f}|^2+|\mathbf{g}|^2-2\langle\mathbf{f},\mathbf{g}\rangle.$$

Следовательно,

$$2\langle \mathbf{f},\mathbf{g}\rangle = |\mathbf{f}|^2 + |\mathbf{g}|^2 - |\mathbf{f}-\mathbf{g}|^2 = \left(f_1^2 + f_2^2\right) + \left(g_1^2 + g_2^2\right) - \left((f_1 - g_1)^2 + (f_2 - g_2)^2\right) =$$
$$= 2\left(f_1 \cdot g_1 + f_2 \cdot g_2\right).$$

Таким образом, для двумерных векторов скалярное произведение можно вычислить по формуле

$$\langle \mathbf{f},\mathbf{g}\rangle = f_1 \cdot g_1 + f_2 \cdot g_2.$$

По – аналогии, для n – мерных векторов $\boldsymbol{f} = \left(f_1, f_2, \dots, f_n\right)$ и $\boldsymbol{g} = (g_1, g_2, \dots, g_n)$ скалярное произведение определяют формулой

$$\langle \mathbf{f},\mathbf{g}\rangle = \sum_{i=1}^{n} f_i \cdot g_i . \tag{11}$$

Скалярное произведение не меняет своего значения при перестановке сомножителей. Также ясно, что скалярное произведение вектора самого на себя равно квадрату его длины $\langle \mathbf{f},\mathbf{f}\rangle = \sum_{i=1}^{n} f_i^2 = |\mathbf{f}|^2$.

Для геометрических векторов определяется условие перпендикулярности, которое означает, что угол θ между ними равен $\frac{\pi}{2}$. Подставляя $\theta = \frac{\pi}{2}$ в (10), получаем, что $\langle \mathbf{f},\mathbf{g}\rangle = 0$. Наоборот, если $\langle \mathbf{f},\mathbf{g}\rangle = 0$, то геометрические вектора ортогональны (перпендикулярны). Теперь можно определить перпендикулярность для n – мерных векторов (дискретных функций), используя скалярное произведение: дискретные функции с n компонентами ортогональны, если $\langle \mathbf{f},\mathbf{g}\rangle = 0$, т.е., если $\sum_{i=1}^{n} f_i\, g_i = 0$.

Пример. Вектора (-2,-1,0,1,2) и (1,1,1,1,1) – ортогональны, т.к. $-2\cdot 1 + (-1\cdot 1) + (0\cdot 1) + (1\cdot 1) + (2\cdot 1) = 0$. Вектора (-1,-1,0,1,1) и (2,1,7,1,2) – ортогональны, т.к. $-1\cdot 2 + (-1\cdot 1) + (0\cdot 7) + (1\cdot 1) + (1\cdot 2) = 0$. Вектора (0,0,0,1,0) и (1,0,0,0,0) – ортогональны, т.к. $0\cdot 1 + (0\cdot 0) + (0\cdot 0) + (1\cdot 0) + (0\cdot 0) = 0$. Вообще, если вектора содержат только по одной ненулевой компоненте, стоящей на разных местах, то они ортогональны.

∎

Теперь формулу (10) для n – мерного случая можно записать следующим образом

$$s = \frac{\langle \mathbf{f},\mathbf{g}\rangle}{|\mathbf{f}|\cdot|\mathbf{g}|} = \frac{\sum_{i=1}^{n} f_i\, g_i}{\sqrt{\sum_{i=1}^{n} f_i^2}\cdot\sqrt{\sum_{i=1}^{n} g_i^2}}, \tag{12}$$

где $\cos\theta$ мы заменили значком s, который характеризует степень «схожести» дискретных функций. Однако у формулы (12) есть существенный недостаток – коэффициент s зависит от единиц измерения физической величины, которую представляют дискретные функции. Например, если функции **f** и **g** представляют температуру, которую можно измерять в градусах Цельсия и Фаренгейта, то получатся разные значения. Для примера предположим, что функции $\mathbf{f} = (2, 5, -3)$ и $\mathbf{g} = (7, -9, 1)$ представляют значения температуры по шкале Фаренгейта. Тогда $s_{f\,g}^{Fahr} = -0.481894$. Формула пересчета от градусов

Фаренгейта в градусы Цельсия имеет вид $c^{\circ} = \left(f^{\circ} - 32\right) \cdot \frac{5}{9}$. Представив те же дискретные функции, но уже в градусах Цельсия, получим $s_{f\,g}^{Cels} = 0.9634765$. Чтобы избавится от указанного недостатка при определении «схожести» дискретных функций, в формуле (12) значения каждой компоненты корректируется на среднее значение соответствующей функции, а именно

$$r(\mathbf{f},\mathbf{g}) = \frac{\sum_{i=1}^{n}\left(f_i - \overline{f}\right)\left(g_i - \overline{g}\right)}{\sqrt{\sum_{i=1}^{n}\left(f_i - \overline{f}\right)^2} \cdot \sqrt{\sum_{i=1}^{n}\left(g_i - \overline{g}\right)^2}}, \qquad (13)$$

где $\overline{f} = \frac{1}{n}\sum_{i=1}^{n} f_i$ и $\overline{g} = \frac{1}{n}\sum_{i=1}^{n} g_i$ представляют средние значения соответствующих функций, определяемые формулой (5).

Параметр r называют *коэффициентом корреляции* Пирсона дискретных функций **f** и **g**, или просто коэффициентом корреляции (Карл Пирсон, английский математик, 1857-1936). Коэффициент корреляции r безразмерная величина и не зависит от единиц, в которых измерялись значения функций. Он не изменится, если ко всем компонентам дискретной функции добавить одно и то же число, или, если все компоненты умножить на одно и то же число, отличное от нуля, т.е.

$$r(a \cdot \mathbf{f} + b, \mathbf{g}) = r(\mathbf{f}, \mathbf{g}),$$

где $a \neq 0$ и b некоторые числа.

Коэффициент корреляции не зависит от порядка следования функций **f** и **g**, т.е. $r(\mathbf{f},\mathbf{g}) = r(\mathbf{g},\mathbf{f})$. Кроме того, $r(\mathbf{f},\mathbf{f}) = 1$.

Смысл коэффициента $r(\mathbf{f},\mathbf{g})$ в формуле (13) такой же, как и $\cos\theta$ в формуле (10). Коэффициент корреляции r характеризует степень «схожести» векторов (дискретных функций). Причем r принимает значения от –1 до +1. Чем больше значение r по абсолютной величине, тем выше корреляция между функциями. Если функций **f** и **g** связаны линейной зависимостью, т.е. $\mathbf{g} = a\mathbf{f} + b$, где a и b числа, то $r(\mathbf{f}, a\mathbf{f} + b) = \pm \operatorname{sign} a$.

Покажем, что коэффициент корреляции, определяемый формулой (13), принимает значения от –1 до +1. Рассмотрим две дискретные функции **f** и **g** и их скалярное произведение $\langle \lambda\mathbf{f} + \mathbf{g}, \lambda\mathbf{f} + \mathbf{g} \rangle = |\lambda\mathbf{f} + \mathbf{g}|^2 \geq 0$. Тогда

$$\lambda^2\langle \mathbf{f},\mathbf{f}\rangle + 2\lambda\langle \mathbf{f},\mathbf{g}\rangle + \langle \mathbf{g},\mathbf{g}\rangle \geq 0.$$

Значит дискриминант многочлена $\lambda^2\langle \mathbf{f},\mathbf{f}\rangle + 2\lambda\langle \mathbf{f},\mathbf{g}\rangle + \langle \mathbf{g},\mathbf{g}\rangle$ неположительный, т.е. $D = 4\left(\langle \mathbf{f},\mathbf{g}\rangle\right)^2 - 4\langle \mathbf{f},\mathbf{f}\rangle\langle \mathbf{g},\mathbf{g}\rangle \leq 0$. Тогда

$$\left|\langle \mathbf{f},\mathbf{g}\rangle\right| \leq |\mathbf{f}| \cdot |\mathbf{g}| \text{ и } \frac{\left|\langle \mathbf{f},\mathbf{g}\rangle\right|}{|\mathbf{f}| \cdot |\mathbf{g}|} \leq 1.$$

Вычитание чисел из дискретных функций $\mathbf{F} = \mathbf{f} - \overline{f}$ и $\mathbf{G} = \mathbf{g} - \overline{g}$ создает новые дискретные функции **F** и **G**, для которых также выполняется последнее неравенство и, следовательно,

$$r=\frac{\left|\left\langle \mathbf{f}-\overline{f},\mathbf{g}-\overline{g}\right\rangle\right|}{\left|\mathbf{f}-\overline{f}\right|\cdot\left|\mathbf{g}-\overline{g}\right|}\le 1.$$

Пример 1. Даны дискретные функции **f**, **h** и **g**, которые представляют среднемесячные температуры в Харькове, Барселоне и Буэнос-Айресе.

	Янв.	Фев.	Март	Апр.	Май	Июнь	Июль	Авг.	Сен.	Окт.	Нояб.	Дек.	Год
Харьков	−4,6	−4,5	0,7	9,2	15,6	19,3	21,3	20,3	14,4	7,9	0,9	−3,5	8,1
Барселона	8,9	10,0	11,3	13,1	16,3	20,0	23,1	23,7	21,1	17,1	12,6	10,0	15,6
Буэнос Айрес	24,8	23,6	22,0	18,2	14,8	11,9	11,1	12,8	14,6	17,8	20,4	23,2	17,9

Последний столбец таблицы содержит среднегодовую температуру, т.е. среднее значение соответствующих $12-$ти компонентных дискретных функций. Определим, насколько похожи температуры в этих городах. Вначале построим графики дискретных функций

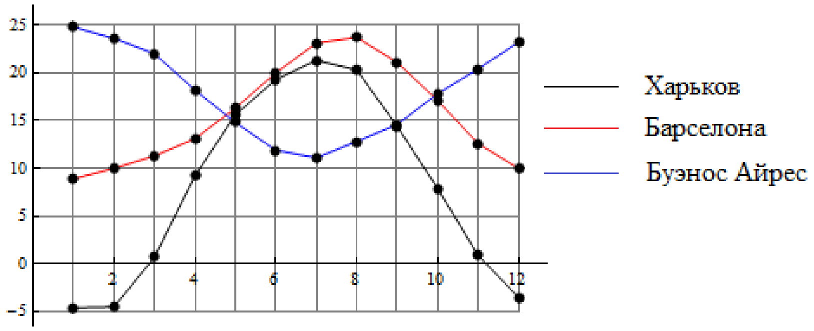


Рис. 8

Коэффициент корреляции между температурой Харькова и Барселоны равен 0.94, что говорит о высокой «схожести» соответствующих дискретных функций (в городах, расположенных в северном полушарии, максимум температуры достигается летом, а минимум – зимой). Коэффициент корреляции между температурой Харькова и Буэнос–Айреса равен –0.99 (почти минус единица). Это говорит о том, что между дискретными функциями имеется линейная связь $\mathbf{g}=a\mathbf{f}+b$, где $a<0$.

∎

Кроме коэффициента корреляции для дискретных функций **f** и **g** часто вычисляют коэффициент ковариации или просто *ковариацию*. Она определяется следующим образом:

$$\operatorname{cov}(\mathbf{f},\mathbf{g})=\frac{1}{n}\sum_{i=1}^{n}\left(f_i-\overline{f}\right)\left(g_i-\overline{g}\right), \qquad (14)$$

где $\overline{f},\overline{g}$ – средние значения соответствующих дискретных функций. Ковариация, с точностью до множителя $1/n$ представляет числитель формулы (13) и, естественно, тесно связана с коэффициентом корреляции. Раскрывая скобки в правой части (14), получаем

$$\operatorname{cov}(\mathbf{f},\mathbf{g})=\frac{1}{n}\sum_{i=1}^{n}\left(f_i-\overline{f}\right)\left(g_i-\overline{g}\right)=\frac{1}{n}\left(\sum_{i=1}^{n}f_i\,g_i-\overline{f}\sum_{i=1}^{n}g_i-\overline{g}\sum_{i=1}^{n}f_i+n\,\overline{f}\,\overline{g}\right)=$$

$$=\frac{1}{n}\left(\sum_{i=1}^{n}f_i\,g_i-\overline{f}\left(n\,\overline{g}\right)-\overline{g}\left(n\,\overline{f}\right)+n\,\overline{f}\,\overline{g}\right)=\frac{1}{n}\left(\sum_{i=1}^{n}f_i\,g_i-n\,\overline{f}\,\overline{g}\right)=$$

$$= \frac{1}{n}\sum_{i=1}^{n} f_i\, g_i - \left(\frac{1}{n}\sum_{i=1}^{n} f_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} g_i\right) = \frac{1}{n}\langle \mathbf{f}, \mathbf{g}\rangle - \overline{f}\cdot\overline{g}\,.$$

Очевидно, что $\mathrm{cov}(\mathbf{f}, \mathbf{g}) = \mathrm{cov}(\mathbf{g}, \mathbf{f})$. Ковариация в отличие от коэффициента корреляции не инвариантна относительно единиц измерения физической величины, что не всегда удобно в приложениях.

Заметим, что ковариация дискретной функции с собой равна ее дисперсии. Действительно, из (8) и (14) следует, что

$$\mathrm{cov}(\mathbf{f}, \mathbf{f}) = \frac{1}{n}\sum_{i=1}^{n}\left(f_i - \overline{f}\right)^2 = D[\mathbf{f}\,].$$

Лемма. Дисперсия суммы двух дискретных величин равна

$$D[\mathbf{f} + \mathbf{g}\,] = D[\mathbf{f}\,] + D[\mathbf{g}\,] + 2\,\mathrm{cov}(\mathbf{f}, \mathbf{g}\,). \tag{15}$$

Доказательство. Поскольку $\overline{\mathbf{f} + \mathbf{g}} = \overline{f} + \overline{g}$, то

$$D[\mathbf{f} + \mathbf{g}\,] = \frac{1}{n}\sum_{i=1}^{n}\left((f_i + g_i) - (\overline{f} + \overline{g})\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left((f_i - \overline{f}) + (g_i - \overline{g})\right)^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(f_i - \overline{f}\right)^2 + \frac{1}{n}\sum_{i=1}^{n}\left(g_i - \overline{g}\right)^2 + \frac{2}{n}\sum_{i=1}^{n}\left(f_i - \overline{f}\right)\left(g_i - \overline{g}\right) = D[\mathbf{f}\,] + D[\mathbf{g}\,] + 2\,\mathrm{cov}(\mathbf{f}, \mathbf{g}\,).$$

∎

Иногда для исследования связей между дискретными величинами числовых коэффициентов недостаточно. Тогда для этого можно использовать дискретные функции, которые в той или иной степени характеризует такую связи. При этом часто приходится использовать компоненты дискретных функций, индексы которых выходят за границы отрезка [1, n]. В такой ситуации говорят о продолжении дискретной функции, которое можно выполнить различными способами.

Рассмотрим дискретную функцию **f**, заданную первой строкой таблицы из примера 1 (она представляет значения среднемесячных температур для Харькова). На рис. 9*а* на отрезке [1,n] (здесь n=12) показан график функции **f** (красный цвет), а левее и правее синим показаны ее повторяющиеся значения. Полученная функция **g**, индексы компонент которой могут принимать любые целые значения, совпадает с **f** при $i = 1,2,\ldots,n$ (т.е. $g_i = f_i$), и для всех значений индексов i удовлетворяет соотношению $g_i = f_{1+\mathrm{mod}(i-1,n)}$. Здесь $\mathrm{mod}(p,q)$ обозначает функцию вычисления остатка от деления целого числа p на целое число q. Дискретная функция **g** (она теперь может содержать бесконечное количество компонент) называется периодическим продолжением дискретной функции **f** (рис. 9*а*). Вообще, последовательность (дискретная функция) $\mathbf{g} = \{\, g_i \,\}_{i=-\infty}^{\infty}$ называется периодической, если для некоторого n выполняется соотношение $g_{i+n} = g_i$ для любого *i*. Число n называется периодом функции.

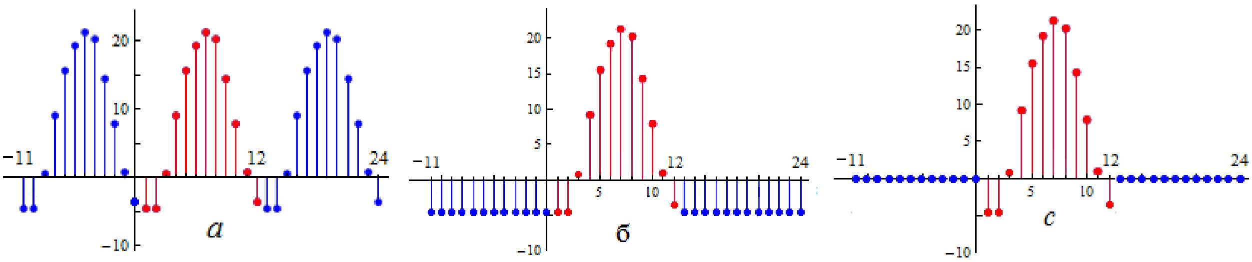


Рис. 9. Способы продолжения дискретных функций.

Другой способ продолжения состоит в повторении крайних значений дискретной функции влево и вправо (рис. 9б), или продолжении функции нулями (рис. 9*c*).

Обратившись еще раз к графику на рис. 8, можем заметить, что если не учитывать среднегодовую температуру и сдвижку на 6 месяцев, то поведение функций **f** и **g** (дискретных функций, представляющих среднемесячные температуры в Харькове и Буэнос – Айресе) довольно похожи. Формула (13) для коэффициента корреляции уже убирает зависимость от среднегодовой температуры. А как определить величину сдвига? Для этого используют *функцию взаимной корреляции* R. Ее удобно определять для периодически продолженных дискретных функций (в примере 1 мы имели именно такую дискретную функцию, поскольку ее значения повторяются через каждые 12 месяцев) следующим образом:

$$R_k = \frac{\sum_{i=1}^{n}\left(f_i - \overline{f}\right)\cdot\left(g_{i+k-1} - \overline{g}\right)}{\sqrt{\sum_{i=1}^{n}\left(f_i - \overline{f}\right)^2}\cdot\sqrt{\sum_{i=1}^{n}\left(g_i - \overline{g}\right)^2}} \qquad (k = 1, 2, \ldots, n). \qquad (16)$$

при этом в случае, когда индекс $i+(k-1)$ становится больше n мы, учитывая периодичность, должны использовать значения с левого конца вектора **g**, т.е. заменять индекс $i+(k-1)$ на $i+(k-1)-n$.

Если использовать функцию mod(·) вычисления остатка от деления и через $\mathbf{g_k} = \left\{g_{1+\mathrm{mod}(i+k-2,3)}\right\}_{i=1}^{n}$ обозначить дискретные функции, получаемые из **g** при фиксированных значениях $k = 1, 2, \ldots, n$, то последнюю формулу можно записать в виде

$$R_k = \frac{\left\langle \mathbf{f} - \overline{f}, \mathbf{g}_k - \overline{g}\right\rangle}{\left|\mathbf{f} - \overline{f}\right|\cdot\left|\mathbf{g} - \overline{g}\right|} = \frac{\sum_{i=1}^{n}\left(f_i - \overline{f}\right)\cdot\left(g_{1+\mathrm{mod}(i+k-2,n)} - \overline{g}\right)}{\sqrt{\sum_{i=1}^{n}\left(f_i - \overline{f}\right)^2}\cdot\sqrt{\sum_{i=1}^{n}\left(g_i - \overline{g}\right)^2}} \qquad (k = 1, 2, \ldots, n), \quad (17)$$

и n – это количество компонент функций **f** и **g** на отрезке одного периода.

Заметим также, что вместо выражения $\left|\mathbf{g} - \overline{g}\right|$ в знаменателе формул (16) и (17) должно было бы стоять выражение $\left|\mathbf{g}_k - \overline{\mathbf{g}_k}\right|$. Но, в силу периодического продолжения функции **g**, они равны и, в частности, равны средние значения функций $\mathbf{g_k}$ и **g**, т.е. $\overline{\mathbf{g_k}} = \overline{\mathbf{g}}$.

Формула (16) (или (17)) выражает величину связи между функциями **f** и **g** при сдвиге их друг относительно друга на k интервалов.

Пример. Вычислим функцию взаимной корреляции дискретных величин $\mathbf{f}=(1,-1,3)$ и **g**=(2,4,3), используя периодическое продолжение.

Для средних значений этих функций имеем $\overline{f}=(1-1+3)/3=1$, $\overline{g}=(2+4+3)/3=3$. Тогда

$$\mathbf{f}-\overline{f}=(1,-1,3)-1=(0,-2,2),\ \mathbf{g}-\overline{g}=(2,4,3)-3=(-1,1,0)$$

и

$$\left|\mathbf{f}-\overline{f}\right|=\sqrt{0^2+(-2)^2+2^2}=\sqrt{8}=2\sqrt{2},\ \left|\mathbf{g}-\overline{g}\right|=\sqrt{(-1)^2+1^2+0^2}=\sqrt{2}.$$

Пусть $\mathbf{g_k}=\left\{g_{1+\mathrm{mod}(i+k-2,3)}\right\}_{i=1}^{3}$ дискретные функции, полученные из **g** при фиксированном k, т.е. $\mathbf{g}_1=(g_1,g_2,g_3)=(2,4,3)$, $\mathbf{g}_2=(g_2,g_3,g_1)=(4,3,2)$, $\mathbf{g}_3=(g_3,g_1,g_2)=(3,2,4)$. Тогда

$$\mathbf{g}_1-\overline{g}=(2,4,3)-3=(-1,1,0);$$
$$\mathbf{g}_2-\overline{g}=(4,3,2)-3=(1,0,-1);$$
$$\mathbf{g}_3-\overline{g}=(3,2,4)-3=(0,-1,1);$$

и

$$R_1=\frac{\left\langle \mathbf{f}-\overline{f},\mathbf{g_1}-\overline{g}\right\rangle}{\left|\mathbf{f}-\overline{f}\right|\cdot\left|\mathbf{g}-\overline{g}\right|}=\frac{0\cdot(-1)+(-2)\cdot1+2\cdot0}{2\sqrt{2}\cdot\sqrt{2}}=-\frac{1}{2};$$

$$R_2=\frac{\left\langle \mathbf{f}-\overline{f},\mathbf{g}_2-\overline{g}\right\rangle}{\left|\mathbf{f}-\overline{f}\right|\cdot\left|\mathbf{g}-\overline{g}\right|}=\frac{0\cdot1+(-2)\cdot0+2\cdot(-1)}{2\sqrt{2}\cdot\sqrt{2}}=-\frac{1}{2};$$

$$R_3=\frac{\left\langle \mathbf{f}-\overline{f},\mathbf{g}_3-\overline{g}\right\rangle}{\left|\mathbf{f}-\overline{f}\right|\cdot\left|\mathbf{g}-\overline{g}\right|}=\frac{0\cdot0+(-2)\cdot(-1)+2\cdot1}{2\sqrt{2}\cdot\sqrt{2}}=1.$$

Т.о., функция взаимной корреляции равна $\mathbf{R}=\left(-\frac{1}{2},-\frac{1}{2},1\right)$. Отсюда видно, что функция $\mathbf{f}=(1,-1,3)$ коррелирует с дискретной функцией $\mathbf{g}_3=(3,2,4)$, полученной из **g** циклическим сдвигом на две позиции влево (или на одну позицию вправо). Если построить графики функции **f** и $\mathbf{g}_3$, то они будут весьма похожи (см. рис. 10).

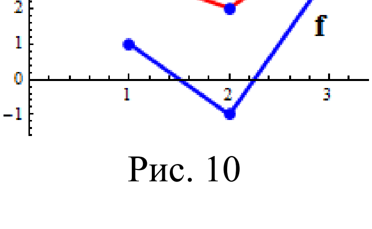


Рис. 10

∎

Пример 2. Пусть дискретные функции **f** и **g**, представляющие среднемесячные температуры в Харькове и Буэнос-Айресе, заданы в таблице примера 1 (первая и третья строки таблицы). Построим график функции взаимной корреляции $R(k)$ дискретных функций **f** и **g**.

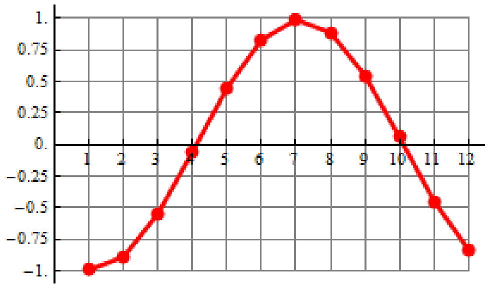


Рис. 11

Пик функции (равный 0.99) приходится на 7 – й месяц. Это означает, что поведение температуры для обоих городов похоже, и изменение температуры сдвинуто на $7-1=6$ месяцев, т.е. на полгода.

∎

Значение функции взаимной корреляции близкие к нулю при всех k указывают на отсутствие корреляции, т.е. функции совершенно независимы, например, если компоненты одной из функций абсолютно случайны. Примером случайной дискретной функции может быть последовательность очков, выпадающая на игральной кости.

Таким образом, с помощью функции взаимной корреляции можно измерить запаздывание во времени двух сигналов (дискретных функций).

Замечание. Иногда функция взаимной корреляции определяется формулой

$$R_k = \frac{1}{n}\sum_{i=0}^{n-1} f_i \cdot g_{i+k} \qquad (k=0,1,\ldots,n-1).$$

Но тогда ее значения могут выходить из диапазона $[-1,+1]$. При этом, для разнообразия, мы записали формулу для случая, когда компоненты дискретных функций нумеруются с нуля.

□

Автокорреляционная функция является частным случаем функции взаимной корреляции периодической дискретной функции с собой. Она определяется формулой

$$A_k = \frac{\sum_{i=1}^{n}\left(f_i-\overline{f}\right)\cdot\left(f_{i+k-1}-\overline{f}\right)}{\sum_{i=1}^{n}\left(f_i-\overline{f}\right)^2} \qquad (k=1,2,\ldots,n), \tag{18}$$

где индекс $i+(k-1)$ заменяется на $i+(k-1)-n$, когда $i+(k-1)$ становится больше n. Аналогом формулы (17) является выражение

$$A_k = \frac{\sum_{i=1}^{n}\left(f_i-\overline{f}\right)\cdot\left(f_{1+\mathrm{mod}(i+k-2,n)}-\overline{f}\right)}{\sum_{i=1}^{n}\left(f_i-\overline{f}\right)^2} \qquad (k=1,2,\ldots,n). \tag{19}$$

Автокорреляционная функция обладает свойством $A_1=1$. Функция взаимной корреляции обычно не принимает максимального значения при $i=1$.

Лемма. Автокорреляция периодической дискретной функции является периодической функцией. Действительно

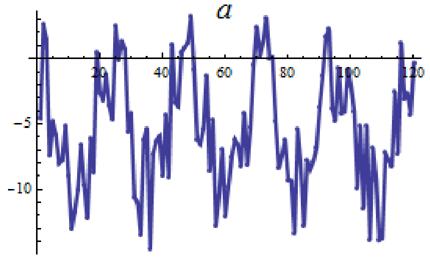
$$A_{k+n}=\frac{\sum_{i=1}^{n}\left(f_i-\overline{f}\right)\cdot\left(f_{1+\mathrm{mod}(i+k+n-2,n)}-\overline{f}\right)}{\sum_{i=1}^{n}\left(f_i-\overline{f}\right)^2}=\frac{\sum_{i=1}^{n}\left(f_i-\overline{f}\right)\cdot\left(f_{1+\mathrm{mod}(i+k-2,n)}-\overline{f}\right)}{\sum_{i=1}^{n}\left(f_i-\overline{f}\right)^2}=A_k.$$

Следовательно, A_k периодическая дискретная функция с периодом n.

∎

С помощью функции автокорреляции можно проанализировать дискретную функцию на наличие в ней периодического свойства.

Пример 3. Задана температура атмосферы в первые 5 дней января с почасовым изменением, т.е. задана дискретная функция с $5\cdot 24=120$ компонентами. График этой функции показан на рис. 13а. Имеют ли периодические свойства изменения температуры?



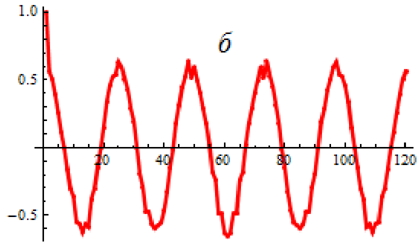


Рис. 13

Температуру в i-й час обозначим f_i. Результат вычисления функции автокорреляции по ф. (18) представлен на рис. 13б. Из него ясно, что функция R_k через каждые 24 значения индекса принимает пиковые значения. Это означает, что дискретная функция обладает периодическими свойствами с периодом 24 часа, что вполне естественно, потому, что температура атмосферы в дневное и в ночное время сильно отличается, но повторяется примерно через сутки. Это и отразила функция автокорреляции.

∎

Свертка дискретных функций. Пусть заданы две функции $\mathbf{f}=\left(f_1,\ldots,f_n\right)$ и $\mathbf{g}=\left(g_1,\ldots,g_n\right)$. Свертка $\mathbf{h}$ функций $\mathbf{f}$ и $\mathbf{g}$ обозначается $\mathbf{h}=\mathbf{f}\otimes\mathbf{g}$ и определяется формулой

$$h_k=\sum_{i=1}^{k} f_i g_{k-(i-1)} \quad \left(k=1,2,\ldots,n\right) \tag{20}$$

Распишем сумму (20)

$$h_k=\sum_{i=1}^{k} f_i g_{k-(i-1)}=f_1 g_k+f_2 g_{k-1}+\ldots+f_{k-1}g_2+f_k g_1$$

и переставим в ней слагаемые в обратном порядке

$$h_k=g_1 f_k+g_2 f_{k-1}+\ldots+g_{k-1}f_2+g_k f_1=\sum_{i=1}^{k} g_i f_{k-(i-1)}.$$

В результате получаем, что $\mathbf{f}\otimes\mathbf{g}=\mathbf{g}\otimes\mathbf{f}$, т.е. свертка коммутативна.

Пример. Вычислим свертку дискретных функций $\mathbf{f}=(1,2,3)$ и $\mathbf{g}=(5,3,1)$. Имеем

$$h_1 = f_1 g_1 = 1\cdot 5 = 5;$$
$$h_2 = f_1 g_2 + f_2 g_1 = 1\cdot 3 + 2\cdot 5 = 13;$$
$$h_3 = f_1 g_3 + f_2 g_2 + f_3 g_1 = 1\cdot 1 + 2\cdot 3 + 3\cdot 5 = 22;$$

В итоге, $\mathbf{f}\otimes\mathbf{g}=\mathbf{h}=(5,13,22)$.

∎

Замечание. Свертку можно определять для дискретных функций разной длины. Тогда придется договариваться о способе продолжения функций.

Для свертки выполняется закон коммутативности (доказан выше), закон дистрибутивности (докажите самостоятельно)

$$\mathbf{f}\otimes(\mathbf{g}+\mathbf{h})=\mathbf{f}\otimes\mathbf{g}+\mathbf{f}\otimes\mathbf{h},$$

и закон ассоциативности (докажите самостоятельно)

$$\mathbf{f}\otimes(\mathbf{g}\otimes\mathbf{h})=(\mathbf{f}\otimes\mathbf{g})\otimes\mathbf{h}.$$

Сглаживание. По своей природе дискретная функция представляет реальную физическую величину, которую необходимо обработать тем или иным способом. Часто такая величина содержит шум, возникающий вследствие погрешностей измерительных приборов, и от которого желательно избавиться. Если мы хотим увидеть поведение дискретной функции, то результаты измерения необходимо сгладить, удалив незначительные случайные отклонения от основной зависимости. Этот вид преобразования дискретной функции называют сглаживанием. Одним из методов, используемых для сглаживания, является преобразование скользящего среднего.

Для нахождения скользящего среднего в окрестности рассматриваемой точки *i* берем среднее арифметическое от K предыдущих и K последующих значений, и значение в самой точке. Т.е.

$$g_i = \frac{1}{2K+1}\left(f_{i-K} + f_{i-(K-1)} + \ldots + f_i + \ldots + f_{i+K}\right) = \frac{1}{2K+1}\sum_{j=-K}^{K} f_{i+j}\,. \tag{21}$$

Одним словом, новое значения g_i представляется как среднее значения 2K+1 соседних точек, включая точку *i*. Обратите внимание, что для K-1 первой и K-1 последней точки невозможно вычислить значение сглаживания, хотя можно выполнить усреднение по тем точкам, которые имеются. Если дискретная функция f_i периодическая (или продолжена периодическим образом), то преобразование скользящего среднего можно определить по формуле

$$g_i = \frac{1}{2K+1}\sum_{j=-K}^{K} f_{1+\mathrm{mod}(i-1+j,n)} \quad (i=1,2,\ldots,n). \tag{22}$$

и n – это количество компонент функции $\mathbf{f}$ на отрезке одного периода.

Пример. Вычислим преобразование скользящего среднего для периодической дискретной функции $\mathbf{f}=(1,3,-1,7,5)$ при K=1. Используя периодичность, имеем

$$g_1 = \frac{1}{3}(5+1+3)=3; \qquad g_2 = \frac{1}{3}(1+3-1)=1; \qquad g_3 = \frac{1}{3}(3-1+7)=3;$$

$$g_4 = \frac{1}{3}(-1+7+5) = \frac{11}{3}; \qquad g_5 = \frac{1}{3}(7+5+1) = \frac{13}{3}.$$

В итоге, $\mathbf{g} = \left(3, 1, 3, \frac{11}{3}, \frac{13}{3}\right)$.

∎

На рисунке 14 показана дискретная функция, имеющая 120 компонент и довольно сильное «зашумление».

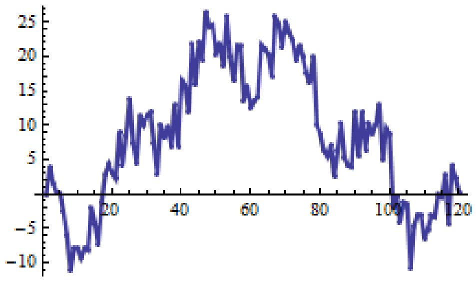


Рис. 14

Определим характерные черты этой дискретной функции в предположении ее периодичности. Для этого построим графики функций, полученных применением преобразования скользящего среднего при значениях K = 2, 4, 6, 12.

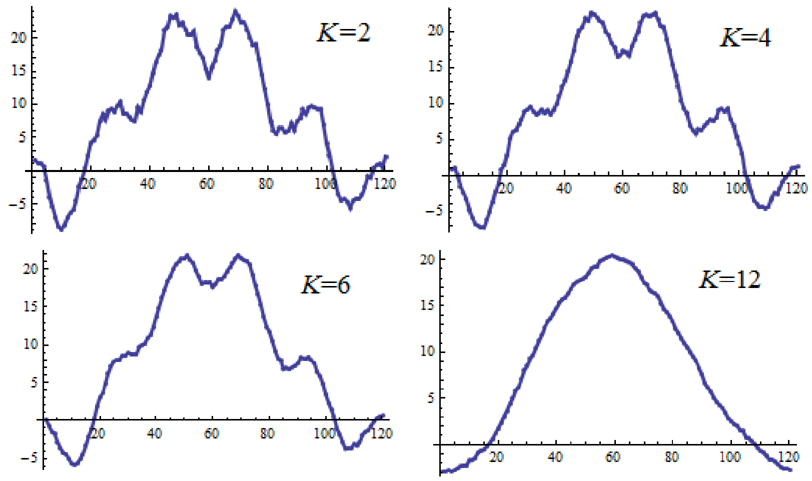


Рис. 15

При значении K=12 наблюдается синусоидальная форма кривой для которой период равен длине отрезка наблюдения, т.е. n=120. При значениях $K \le 6$ мы видим, что на форму основной синусоиды наложена синусоидальное изменение с меньшим периодом. Действительно, исходную дискретную функцию, показанную на рис. 14, мы построили по формуле

$$y = 8 - 13\cos\frac{2\pi x}{120} + 5\cos\frac{2\pi x}{24} + \mathrm{RandomReal}(-5, 5),$$

где функция `RandomReal(-5,5)` генерирует случайные вещественные числа из диапазона [-5, 5].

В формулах (21) и (22) все точки из диапазона $[i-K, \ldots, i-1, i, i+1, \ldots, i+K]$, используемые при вычислении скользящего среднего, мы рассматривали с одинаковой степенью значимости. Однако во многих случаях необходимо учитывать то, что, чем ближе точка к рассматриваемой точке i, тем выше ее значимость. Когда выполняется сглаживание, нужно учитывать «вес» каждой точки в соответствии с ее «значимостью». Это можно записать выражением

$$g_i = \sum_{j=-K}^{K} w_j \, f_{i+j} \, . \tag{23}$$

При этом на «весовые» коэффициенты $\{w_j\}_{-K}^{K}$ накладывается условие $\sum_{j=-K}^{K} w_j = 1$. Случай, когда все весовые коэффициенты одинаковы $w_i = 1/(2K+1)$, является частным случаем и называется методом арифметического скользящего среднего или простого скользящего среднего.

Пример. Вычислим преобразование скользящего среднего для периодической дискретной функции $\mathbf{f} = (1, 3, -1, 7, 5)$ с K=1 и весовыми коэффициентами $w_0 = 1/2$ и $w_{\pm 1} = 1/4$. Используя периодичность, имеем

$$g_1 = \frac{1}{4}5 + \frac{1}{2}1 + \frac{1}{4}3 = \frac{5}{2}; \qquad g_2 = \frac{1}{4}1 + \frac{1}{2}3 + \frac{1}{4}1 = 2; \quad g_3 = \frac{1}{4}3 + \frac{1}{2}(-1) + \frac{1}{4}7 = 2;$$

$$g_4 = \frac{1}{4}(-1) + \frac{1}{2}7 + \frac{1}{4}5 = \frac{9}{2}; \quad g_5 = \frac{1}{4}7 + \frac{1}{2}5 + \frac{1}{4}1 = \frac{9}{2}.$$

В итоге, $\mathbf{g} = \left(\frac{5}{2}, 2, 2, \frac{9}{2}, \frac{9}{2} \right)$.

∎

Фильтрация состоит из выполнения для каждой компоненты дискретной функции некоторой операции, которая использует лишь значения функции в фиксированной окрестности вокруг этой компоненты. Для такой процедуры принято использовать термин окрестностная обработка или фильтрация.

Если операции, совершаемые над компонентами окрестности, являются линейными, то вся процедура называется линейной фильтрацией. Линейные операции состоят из умножения каждой компоненты окрестности на соответствующие коэффициенты и суммирование этих произведений для получения результирующего отклика в каждой точке *i*. Если окрестность имеет размер m, то потребуется столько же коэффициентов. Эти коэффициенты сгруппированы в виде вектора **w**, который называется фильтром (маской, фильтрующей маской, ядром, шаблоном).

Механизм линейной фильтрации заключается в перемещении базовой точки (обычно центра) фильтрующей маски **w** от компоненты к компоненте дискретной функции **f**. В каждой точке откликом фильтра является сумма произведений коэффициентов фильтра и значений соответствующих компонент окрестности. Для маски размера m обычно предполагается, что m – нечетное положительное число, чтобы в качестве базовой точки можно было взять ее центр. В тех ситуациях, когда маска выходит за пределы дискретной функции применяют продолжение. Если нет особых предпочтений, то дискретную функцию продолжают константами, совпадающими с крайними значениями (см. рис. 9б). Фактически формула (23) описывает механизм работы линейного фильтра, а преобразование скользящего среднего является частным случаем окрестностной обработки (фильтрации).

Например, фильтр с маской $\mathbf{w} = (0, -1, 1)$ и центральной базовой точкой выполняет преобразование дискретной функции **f** по формуле $g_i = f_{i+1} - f_i$, что является дискретным аналогом вычисления производной. Другим ее аналогом является применение фильтра $w = \left(-\frac{1}{2}, 0, \frac{1}{2}\right)$, т.е. $g_i = \frac{f_{i+1} - f_{i-1}}{2}$. Аналогом вычисления второй производной дискретной функции является применение фильтра $\mathbf{w} = (1, -2, 1)$, т.е. $g_i = f_{i+1} - 2f_i + f_{i-1}$.

Часто используют фильтры порядка (порядковых статистик), вычисление отклика которых требует упорядочивания значений компонент, заключенных внутри обрабатываемой фильтром области и последующего выбора некоторой из них из полученного вектора. Например, можно выбрать компоненту, находящуюся посередине (это не среднее значение) или на конце такого вектора (максимальное или минимальное значение).

Самым известным из таких фильтров является медианный фильтр, который соответствует выбору среднего элемента маски. Пусть например, дискретная функция содержит последовательность

$$\mathbf{f} = (\ldots, 3, 5, 17, \mathbf{34}, 21, 24, -21, \ldots)$$

и длина фильтра равна 7. Упорядочив элементы в области, обрабатываемой фильтром, мы получаем вектор $(-21, 3, 5, 17, 21, 24, 34)$. Число 17 расположено в его середине и оно выбирается в качестве отклика фильтра для компоненты 34.

Пример. Применим медианный фильтр длины 3 к дискретной функции $\mathbf{f} = (2, 60, 8, 3)$. Имеем (искусственно введенные продолженные значения выделены жирным шрифтом).

$$g_1 = \text{медиана}(\mathbf{2}, 2, 60) = 2; \quad g_2 = \text{медиана}(2, 60, 8) = 8;$$
$$g_3 = \text{медиана}(60, 8, 3) = 8; \quad g_4 = \text{медиана}(8, 3, \mathbf{3}) = 3.$$

В итоге: $\mathbf{g} = (2, 8, 8, 3)$.

∎

На следующем рисунке показан график дискретной функции, имеющей синусоидальное поведение, но значения некоторых из ее компонент изменены случайным образом (это не аддитивный шум).

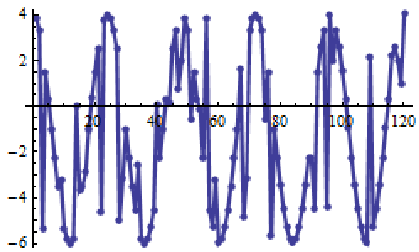


Рис. 16

На рис. 17 слева показан результат фильтрации медианным фильтром при m=17, а справа показан результат применения метода арифметического скользящего среднего при K=8 ($m = 2K + 1$).

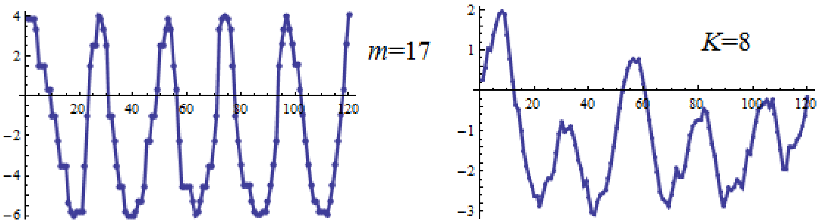


Рис. 17 Результат фильтрации медианным фильтром (слева) и фильтром простого скользящего среднего (справа).

Разные методы фильтрации дают различные результаты в зависимости от природы «шума». На рис. 14 показана дискретная функция с аддитивным шумом (случайные значения добавляются к истинному значению физической величины) и с ним хорошо справляется фильтр арифметического скользящего среднего. На рис. 16 показана дискретная функция со случайным шумом (для некоторых компонент истинные значения физический величины заменяются на случайные) и для нее лучший результат дает медианная фильтрация.

Кроме линейных фильтров, часто используют нелинейные, а также рекурсивные фильтры, которые мы здесь обсуждать не будем.

5. Булева алгебра и ее приложения

Булева алгебра — это название области математики, занимающейся логическим анализом. Операции и законы булевой алгебры применяются к логическим символам так же, как обычная алгебра оперирует символами, представляющими численные величины.

В настоящей главе мы изучим булеву алгебру, а именно множество {0, 1} с определенными на нем операциями дизъюнкции, конъюнкции и отрицания. Мы сформулируем законы булевой алгебры и покажем, как булевы выражения упрощаются. Последние параграфы главы демонстрирует некоторые приложения теории для конструирования логических схем, встречающихся в электронных устройствах: компьютерах, калькуляторах, телефонных системах и ряде других устройств.

5.1 Булева алгебра

Булева алгебра состоит из множества $B=\{0,1\}$ вместе с определенными на нем операциями дизъюнкции (V), конъюнкции ($\wedge$) и отрицания (~). Действие операций ($\vee$) и ($\wedge$) на символах 0 и 1 показаны в следующих таблицах

$\vee$	0	1
0	0	1
1	1	1

$\wedge$	0	1
0	0	0
1	0	1

Действие отрицания на 0 и 1 определяется следующим образом: $\bar{0}=1$ и $\bar{1}=0$.

Переменные, которые могут принимать только два значения 0 и 1, называются булевыми. Тогда для них таблицы, определяющие действия операций $\bar{p}$, $p \vee q$, $p \wedge q$, будут иметь вид

p	$\bar{p}$
0	1
1	0

p	q	$p \vee q$	$p \wedge q$
0	0	0	0
0	1	1	0
1	0	1	0
1	1	1	1

Эти таблицы напоминают таблицы истинности логических операций не, или и и. Действительно, мы можем легко трансформировать их в таблицы истинности, возникающие в логике высказываний, заменив булевы переменные p и q на высказывания P и Q, и используя истинностные значения T и F вместо 1 и 0 соответственно. Таким образом, $\overline{p}$ заменяется на **не P**, $p \vee q$ – на **P или Q**, а $p \wedge q$ – на **P и Q**. Поэтому в контексте булевой алгебры мы будем называть такого сорта таблицы таблицами истинности.

Мы можем комбинировать булевы переменные с помощью операций $\vee, \wedge, \sim$, получая булевы выражения так же, как мы строили составные высказывания из более простых, комбинируя их с помощью логических операций. Также как и в алгебре логики, мы полагаем, что два булевых выражения являются эквивалентными, если они имеют одинаковые таблицы истинности.

Часто, вместо знаков операций $\vee, \wedge, \sim$ используют знаки $+, \cdot\,(\textit{умножение}),\,'$ (штрих – отрицание). В этих обозначениях законы алгебры логики, перенесенные на булевы переменные, имеют вид:

а) *Законы коммутативности*

$$x \cdot y = y \cdot x;$$

$$x + y = y + x.$$

б) *Законы ассоциативности*

$$x \cdot (y \cdot z) = (x \cdot y) \cdot z;$$

$$x + (y + z) = (x + y) + z.$$

в) *Законы дистрибутивности*

$$x \cdot (y + z) = (x \cdot y) + (x \cdot z);$$

$$x + (y \cdot z) = (x + y) \cdot (x + z).$$

г) *Законы тождества*

$$x \cdot 1 = x;$$

$$x + 0 = x.$$

д) *Законы дополнения*

$$x \cdot x' = 0;$$

$$x + x' = 1.$$

Для проверки истинности законов булевой алгебры мы должны построить таблицы истинности для выражений левой и правой частей равенств.

Пример. Докажем закон дистрибутивности:

$$p \cdot (q + r) = p \cdot q + p \cdot r$$

Построим таблицы истинности для левого и правого выражения, а также для промежуточных операндов

p	q	r	$p \bullet q$	$p \bullet r$	$q + r$	$p \bullet (q + r)$	$p \bullet q + p \bullet r$
0	0	0	0	0	0	0	0
0	0	1	0	0	1	0	0
0	1	0	0	0	1	0	0
0	1	1	0	0	1	0	0
1	0	0	0	0	0	0	0
1	0	1	0	1	1	1	1
1	1	0	1	0	1	1	1
1	1	1	1	1	1	1	1

Поскольку два последних столбца таблицы полностью совпадают, то и соответствующие им булевы выражения совпадают (эквивалентны). □

Схожесть названий и форм законов булевой алгебры и соответствующих законов алгебры логики и алгебры множеств, далеко не случайна. В таблице, приведенной ниже, показано соответствие между булевыми операциями, операторами логики высказываний и операциями над множествами.

Логические операторы	Операции над множествами	Булевы операции
не	$-$	~,’
или	$\cup$	$\vee$, +
и	$\cap$	$\wedge$, •

Чтобы в дальнейшем не обращаться к таблицам истинности при доказательстве эквивалентности булевых выражений, надо один раз убедиться в справедливости законов булевой алгебры и применять их при преобразовании булевых выражений.

Пример. Покажем, что булево выражение $\left(\overline{\overline{p} \cdot q}\right) \cdot (p + q)$ эквивалентно p. Здесь удобно использовать кванторы для обозначения булевых операций. Имеем

$$
\begin{aligned}
&\overline{(\bar{p} \wedge q)} \wedge (p \vee q) = \\
&= \left(\overline{(\bar{p})} \vee \bar{q}\right) \wedge (p \vee q) = && \text{(закон де Моргана)} \\
&= (p \vee \bar{q}) \wedge (p \vee q) = && \text{(так как } \overline{(\bar{p})} = p\text{)} \\
&= p \vee (\bar{q} \wedge q) = && \text{(закон дистрибутивности)} \\
&= p \vee 0 = && \text{(так как } \bar{q} \wedge q = 0\text{)} \\
&= p && \text{(по определению } \vee\text{).}
\end{aligned}
$$

Булевой функцией от n булевых переменных $p_1, p_2, \ldots, p_n$ называется функция $f : B^n \to B$. По-другому, результатом вычисления функции $f(p_1, p_2, \ldots, p_n)$, аргументы которой могут принимать только два значения 0 и 1, также является 0 или 1. Очевидно, что количество булевых функций n переменных конечно. Например, булевых функций одной переменной существует всего 4. Булевы функции от одной переменной – это отображения множества {0,1} в себя. Их можно рассматривать как унарные операции на множестве {0,1}. В следующей таблице приведены все четыре булевы функции от одной переменной.

	Название функции	Обозначение функции	Значение аргумента 0	1
$g_0(x)$	Нуль	0	0	0
$g_1(x)$	Тождественная	x	0	1
$g_2(x)$	Отрицание	$-x, x', \bar{x}$	1	0
$g_3(x)$	Единица	1	1	1

Булевы функции от двух переменных можно рассматривать как бинарные операции на множестве {0,1}. В следующей таблице приведены все шестнадцать булевых функции от двух переменных. Для некоторых функций указаны используемые обозначения и названия.

Название функции	Переменная x Переменная y Обозначение функции	0 0	0 1	1 0	1 1
Нуль	0	0	0	0	0
Конъюнкция	$\cdot, \wedge$	0	0	0	1
		0	0	1	0
		0	0	1	1
		0	1	0	0
		0	1	0	1
Сложение по модулю 2	$+, \Delta$	0	1	1	0
Дизъюнкция	$\vee$	0	1	1	1
Стрелка Пирса	$\downarrow$	1	0	0	0
Эквивалентность	$\sim, \equiv$	1	0	0	1
		1	0	1	0
		1	0	1	1
		1	1	0	0
Импликация	$\rightarrow, \Rightarrow, \supset$	1	1	0	1
Штрих Шеффера	$\mid$	1	1	1	0
Единица	1	1	1	1	1

Комбинируя перечисленные функции (с помощью суперпозиций), можно строить более сложные булевы функции, в том числе и большего числа переменных. Булевы отрицание, конъюнкция, дизъюнкция, импликация обладают свойствами, подобными тем, которыми обладают соответствующие логические операции.

Формулы, содержащие кроме переменных (и скобок), только знаки функций дизъюнкции, конъюнкции и отрицания, будем называть булевыми формулами.

Если задана булева функция, например, таблицей истинности, то обычно требуется ее представить с помощью булевых функций двух переменных. Наша ближайшая задача — показать, как произвольную булеву функцию можно представить с помощью бинарных операций $\vee, \wedge, \sim$

Рассмотрим булеву функцию m(p,q,r) от трех булевых переменных p,q и r с таблицей истинности, приведенной ниже

p	q	r	m
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	0
1	1	0	0
1	1	1	0

Функция m принимает значение 1 только на одном наборе значений аргументов. Такая функция называется минтермом или элементарной конъюнкцией. Ее можно выразить через булевы функции двух переменных следующим образом

$$m(p,q,r) = \overline{p} \wedge q \wedge r .$$

Выражения такого типа называются элементарной конъюнкцией, если все входящие в него переменные p,q,…,r (или их отрицания) различны.

Любой минтерм можно записать в виде элементарной конъюнкции, т.е. как конъюнкцию переменных p_i или их отрицаний. Действительно, пусть $m(p_1,p_2,\dots p_r)$ – минтерм. Тогда в последнем столбце таблицы истинности функции m будет стоять только одна единица. Возьмем строку таблицы истинности, соответствующую значению 1. Если в этой строке переменная $p_i=1$, то в элементарной конъюнкции, представляющей функцию m, участвует p_i, а если $p_i=0$, то участвует $\overline{p_i}$. Например, для минтерма из последней таблицы значения переменных p, q, r равны 0, 1, 1. Это значит, что в элементарную конъюнкцию входят $\overline{p}, q$ и r, т.е. $m(p,q,r) = \overline{p} \wedge q \wedge r$.

Используя элементарные конъюнкции, можно записать произвольную булеву функцию как дизъюнкцию минтермов. Например, рассмотрим булеву функцию трех переменных f(p,q,r) с таблицей истинности следующего вида

p	q	r	f
0	0	0	0
0	0	1	1
0	1	0	0
0	1	1	1
1	0	0	1
1	0	1	0
1	1	0	0
1	1	1	0

Единицы последнего столбца в этой таблице соответствуют трем минтермам. Запишем их и соединим операцией дизъюнкции

$$f(p, q, r) = (\bar{p} \wedge \bar{q} \wedge r) \vee (\bar{p} \wedge q \wedge r) \vee (p \wedge \bar{q} \wedge \bar{r})$$

Очевидно, в той же форме можно записать булеву функцию с любым числом переменных. Выражение, представляющее булеву функцию как дизъюнкцию минтермов, называется **дизъюнктивной нормальной формой** (ДНФ). Описанный способ построения формулы по таблице применим к любой функции, не равной тождественно нулю. Т.е. единственной функцией, не имеющей ДНФ, является константа 0. Но ее можно представить булевой формулой $p \wedge \overline{p}$. Можно также считать, что ДНФ тождественного нуля – это «пустая» дизъюнкция, не содержащая ни одного дизъюнктивного слагаемого.

Повторим. Дизъюнктивной нормальной формой (ДНФ) булевой функции называется её представление в виде дизъюнкции некоторых элементарных конъюнкций.

Любую булеву функцию можно представить в виде дизъюнкции минтермов. Значит, каждая булева функция может быть выражена через две функции двух аргументов: $f_1(p,q) = p \vee q$, $f_2(p,q) = p \wedge q$ и одной функции одной переменной $f(p) = \overline{p}$. Множество функций, через которые можно выразить любую булеву функцию, называется полной системой функций. Итак, f_1, f_2, f_3 — полная система функций. Однако можно ограничиться и меньшим количеством функций. Например, по закону де Моргана $\overline{(p \vee q)} = \overline{p} \wedge \overline{q}$. Следовательно $p \vee q = \overline{\overline{p} \wedge \overline{q}}$. Значит, любую булеву функцию можно записать только с помощью двух операций $\wedge$ и $\sim$, т.е. $\{ p \wedge q, \overline{p} \}$ — тоже полная система функций. Расплатой за малое количество операций, посредством которых записывается функция, становится громоздкость формул.

Пример. Функция НЕ–И (обозначим ее как в алгебре логики через |) определяется формулой: $p | q = \overline{p \wedge q}$. Покажем, что { | } — полная система функций. Для этого достаточно показать, что каждая из функций $\overline{p}, p \wedge q, p \vee q$ может быть выражена через НЕ–И. Ввиду закона идемпотентности

$$\overline{p} = \overline{p \wedge p} = p | p \qquad \text{(A)}$$

По закону де Моргана

$$p \vee q = \overline{\overline{p} \wedge \overline{q}} = \left|\, \text{èç}\,(A) \right| = \overline{(p\,|\,p) \wedge (q\,|\,q)} = (p\,|\,p)\;|\;(q\,|\,q)$$

Также

$$p \wedge q = \overline{\overline{p \wedge q}} = \overline{p\,|\,q} = \left|\, \text{èç}\,(A)\;\; \overline{s} = s\,|\,s \;\text{äëÿ}\;\; s = p\,|\,q \;\right| = (p\,|\,q)\,|\,(p\,|\,q)$$

Таким образом, { | } — действительно полная система операций.

□

Двойственной к булевой функции `f(x,y,…,z)` называется функция

$$f^*(x,y,\ldots,z) = f'(x',y',\ldots,z'),$$

которая получается отрицанием всех аргументов булевой функции `f` и последующим отрицанием результата. Из определения видно, что $f^{**}=f$.

Пример **Двойственные функции:**

f	1	0	$x_1 \vee x_2$	$x_1 \wedge x_2$	x	$\overline{x}$
f^*	0	1	$x_1 \wedge x_2$	$x_1 \vee x_2$	x	$\overline{x}$

Функция называется **самодвойственной**, если $f^*=f$. Из приведенной таблицы видно, что тождественная функция и отрицание самодвойственные, а дизъюнкция и конъюнкция – нет.

Суперпозицией функций `f`$_1$`,…,f`$_m$ называется булева функция, полученная с помощью подстановок этих функций друг в друга на места переменных и, возможно, с помощью переименования переменных.

Пример. Составить таблицу истинности функции $h(x,y)=f_2(y,y,f_1(x,y,x))$, где булевы функции трех переменных f_1 и f_2 имеют следующие таблицы истинности

xyz	f_1	f_2
000	1	0
001	0	1
010	0	1
011	1	0
100	0	1
101	1	0
110	1	1
111	1	1

Для составления таблицы функции $h(x,y)$ запишем формулу, задающую функцию $h(x,y)$, выписав под символами переменных все наборы значений, которые эти переменные принимают, а под символами булевых функций выписав значения функций, соответствующие этим наборам. Заключительный столбец, задающий функцию ***h***, обведём двойной рамкой.

$h(x,y) = f_2(y,\ y,\ f_1(x,\ y,\ x))$

	0	0	1	0	0	1	0	0	0
	0	1	1	1	1	0	0	1	0
	1	0	1	0	0	1	1	0	1
	1	1	1	1	1	1	1	1	1

Итак, ***h(x,y)***=(1111).

5.2 Карта Карно

Мы показали, что любую булеву функцию можно выразить в дизъюнктивной нормальной форме. Однако получаемые выражения довольно громоздки. Здесь мы попытаемся решить задачу об «упрощении» выражения для булевой функции. Под «упрощением» мы подразумеваем эквивалентное выражение, использующее меньше символов, чем исходное. Для этого используются карты Карно – наглядные схемы, предназначенные для обнаружения пар минтермов, которые можно сгруппировать и преобразовать в одно простое выражение.

Рассмотрим n булевых переменных $p_1, p_2, \ldots, p_n$. Для них существует 2^n различных элементарных конъюнкций. Например, для двух булевых переменных p и q элементарными конъюнкциями будут $p \wedge q$, $p \wedge \bar{q}$, $\bar{p} \wedge q$, $\bar{p} \wedge \bar{q}$. Карта Карно — это таблица, каждый элемент которой является элементарной конъюнкцией. Для переменных p и q карта Карно имеет вид, изображенный на следующем рисунке слева

	q	$\sim q$
p		
$\sim p$		

	q	$\sim q$
p	$p \wedge q$	$p \wedge \sim q$
$\sim p$	$\sim p \wedge q$	$\sim p \wedge \sim q$

На правой схеме мы для наглядности внутри ячеек представили соответствующие им элементарные конъюнкции.

Для представления картой Карно булевой функции, записанной в дизъюнктивной нормальной форме, необходимо поместить $\times$ (крестик) или 1 (единицу) в прямоугольниках, соответствующих элементарным конъюнкциям (минтермам), входящим в выражение. Например, высказыванию $(p \wedge q) \vee (\sim p \wedge \sim q)$ соответствует карта Карно, изображенная на следующем рисунке

	q	$\sim q$
p	$\times$	
$\sim p$		$\times$

Когда выражению соответствует карта Карно с двумя соседствующими в строке или в столбце крестиками, тогда его можно упростить, сведя две *элементарные конъюнкции к одной, содержащей на одну компоненту меньше* (т.е. либо p, либо q не будут присутствовать в выражении). Например, высказывание $(p \wedge q) \vee (p \wedge \sim q)$, которому соответствует карта Карно, изображенная на следующем рисунке

	q	$\sim q$
p	×	×
$\sim p$		

эквивалентна p, так как

$$(p \wedge q) \vee (p \wedge \sim q) \equiv p \wedge (q \vee \sim q) \equiv p \wedge T \equiv p$$

Для булевой функции трех переменных p, q и r карта Карно имеет следующий вид

	q		$\sim q$	
p				
$\sim p$				
	r	$\sim r$		r

На следующем рисунке для наглядности в прямоугольники мы вписали элементарные конъюнкции

	q		$\sim q$	
p	$p \wedge q \wedge r$	$p \wedge q \wedge \sim r$	$p \wedge \sim q \wedge \sim r$	$p \wedge \sim q \wedge r$
$\sim p$	$\sim p \wedge q \wedge r$	$\sim p \wedge q \wedge \sim r$	$\sim p \wedge \sim q \wedge \sim r$	$\sim p \wedge \sim q \wedge r$
	r	$\sim r$		r

Здесь ячейки карты Карно соответствуют восьми минтермам, которые можно построить из трех булевых переменных. Например, выражению

$$(p \wedge q \wedge \sim r) \vee (p \wedge \sim q \wedge r) \vee (\sim p \wedge q \wedge \sim r)$$

будет соответствовать карта Карно, показанная на следующем рисунке

	q		$\sim q$	
p		×		×
$\sim p$		×		
	r	$\sim r$		r

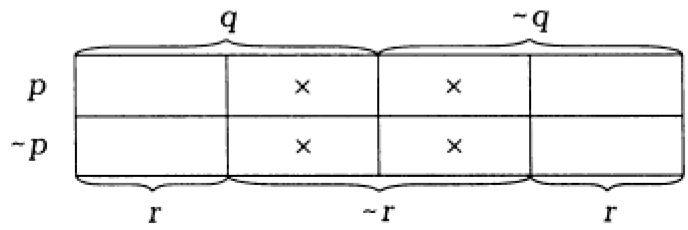
Если два знака (крестика) соседствуют, две элементарные конъюнкции могут быть сведены к одной, содержащей на одну из компонент p, q и r меньше. В данном случае дизъюнкция пары элементарных конъюнкций

$$(p \wedge q \wedge \sim r) \vee (\sim p \wedge q \wedge \sim r) \equiv ((q \wedge \sim r) \wedge p) \vee ((q \wedge \sim r) \wedge \sim p) \equiv$$
$$\equiv (q \wedge \sim r) \wedge (p \vee \sim p) \equiv (q \wedge \sim r) \wedge T \equiv q \wedge \sim r$$

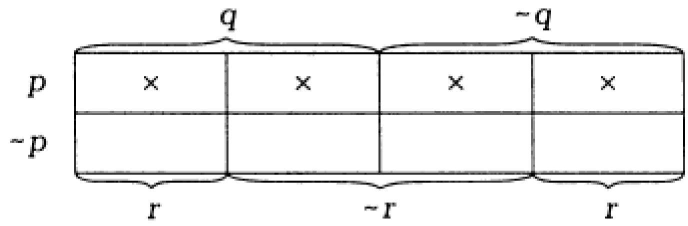
сводится к $q \wedge \sim r$, так что выражение принимает вид

$$(q \wedge \sim r) \vee (p \wedge \sim q \wedge r).$$

Если четыре значка x расположены в прямоугольнике рядом



или



тогда четыре элементарные конъюнкции, отмеченные значками, могут быть сведены к одному члену, содержащему только одну из компонент `p`, `q` или `r`. Например, первая карта Карно представляет выражение

$$\left(p \wedge q \wedge \bar{r}\right) \vee \left(p \wedge \bar{q} \wedge \bar{r}\right) \vee \left(\bar{p} \wedge q \wedge \bar{r}\right) \vee \left(\bar{p} \wedge \bar{q} \wedge \bar{r}\right),$$

которое сводится к `~r`. Действительно

$$\left(p \wedge q \wedge \bar{r}\right) \vee \left(p \wedge \bar{q} \wedge \bar{r}\right) \vee \left(\bar{p} \wedge q \wedge \bar{r}\right) \vee \left(\bar{p} \wedge \bar{q} \wedge \bar{r}\right) \equiv$$
$$\equiv \left(\left(\left(p \wedge \bar{r}\right) \wedge q\right) \vee \left(\left(p \wedge \bar{r}\right) \wedge \bar{q}\right)\right) \vee \left(\left(\left(\bar{p} \wedge \bar{r}\right) \wedge q\right) \vee \left(\left(\bar{p} \wedge \bar{r}\right) \wedge \bar{q}\right)\right) \equiv$$
$$\equiv \left(\left(p \wedge \bar{r}\right) \wedge \left(q \vee \bar{q}\right)\right) \vee \left(\left(\bar{p} \wedge \bar{r}\right) \wedge \left(q \vee \bar{q}\right)\right) \equiv$$
$$\equiv \left(\left(p \wedge \bar{r}\right) \wedge T\right) \vee \left(\left(\bar{p} \wedge \bar{r}\right) \wedge T\right) \equiv \left(p \wedge \bar{r}\right) \vee \left(\bar{p} \wedge \bar{r}\right) \equiv$$
$$\equiv \left(\bar{r} \wedge p\right) \vee \left(\bar{r} \wedge \bar{p}\right) \equiv \bar{r} \wedge \left(p \vee \bar{p}\right) \equiv \bar{r}$$

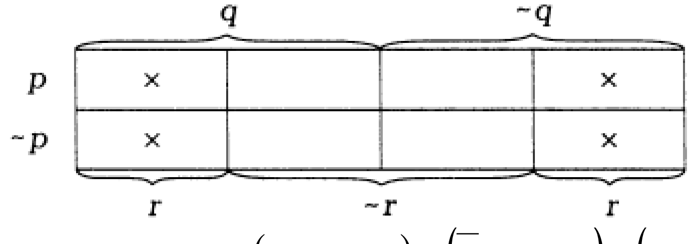
Вторая карта Карно представляет выражение

$$\left(p \wedge q \wedge r\right) \vee \left(p \wedge q \wedge \bar{r}\right) \vee \left(p \wedge \bar{q} \wedge \bar{r}\right) \vee \left(p \wedge \bar{q} \wedge r\right)$$

которое может быть сведено к `p`. Действительно

$$\left(p \wedge q \wedge r\right) \vee \left(p \wedge q \wedge \bar{r}\right) \vee \left(p \wedge \bar{q} \wedge \bar{r}\right) \vee \left(p \wedge \bar{q} \wedge r\right) \equiv$$
$$\equiv \left(\left(\left(p \wedge q\right) \wedge r\right) \vee \left(\left(p \wedge q\right) \wedge \bar{r}\right)\right) \vee \left(\left(\left(p \wedge \bar{q}\right) \wedge \bar{r}\right) \vee \left(\left(p \wedge \bar{q}\right) \wedge r\right)\right) \equiv$$
$$\equiv \left(\left(p \wedge q\right) \wedge \left(r \vee \bar{r}\right)\right) \vee \left(\left(p \wedge \bar{q}\right) \wedge \left(r \vee \bar{r}\right)\right) \equiv$$
$$\equiv \left(\left(p \wedge q\right) \wedge T\right) \vee \left(\left(p \wedge \bar{q}\right) \wedge T\right) \equiv \left(p \wedge q\right) \vee \left(p \wedge \bar{q}\right) \equiv$$
$$\equiv \left(p \wedge q\right) \vee \left(p \wedge \bar{q}\right) \equiv p \wedge \left(q \vee \bar{q}\right) \equiv p \wedge T \equiv p$$

Первый и последний столбец таблицы также считаются соседними (можно считать, что карта «скручена»). Например, карта Карно



представляет выражение $\left(p \wedge q \wedge r\right) \vee \left(\bar{p} \wedge q \wedge r\right) \vee \left(p \wedge \bar{q} \wedge r\right) \vee \left(\bar{p} \wedge \bar{q} \wedge r\right)$, которое может быть приведено к `r`. Действительно

$$\left(p \wedge q \wedge r\right) \vee \left(\bar{p} \wedge q \wedge r\right) \vee \left(p \wedge \bar{q} \wedge r\right) \vee \left(\bar{p} \wedge \bar{q} \wedge r\right) \equiv$$

$$\equiv\left(\left(\left(q\wedge r\right)\wedge p\right)\vee\left(\left(q\wedge r\right)\wedge\overline{p}\right)\right)\vee\left(\left(\left(\overline{q}\wedge r\right)\wedge p\right)\vee\left(\left(\overline{q}\wedge r\right)\wedge\overline{p}\right)\right)\equiv$$
$$\equiv\left(\left(q\wedge r\right)\wedge\left(p\vee\overline{p}\right)\right)\vee\left(\left(\overline{q}\wedge r\right)\wedge\left(p\vee\overline{p}\right)\right)\equiv$$
$$\equiv\left(\left(q\wedge r\right)\wedge T\right)\vee\left(\left(\overline{q}\wedge r\right)\wedge T\right)\equiv\left(q\wedge r\right)\vee\left(\overline{q}\wedge r\right)\equiv$$
$$\equiv\left(r\wedge q\right)\vee\left(\overline{r\wedge q}\right)\equiv r\wedge\left(q\vee\overline{q}\right)\equiv r\wedge T\equiv r$$

Выражение

$$(p\wedge q\wedge r)\vee(\sim p\wedge q\wedge r)\vee(p\wedge q\wedge\sim r)\vee(\sim p\wedge q\wedge\sim r)\vee(p\wedge\sim q\wedge\sim r)$$

представляется следующей картой Карно

	q	q	$\sim q$	$\sim q$
p	×	×	×	
$\sim p$	×	×		
	r	$\sim r$	$\sim r$	r

Четыре крестика в левых двух столбцах после упрощения дадут q, и выражение примет вид $q\vee\left(p\wedge\sim q\wedge\sim r\right)$. Благодаря наличию блока из двух значков в середине первой строки, это выражение можно еще больше упростить, приведя к виду $q\vee\left(p\wedge\sim r\right)$. Действительно

$$q\vee\left(p\wedge\sim q\wedge\sim r\right)\equiv q\vee\left(\sim q\wedge\left(p\wedge\sim r\right)\right)\equiv\left(q\vee\sim q\right)\wedge\left(q\vee\left(p\wedge\sim r\right)\right)\equiv$$
$$\equiv T\wedge\left(q\vee\left(p\wedge\sim r\right)\right)\equiv q\vee\left(p\wedge\sim r\right)$$

Фактически в дизъюнкцию минтермов мы два раза включаем конъюнкцию $p\wedge q\wedge\sim r$. Это, очевидно, не меняет результата, но позволяет включить этот минтерм в дизъюнкцию четырех минтермов в двух левых столбцах и в дизъюнкцию двух минтермов, соответствующих клеткам в середине первой строки. Дизъюнкция четырех минтермов упростилась до q, а двух – до $p\wedge\sim r$.

Для булевой функции четырех переменных также можно строить карты Карно и применять аналогичную технику упрощения булевых выражений. Проиллюстрируем этот процесс путем построения карты Карно для четырех булевых переменных `p,q,r` и `s`. Такая карта Карно может иметь следующий вид

	q	q	$\sim q$	$\sim q$	
p	$p\wedge q\wedge r\wedge s$	$p\wedge q\wedge\sim r\wedge s$	$p\wedge\sim q\wedge\sim r\wedge s$	$p\wedge\sim q\wedge r\wedge s$	s
p	$p\wedge q\wedge r\wedge\sim s$	$p\wedge q\wedge\sim r\wedge\sim s$	$p\wedge\sim q\wedge\sim r\wedge\sim s$	$p\wedge\sim q\wedge r\wedge\sim s$	$\sim s$
$\sim p$	$\sim p\wedge q\wedge r\wedge\sim s$	$\sim p\wedge q\wedge\sim r\wedge\sim s$	$\sim p\wedge\sim q\wedge\sim r\wedge\sim s$	$\sim p\wedge\sim q\wedge r\wedge\sim s$	$\sim s$
$\sim p$	$\sim p\wedge q\wedge r\wedge s$	$\sim p\wedge q\wedge\sim r\wedge s$	$\sim p\wedge\sim q\wedge\sim r\wedge s$	$\sim p\wedge\sim q\wedge r\wedge s$	s
	r	$\sim r$	$\sim r$	r	

Здесь в ячейки мы поместили элементарные конъюнкции, им соответствующие, а не крестики, как это следует делать для конкретного выражения. В качестве примера рассмотрим выражение

$$(p \wedge q \wedge r \wedge s) \vee (p \wedge q \wedge \sim r \wedge s) \vee (p \wedge q \wedge r \wedge \sim s) \vee$$
$$\vee (p \wedge q \wedge \sim r \wedge \sim s) \vee (\sim p \wedge q \wedge r \wedge \sim s) \vee (\sim p \wedge q \wedge \sim r \wedge \sim s) \vee$$
$$\vee (\sim p \wedge q \wedge r \wedge s) \vee (\sim p \wedge q \wedge \sim r \wedge s) \vee (p \wedge \sim q \wedge \sim r \wedge s) \vee$$
$$\vee (p \wedge \sim q \wedge r \wedge s) \vee (p \wedge \sim q \wedge r \wedge \sim s).$$

Размещая соответствующие значки в карте Карно, получаем

	q, r	q, $\sim r$	$\sim q$, $\sim r$	$\sim q$, r	
p	(1)	(2)	(9)	(10)	s
p	(3)	(4)		(11)	$\sim s$
$\sim p$	(5)	(6)			$\sim s$
$\sim p$	(7)	(8)			s

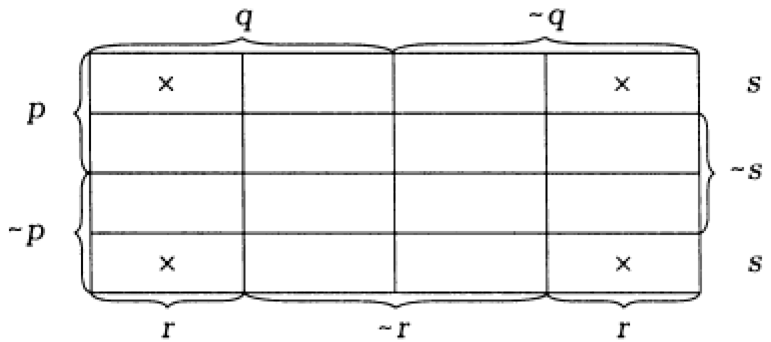
где вместо крестиков мы поместили в ячейки номера минтерма. Например, минтерм (1) – это $p \wedge q \wedge r \wedge s$, а минтерм (10) – это $p \wedge \sim q \wedge r \wedge s$. Все булево выражение может быть представлено в виде дизъюнкции минтермов

$$(1) \vee (2) \vee (3) \vee (4) \vee (5) \vee (6) \vee (7) \vee (8) \vee (9) \vee (10) \vee (11) \equiv$$
$$\equiv ((1) \vee (2) \vee (3) \vee (4) \vee (5) \vee (6) \vee (7) \vee (8)) \vee ((1) \vee (2) \vee (9) \vee (10)) \vee ((10) \vee (11))$$

Здесь во второй строке мы повторили минтермы (1), (2) и (10) и перегруппировали операнды. Очевидно, от этого булево выражение не изменилось. Но теперь видно, что минтермы (1) – (8) образуют восьмиячеечный блок (назовем его A), минтермы (1), (2), (9), (10) образуют четырехъячеечный блок (назовем его B) и минтермы (10) и (11) образуют двухъячеечный блок C. Тогда все булево выражение можно представить в виде $A \vee B \vee C$.

Для каждого блока нужно записать булево выражение. Восьмиячеечный блок представляется только одним из аргументов булевой функции. В данном случае блок A представляется переменной q. Четырехъячеечный блок можно описать, используя только две булевых переменные. В нашем случае четырехъячеечный блок B, составленный из ячеек верхней строки, можно описать выражением $p \wedge s$. Двухъячеечный блок можно описать, используя три булевых переменные. В нашем случае для блока C, составленного из двух правых верхних ячеек, соответствующее выражение имеет вид $p \wedge \sim q \wedge r$. Следовательно, исходное высказывание можно упростить и привести к виду $q \vee (p \wedge s) \vee (p \wedge \sim q \wedge r)$. Для последних двух скобок имеем $(p \wedge s) \vee (p \wedge \sim q \wedge r) \equiv p \wedge (s \vee (\sim q \wedge r))$ и все выражение станет еще на один символ короче $q \vee (p \wedge (s \vee (\sim q \wedge r)))$.

Карту Карно для булевой функции четырех аргументов можно рассматривать как «скрученную» в обоих направлениях, т.е. следует считать, что первый и последний столбцы соседствуют, также как и первая и последняя строка прилегают друг к другу. Тогда, например, четырехъячеечный блок на следующей карте Карно

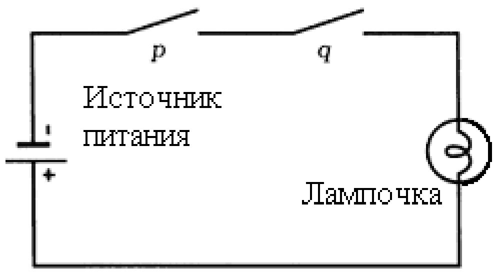


можно описать посредством $r \wedge s$.

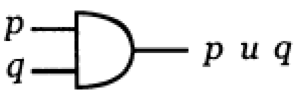
Наглядный способ поиска минтермов для группировки, который предоставляет карта Карно, можно обобщить на булевы функции пяти и даже шести переменных. Однако, возникающие при этом «трехмерные» диаграммы, вносят в метод дополнительные осложнения и делают его малопроизводительным.

5.3 Функциональные схемы

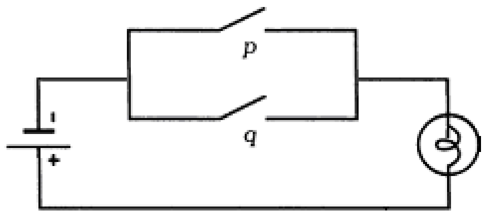
В 1938 г. Клод Шеннон заметил связь между таблицами истинности и электрическими цепями. Рассмотрим схему переключения, приведенную на следующем рисунке



Она состоит из источника питания, двух переключателей и электрической лампочки. Присвоим значение 1 переключателям p и q, если они замкнуты, т.е. электрический ток проходит через них. Если переключатель разомкнут, то считаем, что переключателю (точнее булевой переменной) присвоено значение ноль. Присвоим значение 1 схеме, когда лампочка светится (т.е. электрический ток через нее проходит). При последовательном соединении элементов цепи p и q лампочка загорается, и значение схемы становится равным 1 только в случае, когда оба переключателя замкнуты, т.е. и p и q имеют значение 1. Таким образом, схема соответствует булеву выражению $p \wedge q$ (или $p \cdot q$). В этом параграфе удобнее использовать обозначения логических операций $+, \cdot$ и ' (штрих). Такое расположение переключателей называется логическим элементом p И q, или схемой логического умножения. Этот логический элемент обозначается следующим символом



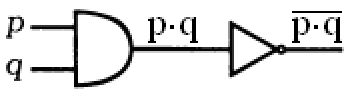
Теперь рассмотрим схему переключения, показанную на следующем рисунке, где переключатели p и q соединены параллельно.



Отметим, что теперь лампочка загорается, и значение схемы становится равным 1, когда один (или оба) из двух переключателей p или q замкнут, т.е. либо значение p = 1, либо q = 1 (либо оба они равны 1). Эта схема соответствует высказыванию p + q и называется логическим элементом p ИЛИ q, или схемой логического сложения. Этот логический элемент обозначается следующим символом

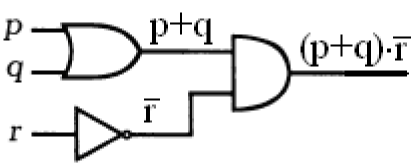
Предположим, имеется схема, с одним переключателем p, который обладает таким свойством, что лампочка загорается тогда и только тогда, когда p разомкнут. Следовательно, схема имеет значение 1, когда p имеет значение 0, и имеет значение 0, когда p имеет значение 1. Эта схема соответствует отрицанию p, т.е. p', а соответствующий элемент называется логическим элементом не, или инвертором. Логический элемент не обозначается следующим символом

Пример 1. Схема на следующем рисунке



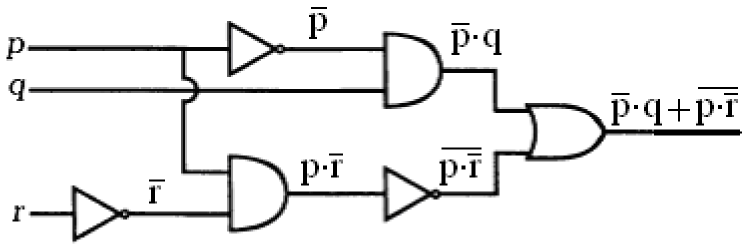
содержит логический элемент p и q, за которым следует инвертор, так что схема соответствует выражению $(p \cdot q)'$. Заметим, что инвертор отрицает всю предшествующую ему схему.

Пример 2. Схема на следующем рисунке

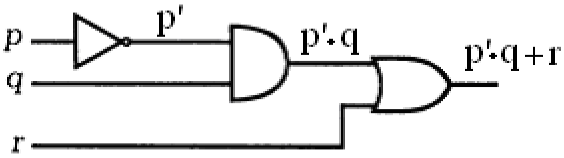


содержит соединение логического элемента p или q с логическим элементом не r посредством логической схемы умножения. Следовательно, она соответствует выражению $(p+q)\cdot r'$.

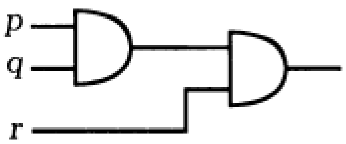
Пример 3. Булево выражение, соответствующее схеме на следующем рисунке, имеет вид $(p'\cdot q)+(p\cdot r')'$

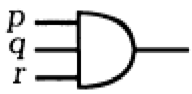


Пример 4. Схема, соответствующая выражению $(p'\cdot q)+r$, показана на следующем рисунке.

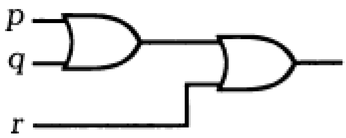


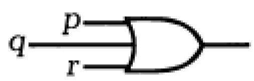
Пример 5. Построим схему трехклавишного переключателя, при помощи которого свет включается тремя различными двухпозиционными переключателями. Для этого построим соответствующее ему булево выражение. Свет должен включаться, когда все три переключателя замкнуты, т.е. необходимо иметь $p\cdot q\cdot r$. Если один из переключателей разомкнут, то свет должен быть выключен. Однако, если разомкнуть другой переключатель, то свет должен включиться. Следовательно, искомое выражение имеет вид $p\cdot q\cdot r+p\cdot q'\cdot r'+p'\cdot q'\cdot r+p'\cdot q\cdot r'$. Для простоты, вместо схемы, представленной на следующем рисунке слева, для выражения $p\cdot q\cdot r$ будем использовать схему, изображенную на рисунке справа.



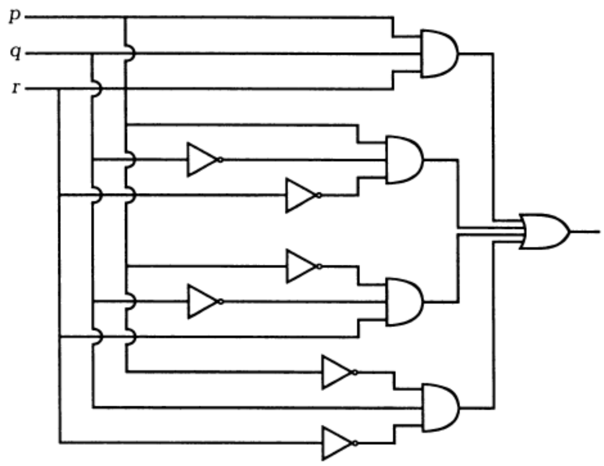


А для выражения p+q+r вместо схемы, показанной на следующем рисунке слева, мы будем использовать схему, показанную справа.



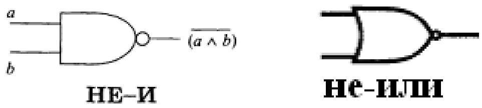


Аналогично будем изображать схемы с большим количеством входных сигналов, т.е. правые схемы на предыдущих рисунках могут иметь по четыре и более входных линий, и будут соответствовать левым схемам с однотипными элементами (только И или только ИЛИ). Используя такие сокращающие значки, искомая схема примет вид

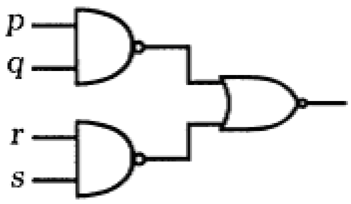


□

Ранее было отмечено, что штрих Шеффера, обозначаемый |, имеет ту же таблицу истинности, что и $(pq)'$, поэтому мы и упоминали его как логическую связку `не–и`. В свою очередь, стрелка Пирса $\downarrow$, имеет ту же самую таблицу истинности, что и $(p+q)'$, поэтому она упоминалась как связка `не–или`. Логические элементы `не–и` и `не–или` обозначаются символами, показанными на следующем рисунке, соответственно слева и справа

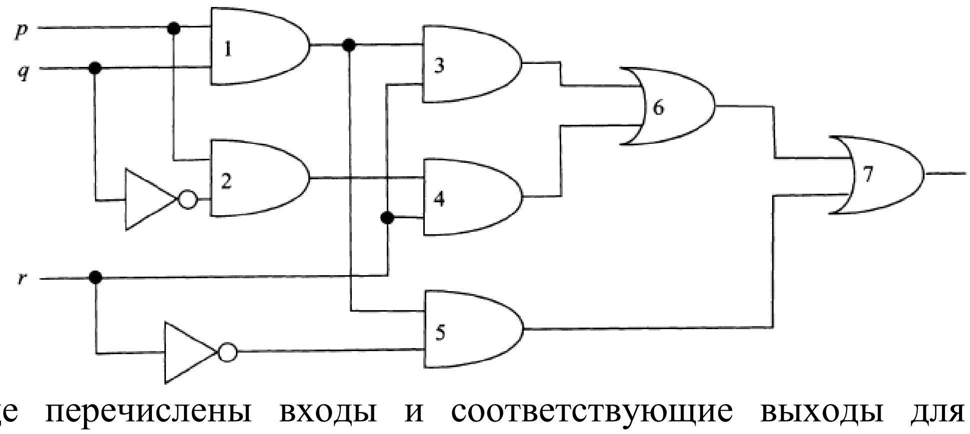


Тогда, например, выражению $(p\,|\,q)\downarrow(p\,|\,r)$ соответствует следующая схема



Таким образом, соединяя функциональные элементы вместе, мы получаем функциональную схему. С ее помощью можно реализовать любую булеву функцию.

Пример 6. Что получится на выходе функциональной схемы, представленной на рисунке



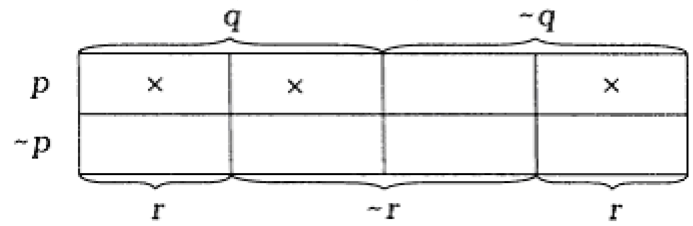
В таблице перечислены входы и соответствующие выходы для каждого функционального элемента в соответствии с нумерацией из рисунка

Вентиль	Вход	Выход
1	$p,\ q$	pq
2	$p,\ \bar q$	$p\bar q$
3	$pq,\ r$	pqr
4	$p\bar q, r$	$p\bar q r$
5	$pq,\ \bar r$	$pq\bar r$
6	$pqr,\ p\bar q r$	$pqr \vee p\bar q r$
7	$pqr \vee p\bar q r,\ pq\bar r$	$pqr \vee p\bar q r \vee pq\bar r$

Таким образом, на выходе схемы получится функция $pqr \vee p\bar{q}r \vee pq\bar{r}$.

Как мы видели, диаграммы функциональных схем можно упростить, если разрешить функциональным элементам И и ИЛИ иметь не по два входа, а больше. Но более впечатляющего упрощения можно добиться, если привлечь карту Карно для преобразования функции, полученной на выходе сложной схемы.

Упростим булеву функцию, генерируемую нашей схемой, и найдем более простую функциональную схему, ее реализующую. Карта Карно требуемого выражения представлена на следующем рисунке.



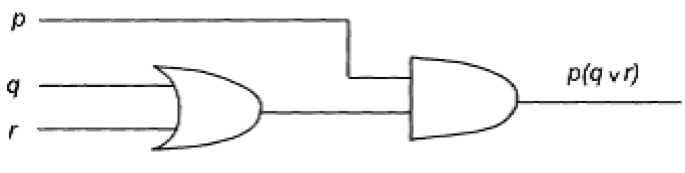
Она имеет две пары минтермов для группировки. Итак,

$$pqr \vee pq\bar r = pq(r \vee \bar r) = pq$$

и

$$pqr \vee p\bar q r = (q \vee \bar q)pr = pr.$$

Это сводит функцию к выражению $pq \vee pr$, которое ввиду дистрибутивности, сводится к функции $p \cdot (q \vee r)$. Более простая схема, реализующая булеву функцию, показана на следующем рисунке



□

При вычерчивании функциональных схем нет необходимости использовать все типы функциональных элементов. Как мы уже видели, множество {V, ~} является полной системой функций. Поэтому мы можем построить любую схему, ограничившись функциональными элементами И и НЕ. Более того, если по той или иной причине нам неудобно использовать большое число компонент, мы могли бы использовать только функциональный элемент НЕ–И.

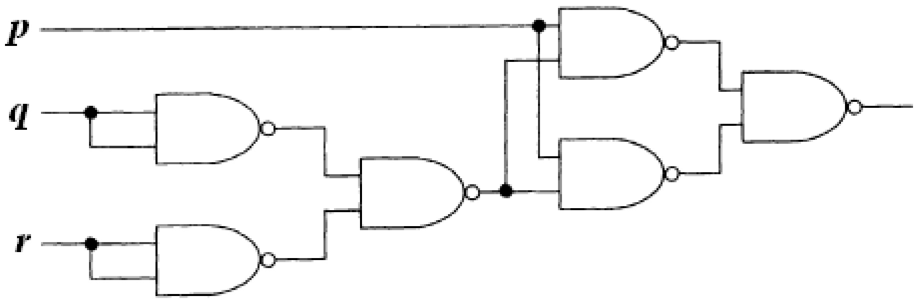
Пример 7. Начертим функциональную схему, реализующую булеву функцию $p(q \vee r)$, используя только НЕ–И. Заметим, что

$$p(q \vee r) = (p \mid (q \vee r)) \mid (p \mid (q \vee r))$$

А также,

$$q \vee r = (q \mid q) \mid (r \mid r)$$

Искомая схема показана на следующем рисунке



5.4 Системы счисления

Со школьной скамьи, имея дело с числами, мы привыкли пользоваться одной системой счисления десятичной. Название *десятичная* объясняется тем, что в основе этой системы лежит основание десять. Это означает, что в десятичной системе любое число выражается упорядоченной последовательностью десяти разных цифр 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. Десятичная система является позиционной, так как значение цифры в записи десятичного числа зависит от ее позиции или местоположения в числе. Например, в десятичном числе 2749 цифра 9 выражает количество единиц, цифра 4 количество десятков, цифра 7 - количество сотен и цифра 2 - количество тысяч. При этом цифра 2 имеет наибольший вес и называется старшей цифрой числа, а цифра 9 - наименьший вес и называется младшей цифрой этого же числа. Различие весов цифр в числе 2749 становится очевидным, если это число записать в виде суммы $2 \cdot 10^3 + 7 \cdot 10^2 + 4 \cdot 10^1 + 9 \cdot 10^0$, в которой число 10 основание системы счисления. Для каждой цифры числа основание 10 возводится в степень, зависящую от позиции цифры, и умножается на эту цифру. Степень основания для единиц равна нулю, для десятков - единице, для сотен - двум и т.д. Если десятичное число дробное, то оно тоже легко записывается в виде суммы, в которой степень основания для каждой цифры дробной части отрицательна и равна –1 для старшей цифры дробной части, –2 для следующей цифры дробной части и т.д. Например, десятичное число 384.9506 выразится суммой $3 \cdot 10^2 + 8 \cdot 10^1 + 4 \cdot 10^0 + 9 \cdot 10^{-1} + 5 \cdot 10^{-2} + 0 \cdot 10^{-3} + 6 \cdot 10^{-4}$. Таким образом, вес любой цифры десятичного числа представляет собой определенную целую степень десяти, а значение степени диктуется позицией соответствующей цифры. В качестве начала счета позиций используется положение десятичной запятой (в компьютерной литературе целая часть отделяется от дробной точкой, так же как и в англоязычных математических источниках). При этом при счете позиций цифр от десятичной запятой налево степень десяти принимает последовательно значения 0, 1, 2, 3 и т.д., а при счете направо – значения –1, –2 , –3 и т.д.

В компьютерах применяется, как правило, не десятичная, а позиционная ***двоичная*** система счисления, т.е. система счисления с основанием 2. В этой системе любое число записывается с помощью двух цифр 0 и 1 и называется двоичным числом. Для того чтобы отличить двоичное число от десятичного числа, содержащего только цифры 0 и 1, в записи двоичного числа нередко добавляется признак двоичной системы счисления, например, 110101.111_2. Каждый разряд (цифру) двоичного числа называют битом. Важное достоинство двоичной системы – удобство физического представления цифр. Например, при передаче информации, цифра 1 может соответствовать наличию электрического напряжения в проводнике, а цифра 0 – отсутствию напряжения. При хранении информации цифра 1 или 0 может соответствовать наличию или отсутствию заряда на конденсаторе или другом электронном элементе. Это приводит к относительной простоте аппаратуры компьютера, в частности арифметико – логического устройства, предназначенного для выполнения арифметических и логических операций над двоичными числами. Как и десятичное число, любое двоичное число можно записать в виде суммы, явно отражающей различие весов цифр, входящих в двоичное число 2. Например, двоичное число 110101.101 в форме суммы имеет вид $1{\cdot}2^5 + 1{\cdot}2^4 + 0{\cdot}2^3 + 1{\cdot}2^2 + 0{\cdot}2^1 + 1{\cdot}2^0 + 1{\cdot}2^{-1} + 0{\cdot}2^{-2} + 1{\cdot}2^{-3}$. Эта сумма записывается по тем же правилам, что и сумма для десятичного числа. В данном примере двоичное число имеет целую шестизначную и дробную трехзначную части. Поэтому старшая цифра целой части, т.е. единица, умножается на $2^{6-1}=2^5$, а следующая цифра целой части, тоже равная единице, умножается на 2^4 и т.д. по убывающим степеням двойки до младшей, третьей, цифры дробной части, которая будет умножена на 2^{-3}. Выполняя в этой сумме арифметические операции по правилам десятичной системы, получим десятичное число 53.625. Таким образом, двоичное число 110101.101 совпадает с десятичным числом 53.625, или $110101.101_2=53.625_{10}$.

Существенным недостатком двоичной системы является то, что для записи числа в этой системе требуется довольно много цифр – нулей и единиц. Это затрудняет восприятие двоичных чисел человеком. Например, десятичное число 173 в двоичной системе имеет вид 10101101. Поэтому двоичную систему применяют, как правило, для *внутренних нужд* компьютера, а для целей коммуникации человека с компьютером выбирают систему счисления с большим основанием. При этом часто используются восьмеричная и шестнадцатеричная системы, поскольку, как будет показано дальше, между этими двумя системами и двоичной системой существует простая связь, облегчающая перевод чисел из одной системы в другую. В **восьмеричной** системе, т.е. системе счисления с основанием 8, числа выражаются с помощью восьми цифр: 0, 1, 2, 3, 4, 5, 6, 7. Например, в восьмеричном числе 257 есть семь единиц, пять восьмерок и две восьмерки в квадрате, т.е. $257_8=2{\cdot}8^2+5{\cdot}8^1+7{\cdot}8^0$, где индекс «8» у числа 257 означает систему счисления. Выполняя в записанной сумме арифметические действия по правилам десятичной системы, получим, что $257_8=175_{10}$, т.е. восьмеричное число 257 совпадает с десятичным числом 175.

В **шестнадцатеричной системе** для записи чисел используются цифры 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 и прописные латинские буквы A, B, C, D, E, F, имеющие значения десятичных чисел 10, 11, 12, 13, 14 и 15 соответственно. Таким образом, «цифрами» шестнадцатеричной системы являются все цифры десятичной системы и, кроме того, шесть латинских букв. Поэтому шестнадцатеричное число может иметь, например, вид 3E5A1. Расписывая это число суммой с учетом основания 16, получим $3E5A1_{16}=3\cdot16^4+E\cdot16^3+5\cdot16^2+A\cdot16^1+1\cdot16^0$. Выполняя арифметические операции по правилам десятичной системы и, принимая во внимание, что A=10, E=14, получим $3E5A1_h=255393_{10}$.

ПЕРЕВОД ЧИСЕЛ ИЗ ОДНОЙ СИСТЕМЫ В ДРУГУЮ. Нередко возникает необходимость переводить числа из одной системы в другую. Перевод чисел из двоичной, восьмеричной или шестнадцатеричной системе десятичную систему состоит в последовательном разложении числа в сумму по степеням основания системы счисления. Для перевода целых чисел из десятичной системы, например, в восьмеричную можно применить следующий прием. Преобразуемое число делят на 8 по правилам десятичной системы, с запоминанием остатка, который, конечно, не превышает 7. Если полученное частное больше 7,его тоже делят на 8,сохраняя остаток. Новое частное, если оно больше 0, в свою очередь делят на 8. Этот процесс деления на 8 продолжается до тех пор, пока полученное частное не станет равно нулю. Затем выписывают подряд все остатки, начиная с последнего. Это и будет результирующее восьмеричное число.

Пример. Выполняя описанный процесс, переведем число 891 из десятичной системы в восьмеричную систему счисления. Решение. 891/8=111, остаток 3; 111/8=13, остаток 7; 13/8=1, остаток 5; 1/8=0, остаток 1 (старшая цифра восьмеричного числа). Ответ: $891_{10}=1573_8$.

Аналогично преобразуют десятичное число в шестнадцатеричное с той лишь разницей, что это число вместо 8 делят на 16.

Пример. Число 891 перевести из десятичной системы в шестнадцатеричную систему счисления. Имеем 891/16=55, остаток 11, т.е. «цифра» B. Далее 55/16=3, остаток 7; 3/16=0, остаток 3 (старшая цифра шестнадцатеричного числа). Ответ: $891_{10}=37B_{16}$.

При переводе десятичного числа в двоичное нужно делить на 2.

Пример. Число 891 перевести из десятичной системы в двоичную систему счисления. Имеем 891/2=445, остаток 1; 445/2=222, остаток 1; 222/2=111, остаток 0; 111/2=55, остаток 1; 55/2=27, остаток 1; 27/2=13, остаток 1; 13/2=6, остаток 0; 3/2=1; остаток 1; 1/2=0; остаток 1 (старшая цифра двоичного числа). Т.о. $891_{10}=110111011_2$.

Весьма прост процесс преобразования двоичного числа в восьмеричное или шестнадцатеричное число. Известно, что для записи любой цифры восьмеричного числа необходимы три двоичные цифры. Поэтому преобразуемое двоичное число разделяют справа налево на группы по три двоичные цифры; при этом самая левая группа может содержать меньше трех двоичных цифр. Затем каждую группу двоичных цифр выражают в виде восьмеричной цифры с учетом того, что $000_2=0_8$, $001_2=1_8$, $010_2=2_8$, $011_2=3_8$, $100_2=4_8$, $101_2=5_8$, $110_2=6_8$ и $111_2=7_8$. Например, двоичное число 1101111011, разбитое на группы по три двоичные цифры, можно записать как 1 101 111 011 и затем, после записи каждой группы одной восьмеричной цифрой, получить восьмеричное число 1573. Аналогично преобразуют двоичное число в шестнадцатеричное с той лишь разницей, что преобразуемое двоичное число делят на группы по четыре двоичных цифры в каждой, поскольку для записи любой цифры шестнадцатеричного числа необходимы четыре двоичных цифры. Поэтому двоичное число 1101111011 , использованное в предыдущем примере, после разбиения на группы по четыре двоичных цифры, можно записать как 11 0111 1011 и, после выражения каждой группы одной шестнадцатеричной цифрой получить шестнадцатеричное число 37B. При этом используем тот факт, что $0000_2 = 0_{16}$, $0001_2 = 1_{16}$, $0010_2 = 2_{16}$, $0011_2 = 3_{16}$, $0100_2 = 4_{16}$, $0101_2 = 5_{16}$, $0110_2 = 16_{16}$, $0111_2 = 7_{16}$, $1000_2 = 8_{16}$, $1001_2 = 9_{16}$, $1010_2 = A_{16}$, $1011_2 = B_{16}$, $1100_2 = C_{16}$, $1101_2 = D_{16}$, $1110_2 = E_{16}$, $1111_2 = F_{16}$. Эти тождества легко проверить. Например, $1011_2 = 1\cdot2^3+0\cdot2^2+1\cdot2^1+1\cdot2^0= 11_{10} = B_{16}$.

Преобразование восьмеричного или шестнадцатеричного числа в двоичное число осуществляется простым переводом каждой цифры исходного числа в группу из трех (для восьмеричного числа) или четырех (для шестнадцатеричного числа) двоичных цифр. Например, 123_8= 001 010 011=1010011_2, $A17_{16}$=1010 0001 0111=101000010111_2.

Во всех компьютерах число двоичных разрядов для записи чисел фиксировано и обычно равно 8, 16, 32 или 64. Это количество двоичных разрядов называется разрядностью или длиной разрядной сетки компьютера. Например, в первых персональных компьютерах IBM целые числа занимали 16 двоичных разрядов. При этом старшие разряды небольших положительных чисел заполняются нулями. Например, число 45_{10}= 101101_2 в памяти такого компьютера, занимая 2 байта, имеет вид 00000000 00101101. Для упрощения записи в этом параграфе мы будем использовать 8 бит для чисел в двоичной форме и 16 бит для чисел в шестнадцатеричной форме. Т.о. для представления целого положительного числа с помощью 8 бит перед ним при необходимости дописывается нужное количество нулей.

Числа со знаком. В современных компьютерах для представления отрицательных чисел используется методом двоичного дополнения. Он состоит в следующем. Берем модуль отрицательного целого числа (т.е. положительное число), записываем его в двоичном виде так, как описано выше. Затем каждый ноль заменяем единицей, а единицу – нулем и к полученному двоичному числу прибавляем единицу. Целое положительное число всегда будет начинаться с 0, а целое отрицательное число будет начинаться с 1. Поэтому 10110011 будет представлять целое отрицательное число, а 01110101 будет представлять целое положительное число (при восьмиразрядном представлении). Поскольку при таком способе записи наибольшее возможное целое положительное число есть 01111111, то всего имеется 127 целых положительных чисел.

Если дано целое положительное число, представленное с использованием n бит, то соответствующее ему отрицательное число находят путем вычитания этого числа из 2^n, выраженного в двоичной форме. Это осуществляется вычитанием числа из $2^n - 1$ (это максимальное двоичное число $\underbrace{11...11}_{n}$, которое имеет n разрядов) и последующим прибавлением 1. Поэтому в случае, когда используется 8 бит, положительное целое число вычитается из

$$2^8 - 1 = 11111111,$$

а затем прибавляется 1. Например, чтобы найти отрицание числа 00101011, мы вычитаем его из 11111111:

$$\begin{array}{r} 11111111 \\ -\underline{00101011} \\ 11010100 \end{array},$$

а затем к полученному числу 11010100 прибавляем 1. Таким образом, отрицанием числа $00101011_2=43_{10}$ будет двоичное число 11010101, которое, таким образом, представляет в памяти число -43_{10}. Заметим, что вычитание 00101011 из 11111111 просто меняет нули на единицы, а единицы — на нули.

Пример. Найти представление числа – 56 в 8-битовом двоичном дополнительном коде. Сначала число 56 представляется как 00111000. Вычитая его из 11111111, получаем 11000111. Прибавление 1 дает 11001000, так что – 56 имеет представление 11001000.

Если задано целое отрицательное число, можно найти целое положительное число, для которого оно является отрицанием, просто выполняя те же самые операции с отрицательным числом. Например, чтобы найти то положительное число, для которого 10110111 представляет отрицание, просто вычтем его из 11111111, получая 01001000. После прибавления 1 получаем 01001001, что является двоичным представлением числа 73. Отсюда 10110111 изображает – 73.

Отрицательные двоичные дополнения изменяются от 11111111 ($= -1_{10}$) до 10000000 ($= -128_{10}$), так что их имеется 128. Таким образом, если использовать 8 двоичных разрядов, можно представить все числа от –128 до 127. В общем случае, если имеется n двоичных разрядов, можно представить все целые числа от -2^{n-1} до $2^{n-1}-1$.

Если для представления целых чисел используется 16 (или больше) бит, то удобнее всего пользоваться шестнадцатеричной системой счисления. Поскольку каждый шестнадцатеричный разряд состоит из четырех двоичных, то для 16 бит необходимо 4 шестнадцатеричных разряда. Таким образом, 21A5, F9AB и 004F являются примерами шестнадцатиразрядных целых чисел со знаком, представленных в шестнадцатеричной системе счисления. Поскольку двоичное представление целого положительного числа должно начинаться с 0, то наибольшее число, которое может стоять в первом разряде его шестнадцатеричного представления, это 7 ($= 0111_2$). Поэтому 691F, 7FFF и 0031 изображают целые положительные числа, а 8124, A105 и FFF1 изображают целые отрицательные числа. Как и ранее, если целое положительное число представлено с использованием n битов, отрицание числа находят путем его вычитания из 2^n. В случае использования 16 битов это осуществляется путем вычитания целого положительного числа в шестнадцатеричной форме из FFFF и последующего прибавления 1 (или, что тоже, вычитанием из 10000_{16}). Как и раньше, для преобразования отрицательного числа обратно в его положительного двойника выполняем операцию дополнения, вычитая его из $2^{16} - 1 = FFFF$, а затем прибавляя 1.

Пример. Найти представление в шестнадцатеричной системе счисления отрицания шестнадцатеричного числа 78F с использованием 16 бит. Сначала заменим запись числа на 16–битовое шестнадцатеричное число 078F. Затем вычтем 078F из FFFF, получив F870, и далее прибавляем 1. В итоге получаем представление F871.

Пример. Найти 16-битовое представление в шестнадцатеричной форме числа – 1158_{10}. Поскольку $1158_{10}=486_{16}=0486$, представленное в 16-битовой шестнадцатеричной форме, вычитаем 0486 из FFFF, получая FB79, после чего, прибавляя 1, получаем FB7A. Следовательно, 16-битовое шестнадцатеричное представление числа – 1158_{10} есть FB7A..

Арифметические действия в двоичной, восьмеричной и шестнадцатеричной системах счисления выполняют по тем же правилам, как и в десятичной системе, с той лишь разницей, что основание системы счисления равно двум, восьми или шестнадцати. Применяя эти правила, можно убедиться в справедливости суммирования двух двоичных чисел:

$$\begin{array}{r} 10111 \\ +\ \underline{10110} \\ 101101 \end{array} \quad \text{или} \quad \begin{array}{r} 23 \\ +\ \underline{22} \\ 45 \end{array}$$

Справа мы выполнили сложения для соответствующих десятичных чисел.

Сложение чисел фиксированной длины со знаком на основе двоичного или шестнадцатеричного представления ничем не отличается от обычного сложения двоичных и шестнадцатеричных чисел, за исключением того, что если сумма порождает дополнительный (девятый) разряд, то этот разряд удаляется.

Например, вычтем из двоичного числа 10111 двоичное число 1101. Для простоты будем считать, что имеем 8 двоичных разрядов для хранения. Припишем спереди к вычитаемому нули, получим 00001101. Заменим в вычитаемом нули единицами и единицы нулями, получим 11110010. К этому числу прибавим единицу, получим 11110011 - дополнительный код вычитаемого. Теперь сложим уменьшаемое с дополнительным кодом вычитаемого, игнорируя единицу переноса (переполнения) из старшего разряда:

$$\begin{array}{r} 00010111 \\ + \\ \underline{11110011} \\ [1]\ 00001010 \end{array}$$

Таким образом, получено двоичное число 00001010, совпадающее с десятичным числом 10. Так как десятичное значение уменьшаемого равно 23, а вычитаемого – 13, то получен верный результат.

Если шестнадцатеричные числа со знаком FF14 и 0F1A складываются как шестнадцатеричные:

$$\begin{array}{r} FF14 \\ \underline{0F1A} \\ [1]\,0E2E \end{array},$$

то цифра 1 слева также удаляется. Если попытаться сложить 7F14 и 6F23, то получаем

$$\begin{array}{r} 7F14 \\ \underline{6E23} \\ ED37 \end{array},$$

т.е. сложив два положительных числа, мы получили отрицательное число. В таком случае говорят, что имеет место **переполнение**. Значение суммы слишком велико для представления с использованием 16 бит. Числа со знаком, выражаемые с использованием 16 двоичных разрядов или 4 шестнадцатеричных разрядов, должны находиться между -32768 и 32767.

При вычитании одного целого числа из другого находим двоичное дополнение числа, которое нужно вычесть, и производим сложение. Например, для нахождения разности

$$\begin{array}{r} 01010110 \\ -\ \underline{01110001} \end{array}$$

мы сначала находим двоичное дополнение числа 01110001, которое равно 10001111, после чего выполняем сложение:

$$\begin{array}{r} 01010110 \\ +\ \underline{10001111} \\ 11100101 \end{array}.$$

Как и следовало ожидать, результат — отрицательное число. Используя тот же самый подход при решении следующей задачи:

$$\begin{array}{r} 01010001 \\ -\ \underline{11101001} \end{array},$$

мы сначала находим двоичное дополнение числа 11101001, которое равно 00010111, после чего выполняем сложение:

$$\begin{array}{r} 01010001 \\ +\ \underline{00010111} \\ 01101000 \end{array} \qquad \begin{array}{r} -\ 81_{10} \\ \underline{-23_{10}} \\ 104_{10} \end{array}$$

В данном случае отрицательное число вычиталось из положительного, поэтому отрицательное число мы заменили положительным и произвели сложение.

Пример. Вычесть 56 из 23, используя 8-битовые числа со знаком. Поскольку 56 = 00111000 и 23 = 00010111, имеем

$$\begin{array}{r} 00010111 \\ -\ \underline{00111000} \end{array}.$$

Сначала находим двоичное дополнение числа 00111000, которое равно 11001000, а затем выполняем сложение:

$$\begin{array}{r} 00010111 \\ +\ \underline{11001000} \\ 11011111 \end{array}.$$

Так как двоичным дополнением числа 11011111 является 00100001, которое равно 33, число 11011111 равно -33.

Пример. Вычесть 7328 из 3614, воспользовавшись 16-битовыми шестнадцатеричными числами со знаком. Поскольку $7328_{10} = 1CA0_{16}$ и 3614 = E1E16, имеем

$$\begin{array}{r} 0E1E \\ -\ \underline{1CA0} \end{array}.$$

Сначала находим двоичное дополнение числа 1CA0, которое равно E35F, а затем выполняем сложение:

$$\begin{array}{r} 0E1E \\ +\ \underline{E360} \\ F17E \end{array}.$$

Поскольку двоичным дополнением числа F1E является 0E82 и $0E82 = 3714_{10}$, то 3614 - 7328 = –3714.

5.5 Проектирование сумматоров

Ранее мы рассматривали функциональные схемы с одним выходом, т.е. схемы, которые соответствовали одной булевой функции нескольких булевых аргументов. Но часто требуется построить схемы с несколькими выходами, сигнал на каждом из которых соответствует своей булевой функции, а значения аргументов (сигналов на входе) общие. Важным примером таких схем являются арифметические регистры микропроцессоров. Здесь мы рассмотрим способ конструирования функциональных схем простейших сумматоров – полубитного, одноразрядного, двухбитного, трехбитного и n – битного сумматоров.

Полубитный сумматор. Пусть p и q обозначают двоичные цифры, которые предстоит сложить, а d_1 и d_0 — двоичные цифры суммы, получающейся на выходе сумматора (здесь d_1 будет представлять старший разряд, а d_0 – младший разряды суммы)

Полусумматор находит сумму двух двоичных чисел (единиц и/или нулей) согласно таблице сложения

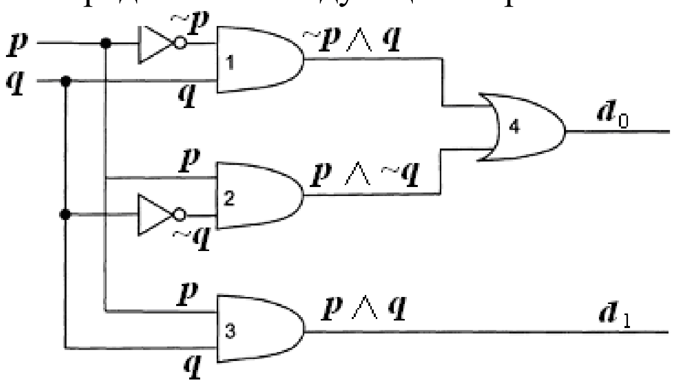
+	0	1
0	00	01
1	01	10

Таблицы истинности для булевых функций d_0 и d_1 должны иметь вид

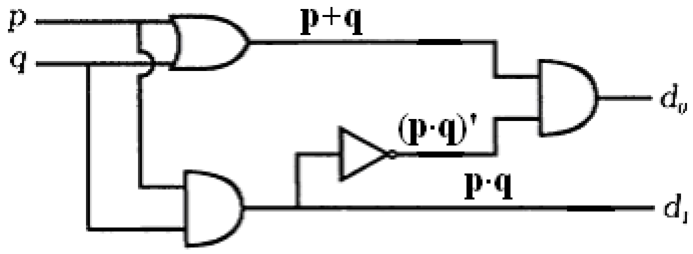
Случай	p	q	d_0
1	1	1	0
2	1	0	1
3	0	1	1
4	0	0	0

Случай	p	q	d_1
1	1	1	1
2	1	0	0
3	0	1	0
4	0	0	0

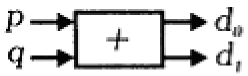
Следовательно, $d_0 \equiv p \cdot q' + p' \cdot q$ и $d_1 = p \cdot q$. Тогда функциональную схему полусумматора можно представить следующим образом



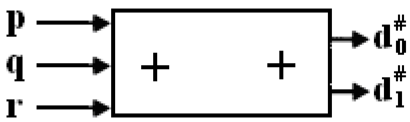
Поскольку $p \cdot q' + p' \cdot q \equiv (p+q) \cdot (p \cdot q)'$ (это проверяется, например, построением таблиц истинности), то для булевых функций $d_0 \equiv (p+q) \cdot (p \cdot q)'$ и $d_1 = p \cdot q$ схему полусумматора можно построить также в следующем виде



Полусумматор часто обозначается символом, изображенным на следующем рисунке



Одноразрядный (полный) двоичный сумматор складывает три двоичных одноразрядных числа. Он производит сложение двух бинарных чисел `p, q` (обычно это `i` – е двоичные разряды многоразрядных двоичных чисел) и учитывает еще перенос `r` из предыдущего разряда (добавляет двоичную цифру). На выходе одноразрядного сумматора выдается сумма $p +_2 q +_2 r = d_0^{\#}$ и еще вычисляется перенос $d_1^{\#}$ в следующий разряд. На схемах его часто обозначают следующим образом



Пусть p,q и r обозначают одноразрядные двоичные числа, которые необходимо сложить, а $d_1^{\#}, d_0^{\#}$ — первый (старший) и второй (младший) разряды их суммы. Построим следующие таблицы истинности:

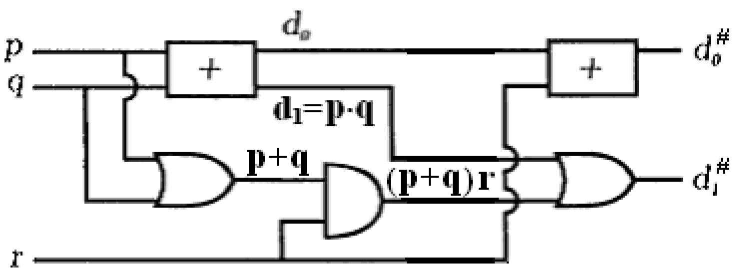
				p+q		(p+q)+r	
Случай	p	q	r	d_1	d_0	$d_1^{\#}$	$d_0^{\#}$
1	1	1	1	1	0	1	1
2	1	1	0	1	0	1	0
3	1	0	1	0	1	1	0
4	1	0	0	0	1	0	1
5	0	1	1	0	1	1	0
6	0	1	0	0	1	0	1
7	0	0	1	0	0	0	1
8	0	0	0	0	0	0	0

Здесь $d_0 \equiv (p+q)\cdot(p\cdot q)'$ младший разряд суммы $p +_2 q$ и $d_1 = p \cdot q$ - старший разряд этой суммы. Тогда $d_0^{\#}$ (младший разряд суммы трех чисел) есть младший разряд результата сложения $d_0 +_2 r$ – это видно из приведенной таблицы.

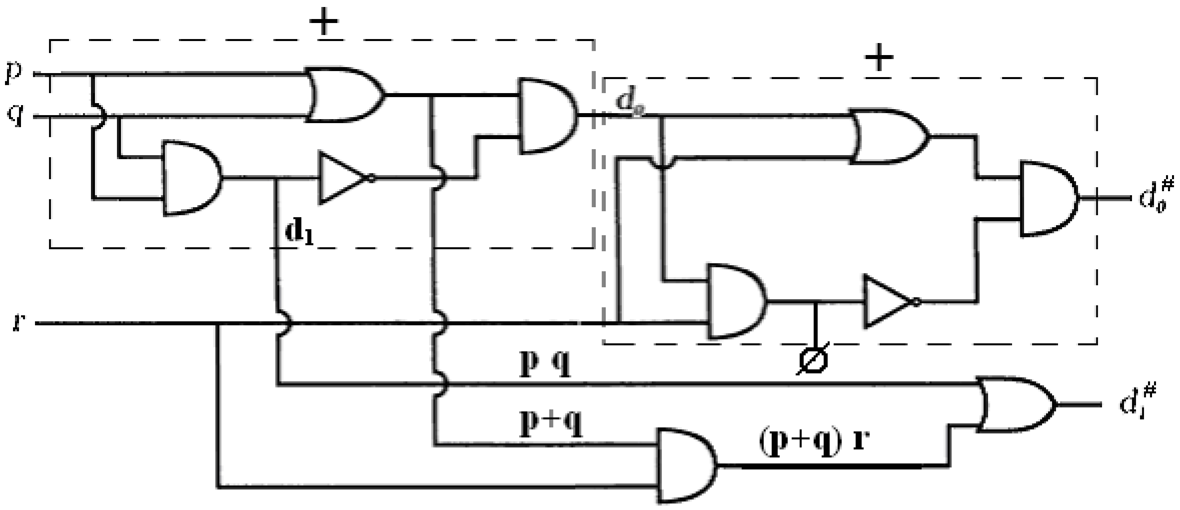
Из приведенной таблицы для $d_1^{\#}$ (старший разряд суммы трех чисел) имеем такое представление в виде дизъюнкции минтермов $d_1^{\#} = pqr + pqr' + pq'r + p'qr$. Для упрощения этого выражения составляем карту Карно.

	q	q	$\sim q$	$\sim q$
p	×	×		×
$\sim p$	×			
	r	$\sim r$	$\sim r$	r

Из нее получаем $d_1^{\#} \equiv p\cdot q + p\cdot r + q\cdot r \equiv p\cdot q + (p+q)r$. Поэтому схема может быть приведена к виду, показанному на следующем рисунке, в который мы включили два полубитных сумматора.



Заметим, что старший разряд (выход) второго полубитного сумматора мы игнорируем. Более подробно эта схема представлена на следующем рисунке



Здесь схемы полусумматоров мы отметили пунктирными прямоугольниками, и символ $\varnothing$ представляет игнорируемый выход второго полусумматора, который, конечно, можно не делать.

Двухбитный сумматор — это устройство, которое вычисляет сумму двух двоичных чисел, выдавая в качестве ответа трехзначное двоичное число. На вход он получает два двузначных двоичных числа, а на выходе у него оказывается трехзначное число, равное сумме вводимых чисел. Иными словами, 2-битный сумматор складывает числа в двоичной системе счисления. Например, $10_2+11_2=101_2$.

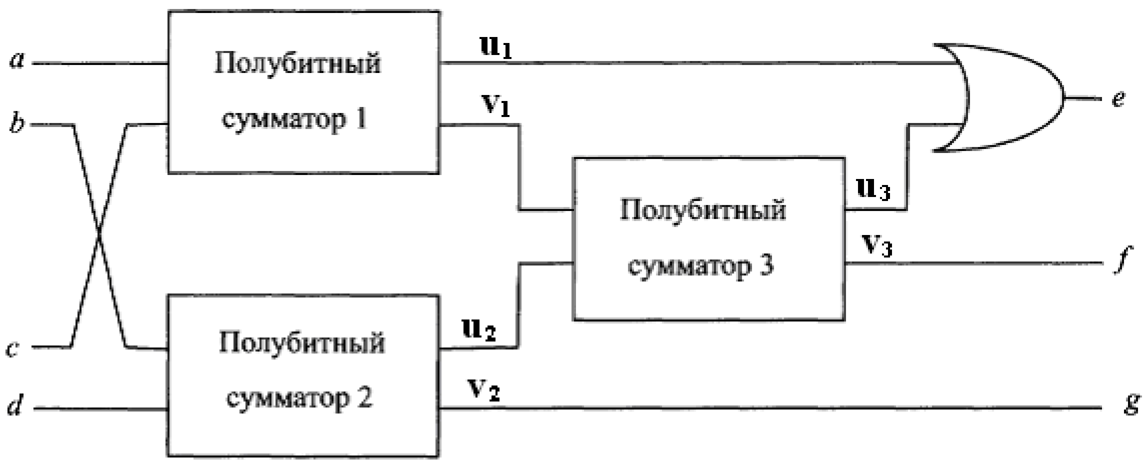
Обозначим через *a* и *b* цифры первого вводимого в сумматор числа, а через *c* и *d* — цифры второго (см. рисунок). Пусть `e,f,g` — цифры вычисляемой суммы.

Составим следующую таблицу истинности

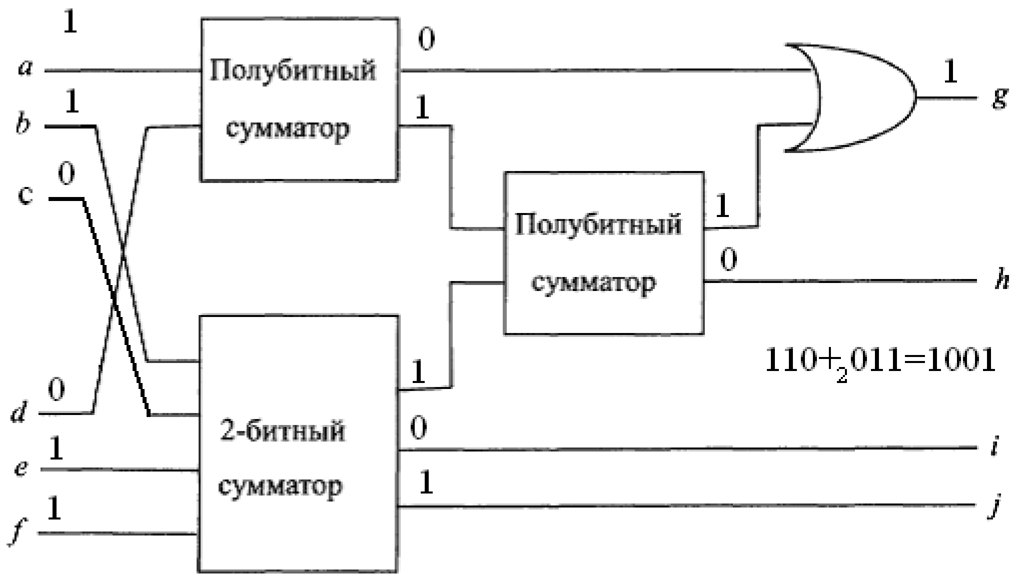
				$ab+_2 cd$			$a+_2 c$		$b+_2 d$		$u_2+_2 v_1$		
a	b	c	d	e	f	g	u_1	v_1	u_2	v_2	u_3	v_3	u_1 V u_3
1	1	1	1	1	1	0	1	0	1	0	0	1	1
1	1	1	0	1	0	1	1	0	0	1	0	0	1
1	1	0	1	1	0	0	0	1	1	0	1	0	1
1	1	0	0	0	1	1	0	1	0	1	0	1	0
1	0	1	1	1	0	1	1	0	0	1	0	0	1
1	0	1	0	1	0	0	1	0	0	0	0	0	1
1	0	0	1	0	1	1	0	1	0	1	0	1	0
1	0	0	0	0	1	0	0	1	0	0	0	1	0
0	1	1	1	1	0	0	0	1	1	0	1	0	1
0	1	1	0	0	1	1	0	1	0	1	0	1	0
0	1	0	1	0	1	0	0	0	1	0	0	1	0
0	1	0	0	0	0	1	0	0	0	1	0	0	0
0	0	1	1	0	1	1	0	1	0	1	0	1	0
0	0	1	0	0	1	0	0	1	0	0	0	1	0
0	0	0	1	0	0	1	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0

Здесь *a*b – первый двоичный операнд и *a* – его старшая двоичная цифра, а b – младшая. Также *cd* – второй двоичный операнд и *c* – его старшая двоичная цифра, а d – младшая. Результат efg=*a*b$+_2$cd, где g, f, e – младшая, средняя и старшая цифры результата суммирования. В таблице также заполнены столбцы для двоичных сумм $a+_2 c$, $b+_2 d$, $u_2+_2 v_1$. Очевидно, что в функциональной схеме они могут быть получены с использованием полубитных сумматоров. Значения последнего столбца u_1 V u_3 могут быть получены с использованием двоичного элемента ИЛИ.

Из таблицы видно, что столбец v_2 совпадает со столбцом g. Поэтому младший разряд результата может быть получен как младший разряд суммы цифр $b+_2 d$. Столбец v_3 совпадает со столбцом f. Поэтому средняя цифра результата может быть получена как сумма $u_2+_2 v_1$. Также из таблицы видно, что старшая цифра суммы может быть получена бинарным сложением $e=u_1 \vee u_3$. Исходя из всего сказанного функциональная схема 2-битного сумматора может быть представлена следующей схемой

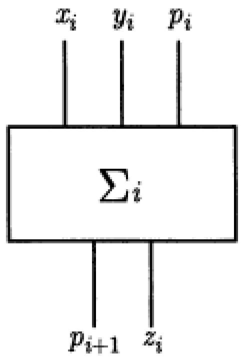


На следующем рисунке представлена функциональная схема **3-битного сумматора**, складывающего два трехзначных двоичных числа с цифрами a,b,c и d,e,f соответственно. В качестве суммы получается четырехзначное число с цифрами g,h,i,j .

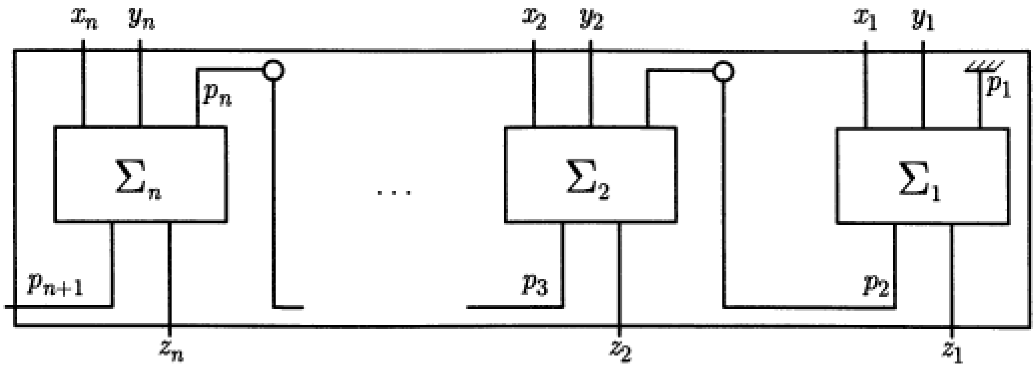


Мы не будем строить таблицы истинности для этого сумматора, а проанализируем его работу на примере сложения двух трехзначных двоичных чисел, например, $110+_2 011=1001$. Для этого мы проставили значения сигналов на предыдущей схеме. В общем случае мы, конечно, должны составить таблицы истинности для булевой функции, соответствующей каждому выходу, и убедиться, что при любых входных значениях (двоичных аргументах) мы получаем двоичные цифры суммы на соответствующих выходах.

n – разрядный двоичный сумматор суммирует два n – разрядных двоичных числа и возвращать их сумму, и еще, возможно, разряд переполнения (переноса). Для построения его схемы мы используем одноразрядный двоичный сумматор. Напомним, что он производит сложение двух двоичных цифр x_i, y_i и учитывает еще перенос p_i из предыдущего разряда. На выходе одноразрядного сумматора выдается сумма $x_i +_2 y_i +_2 p_i = z_i$ и вычисляется перенос p_{i+1} в следующий разряд. На схемах такой сумматор часто обозначается следующим образом



Пусть первое слагаемое **n** – разрядного сумматора имеет вид $x_n x_{n-1} x_{n-2} .. x_2 x_1$, второе – $y_n y_{n-1} y_{n-2} .. y_2 y_1$ и сумма – $z_n z_{n-1} z_{n-2} .. z_2 z_1$, где x_i, y_i, z_i - i-е двоичные цифры (0 или 1) слагаемых и результата. Построим последовательную схему из одноразрядных сумматоров следующего вида



n-Разрядный двоичный сумматор

Здесь $p_1 = 0$, $z_1 = x_1 \oplus y_1$ – младшая двоичная цифра результата сложения младших разрядов слагаемых (символ $\oplus$ обозначает сложение по модулю 2, т.е. младшую двоичную цифру сложения), p_2 – цифра переноса из младшего разряда. Далее $z_2 = x_2 \oplus y_2$ – вторая цифра результата и p_3 – цифра переноса из второго разряда. И так далее. Если значение $p_{n+1}=1$, то это сигнал о переполнении сумматора. Важно, что каждый очередной одноразрядный сумматор вступает в работу только после окончания работы предыдущего одноразрядного сумматора, который вырабатывает перенос. Поэтому на рисунке изображены элементы задержки в виде кружков на входе переноса очередного одноразрядного сумматора. Этот элемент запирает каждый одноразрядный сумматор до тех пор, пока не закончилось суммирование в предыдущем разряде. Синхронизация этого процесса характеризуется тактовой частотой процессора. Быстродействие компьютера зависит от тактовой частоты процессора и от величины разрядности n сумматора.

6. Комбинаторика.

Комбинаторика представляет собой раздел математики, изучающий способы подсчета количества элементов конечных множеств.

6.1 Основные комбинаторные принципы

Решение многих комбинаторных задач основывается на двух фундаментальных правилах, называемых правилами суммы и произведения.

Правило суммы. Если X_1 и X_2 – непересекающиеся конечные множества, содержащие n_1 и n_2 элементов соответственно, то объединение $X_1 \bigcup X_2$ содержит $n_1 + n_2$ элементов.

Сформулированное правило можно распространить на случай произвольного числа слагаемых: если множества $X_1, X_2, \ldots, X_k$ образуют разбиение множества X, то $n = n_1 + n_2 + \ldots + n_k$, где $n=|X|$ – число элементов множества X, а $n_i=|X_i|$ – число элементов (мощность) множества X_i, $i = 1, 2, \ldots, k$. Напомним, что разбиением множества X называется набор его непересекающихся подмножеств, объединение которых дает все исходное множество X.

Правило суммы можно приспособить и для подсчета числа элементов объединения двух множеств с непустым пересечением:

$$|X_1 \cup X_2| = |X_1| + |X_2| - |X_1 \cap X_2|. \qquad (1)$$

В самом деле, множества $X_1 \backslash X_2$ и $X_1 \cap X_2$ образуют разбиение множества X_1, поэтому

$$|X_1| = |X_1 \backslash X_2| + |X_1 \cap X_2|.$$

Аналогично

$$|X_2| = |X_2 \backslash X_1| + |X_1 \cap X_2|.$$

Кроме того, множества $X_1 \backslash X_2$, $X_2 \backslash X_1$ и $X_1 \cap X_2$ образуют разбиение множества $X_1 \cup X_2$, так что

$$|X_1 \cup X_2| = |X_1 \backslash X_2| + |X_2 \backslash X_1| + |X_1 \cap X_2|.$$

Подставляя сюда выражения $|X_1 \setminus X_2| = |X_1| - |X_1 \bigcap X_1|$, и $|X_2 \setminus X_1| = |X_2| - |X_1 \bigcap X_1|$, приходим к утверждению (1).

□

Пример. Найдем количество положительных целых чисел, меньше или раных 1000, которые делятся на 3 или на 5.

Количество элементов множества S положительных целых чисел, меньших 1000, которые делятся на 3, равно $\left\lfloor \frac{1000}{3} \right\rfloor$ или 333. Количество элементов множества T положительных целых чисел, меньших 1000, которые делятся на 5, равно $\left\lfloor \frac{1000}{5} \right\rfloor$ или 200. Элементами множества $S \cap T$ являются целые числа, меньшие 1000, которые делятся на 5 и на 3, и поэтому делятся на 15. Следовательно, $|S \cap T| = \left[\dfrac{1000}{15}\right] = 66$. Значит

$$|S \cup T| = |S| + |T| - |S \cap T| = 333 + 200 - 66 = 467$$

■

Пример. Предположим, что на курсе из 100 студентов 60 человек изучают математику, 75 — историю, а 45 человек — и то, и другое.

а) Сколько студентов изучают математику или историю?

Пусть универсум U — группа из 100 студентов, М — множество студентов, изучающих математику, Н — множество студентов, изучающих историю. Тогда количество студентов, изучающих математику или историю, равно $|M \cup H| = |M| + |H| - |M \cap H| = 60 + 75 - 45 = 90$.

б) Сколько студентов не изучают ни математику, ни историю?

Количество студентов, не изучающих ни математику, ни историю, равно $\left|\overline{M} \cap \overline{H}\right|$. Но $\overline{M} \cap \overline{H} = \overline{M \cup H}$, поэтому

$$\left|\overline{M} \cap \overline{H}\right| = \left|\overline{M \cup H}\right| = 100 - 90 = 10.$$

■

Существует альтернативный метод решения приведенных выше задачи, который является более информативным. Рассмотрим его на следующем примере.

Пример. Предположим, что из 100 студентов курса 50 изучают химию, 53 — математику, 42 — физику, 15 — химию и физику, 20 занимаются физикой и математикой, 25 — математикой и химией и 5 студентов изучают все три предмета.

а) Сколько студентов изучают хотя бы один из трех перечисленных предметов?

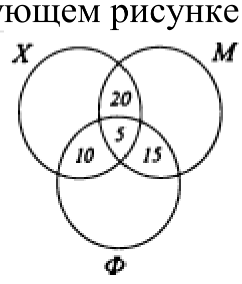
б) Сколько студентов не изучают ни один из трех перечисленных предметов?

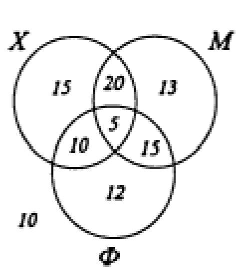
в) Сколько студентов изучают только математику?

г) Сколько студентов изучают физику или химию, но не изучают математику?

д) Сколько студентов не изучают ни математику, ни химию?

Поскольку 5 человек изучают все три предмета, а 15 человек — химию и физику, остаются 10 человек, изучающих химию и физику, но не изучающих математику. Аналогично, 25 — 5 = 20 человек занимаются математикой и химией, но не физикой, и 20 — 5 = 15 человек изучают математику и физику, но не изучают химию. Данную ситуацию изображает диаграмма Эйлера, приведенная на следующем рисунке слева





Поскольку 50 студентов изучают химию и 35 из них уже учтены, то оставшиеся 15 изучают только химию. Аналогично, 53 студента занимаются математикой и

40 из них уже учтены. Поэтому 13 человек изучают только математику. Наконец, 42 студента изучают физику, и 30 из них уже учтены, поэтому 12 человек изучают только физику.

а) Суммируя количество людей, принадлежащих семи непересекающимся подмножествам, получаем 90 тех, кто изучает хотя бы один из трех предметов.

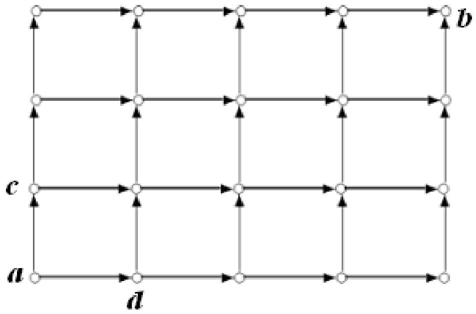
б) Поскольку 90 из 100 студентов изучают хотя бы один предмет, то 100-90 = 10 человек не изучают ни один из этих трех предметов.

в) Из диаграммы Эйлера следует, что 13 человек изучают только математику.

г) Тридцать семь студентов занимаются химией или физикой, но не изучают математику.

д) Из диаграммы Эйлера, изображенной на предыдущем рисунке справа, следует, что 75 человек изучают математику или физику. Поэтому 100 — 75 = 25 студентов не изучают ни математику, ни физику.

Пример. Сколько имеется путей из вершины *a* в вершину *b* в сети, показанной на следующем рисунке (движение возможно только вправо или вверх)?



Обозначим множество всех путей из *a* в *b* через L_{ab} и разобьем его на два непересекающихся подмножества:

L_{acb} – пути, проходящие через вершину c;

L_{adb} – пути, проходящие через вершину d.

Имеем $|L_{ab}| = |L_{acb}| + |L_{adb}|$. Очевидно, что интересующее нас количество путей зависит только от размеров решетки, поэтому обозначим через $l_{m,n}$ количество путей в сети, имеющей m горизонтальных и n вертикальных вершин. Тогда последнее равенство можно записать в виде $l_{4,5} = l_{3,5} + l_{4,4}$. И вообще, $l_{m,n} = l_{m-1,n} + l_{m,n-1}$. Пользуясь этим соотношением, подсчитаем количество путей на рисунке:

$$l_{4,5} = l_{3,5} + l_{4,4} = (l_{2,5} + l_{3,4}) + (l_{3,4} + l_{4,3}) = l_{2,5} + 3 \cdot l_{3,4} =$$
$$= (l_{1,5} + l_{2,4}) + 3(l_{2,4} + l_{3,3}) = l_{1,5} + 4 \cdot l_{2,4} + 3 \cdot l_{3,3} =$$
$$= l_{1,5} + 4\,(l_{1,4} + l_{2,3}) + 3\,(l_{2,3} + l_{3,2}) = l_{1,5} + 4\,l_{1,4} + 10\,l_{2,3} =$$
$$= 1 + 4 + 10 \cdot 3 = 35$$

поскольку $l_{1,5} = 1,\ l_{1,4} = 1, l_{2,3} = 3$.

■

Другое фундаментальное правило дает, доказанная ранее

Теорема. Мощность декартового произведения двух конечных множеств равна произведению их мощностей $|A\times B|=|A|\cdot|B|$.

Обобщение ее на декартово произведение n множеств называется **правилом произведения**. Рассмотрим упорядоченные наборы $(a_1, a_2, \ldots, a_k)$ заданной длины k. Предположим, что элемент a_1 из множества X_1 может быть выбран n_1 способами, т.е. $|X_1|=n_1$. При уже выбранном элементе a_1, элемент a_2 из множества X_2 может быть выбран n_2 способами, т.е. $|X_2|=n_2$. При фиксированных a_1 и a_2 элемент a_3 из множества X_3 может быть выбран n_3 способами, т.е. $|X_3|=n_3$ и т.д. При фиксированных $a_1, a_2, \ldots, a_{k-1}$ элемент a_k из множества X_k может быть выбран n_k способами. Тогда число различных упорядоченных наборов равно произведению $n_1\times n_2\times\ldots\times n_k$. Это означает, что

$$|X_1\times X_2\times\ldots\times X_k| = |X_1|\times|X_2|\times\ldots\times|X_k|.$$

Это правило часто называют комбинаторным принципом умножения.

Доказательство. Будем доказывать теорему методом математической индукции. *Базис индукции*. Пусть n=2. В этом случае все элементы множества X_1 x X_2, т. е. упорядоченные пары (x_1, x_2), можно расположить в виде прямоугольной таблицы со строками — элементами X_1 и столбцами — элементами X_2. В этой таблице, очевидно, будет $|X_1|$ x $|X_2|$ элементов.

Индуктивный переход. Предположим справедливость утверждения теоремы для n. Покажем, что для n+1 оно будет тоже справедливо. В самом деле, добавляя еще одно множество в декартово произведение, видим, что

$$|X_1\times X_2\times\ldots\times X_n\times X_{n+1}| = |(X_1\times X_2\times\ldots\times X_n)\times X_{n+1}| =$$

$$= |X_1\times X_2\times\ldots\times X_n||X_{n+1}| = |X_1||X_2|\ldots|X_n||X_{n+1}|.$$

■

В случае, когда $X_1=X_2=\ldots=X_k$ эта формула принимает вид $|X^k|=|X|^k$. В этом случае множество X называется алфавитом, его элементы – буквами, а элементы декартова произведения X^k, т.е. упорядоченные пары из k букв $(x_1,x_2,\ldots,x_k)$ называют словами в алфавите X. Число k называется длиной слова.

Пусть $|X|=m$, тогда формулу $|X^k|=|X|^k$ можно сформулировать следующим образом:

число слов длины k в алфавите из m букв равно m^k.

Пример. Битовая строка – это строка, состоящая из элементов множества {0, 1}, т.е. каждый из элементов имеет значение 0 или 1. Сколько существует битовых строк длины 5? Сколько существует битовых строк длины k?

Поскольку каждый символ строки может иметь значение 1 или 0, то существует два варианта выбора для каждой позиции. Следовательно, существует 2 x 2 x 2 x 2 x 2 = 2^5 битовых строк длины 5. По аналогичным соображениям, имеется 2^k битовых строк длины k.

Пример. Используя правило произведения и правило суммы, найдем число различных трехзначных чисел, не содержащих одинаковых цифр, и число различных трехзначных чисел, содержащих хотя бы две одинаковые цифры

Пусть a_1, a_2, a_3 – цифры трехзначного числа. Первую цифру a_1 можно выбрать девятью способами (в качестве нее можно взять любую цифру, кроме нуля); при фиксированной первой цифре вторую цифру a_2 можно выбрать также девятью способами (в качестве нее можно взять любую цифру, кроме a_1); наконец, при фиксированных первой и второй цифрах третью цифру можно выбрать восемью способами. По правилу произведения число трехзначных чисел, не содержащих одинаковых цифр, равно $9 \cdot 9 \cdot 8=648$.

Всего имеется 900 различных трехзначных чисел (от 100 до 999). Каждое из них либо содержит две одинаковые цифры, либо нет. Следовательно, имеется 900 – 648 = 252 различных трехзначных числа, содержащих хотя бы две одинаковые цифры.

■

Третьим важным принципом является **правило взаимно однозначного соответствия** (правило биекции или правило равенства): множества А и В имеют одинаковое количество элементов | А | = | В | тогда и только тогда, когда между ними можно установить взаимно однозначное соответствие.

В качестве примера применения этого принципа подсчитаем количество всевозможных подмножеств множества М, состоящего из n элементов. Так это количество не зависит от природы элементов множества М, возьмем М={1,2,…,n}. Произвольному подмножеству $A \subseteq M$ поставим в соответствие двоичное слово $\alpha_1\alpha_2..\alpha_n$ по следующему правилу:

$$\alpha_i = \begin{cases} 1, \text{если } i \in A \\ 0, \text{если } i \notin A \end{cases}$$

Это соответствие взаимно однозначное. Отсюда число всех подмножеств n – элементного множества равно числу двоичных слов длины n, т.е. равно 2^n.

Пример. Сколькими способами можно разложить 10 монет разной стоимости по двум карманам?

Формализуя задачу, отметим, что два кармана дадут два подмножества 10 – элементного множества; объединение множеств монет, лежащих в этих двух карманах, дает все множество; пересечение этих множеств пусто. Это значит, что те монеты, которые лежат в одном кармане, полностью определяют содержимое второго кармана.

Используя правило биекции, можем сказать, что требуемых способов столько, сколько подмножеств в 10 – элементном множестве. Т.о. можем написать ответ: 2^{10}=1024.

■

Напомним также принцип Дирихле, который мы формулировали следующим образом. Пусть f : A —> B — функция, причем как A, так и B — конечные множества. Предположим, что A состоит из n элементов: $a_1, a_2,\dots a_n$. Принцип Дирихле гласит, что если |A|>|B|, то по крайней мере одно значение f встретится более одного раза. Проще говоря, найдется пара элементов $a_i \neq a_j$, для которых $f(a_i) = f(a_j)$.

Пример. Показать, что если в прямоугольнике со сторонами 6 и 8 сантиметров помещены пять точек, то существуют две точки, расстояние между которыми не более 5 сантиметров.

Разделим исходный прямоугольник на четыре прямоугольника, размером 3 на 4 сантиметра каждый. Поскольку пять точек должны находиться либо внутри, либо на границах четырех прямоугольников, то хотя бы две точки должны быть либо внутри, либо на границе одного и того же прямоугольника размера 3 x 4. Но любые такие точки находятся на расстоянии не более 5 сантиметров.

6.2 Модельные схемы

Задачи, ответы к которым часто используются, называют модельными. Они часто касаются конкретных вещей – карточных наборов, размещение шаров в урнах, множества дорог на карте и т.д. Поскольку природа множества при подсчете количества его элементов несущественна, то обычно ограничиваются объектами какой-нибудь одинаковой природы. Ответы к модельным задачам называются комбинаторными числами. Используя взаимно однозначное соответствие между элементами исследуемого множества и элементами модельного множества, устанавливается количество элементов того множества, которое требуется сосчитать.

Дадим некоторые определения. Пусть задано множество X. Последовательная запись нескольких элементов из X, в которой каждый элемент может встретиться только один раз и порядок записи элементов существенный, называется упорядоченным подмножеством в X.

Пусть $X = \{ x_1, x_2, ..., x_n \}$ – конечное множество. Любой упорядоченный набор $(x_{i_1}, x_{i_2}, ..., x_{i_n})$ из n различных элементов множества X называется перестановкой элементов множества X.

Любое упорядоченное подмножество $(x_{i_1}, x_{i_2}, ..., x_{i_m})$ из m различных элементов множества X $(m \leq n)$ называется размещением из n элементов по m. В частности, для n=m размещение будет перестановкой. Количество m – элементных упорядоченных подмножеств в n – элементном множестве обозначается через A_n^m и называется количеством размещений из n по m. Если m=n, то эту величину называют количеством перестановок из n элементов и обозначают P_n.

Когда речь идет про подсчет количеств, слово «размещение» понимаем как синоним словосочетания «упорядоченное множество». Также слово «перестановка» понимается как размещение всех элементов множества в определенном порядке.

Любой неупорядоченный набор, т.е. подмножество $\left(x_{i_1}, x_{i_2}, \ldots, x_{i_m}\right)$ из m, различных элементов множества X $(m \le n)$ называется сочетанием m элементов из n , а их количество обозначается C_n^m. Это количество еще называют количеством комбинаций m элементов из n. В контексте вычислений слово «подмножество» и «комбинация» являются синонимами.

Если в размещении и сочетании из n элементов по m убрать требование различия элементов и разрешить повторение элементов из множества X, то такие размещение и сочетание называются с повторениями.

Основными модельными являются следующие комбинаторные задачи. Найти:

- число P_n всех перестановок из n элементов;
- число A_n^m всех размещений m элементов из n элементов;
- число $C_n^m = \binom{n}{m}$ всех сочетаний m элементов из n элементов;
- число $\overline{A}_n^m$ всех размещений *m* элементов из *n* с повторениями;
- число $\overline{C}_n^m$ всех сочетаний *m* элементов из *n* элементов с повторениями;

Во всех формулах будет встречаться факториал целого неотрицательного числа. Напомним, что факториалом целого положительного числа n (обозначение n!) называется произведение $1 \cdot 2 \cdot \ldots \cdot n$. По определению считается, что 0! = 1.

Перестановки. Переставляя объекты некоторого множества, мы обычно располагаем их в различном порядке. В этом смысле перестановка — это переупорядочение элементов множества. Исследуем, сколько существует способов переупорядочения элементов множества.

Рассмотрим количество возможных способов формирования числа, переставляя цифры 1, 2, 3, 4 и 5. Варианты возможных перестановок — это, например, числа 12345, 15342, 32415 и 32415. Для нахождения количества возможных перестановок заметим, что первую цифру можно выбрать пятью способами, вторую — четырьмя способами, третью — тремя способами, четвертую — двумя, и только один вариант остается для выбора пятой цифры. Поэтому существует 5! возможных перестановок. Точно так же, если необходимо переупорядочить n объектов, то для этого существует n! способов. Таким образом получена

Теорема: *число перестановок P_n равно* $P_n = n!$

Доказательство. На первое место в перестановке можно поставить любой из n элементов множества X, на второе место — уже любой из n−1 оставшихся и т. д. На последнее место остается только один элемент. Тем самым всего будет $P_n = n(n—1)... 1 = n!$ перестановок. □

В перестановках важен порядок. Числа 51342 и 32415, образованные перестановкой цифр 1, 2, 3, 4 и 5, не совпадают. Кроме того, поскольку перестановки рассматриваются как переупорядочения, то каждый элемент множества можно использовать только один раз. Если бы повтор цифр допускался, то при формировании числа для каждой цифры существовало бы пять вариантов выбора, поэтому существовало бы 5^5 возможных чисел.

Перестановки элементов 1,2,…,n записываются в матричной форме

$$\pi = \begin{pmatrix} 1 & 2 & \dots & n \\ \pi_1 & \pi_2 & \dots & \pi_n \end{pmatrix},$$

где верхняя строка – это порядковые номера 1,2,…,n позиций элементов в перестановке; нижняя строка — тот же набор чисел 1,2,..., n, взятых в каком-либо порядке; π_j — номер элемента на j-м месте перестановки. Порядок столбцов в перестановках, записанных в матричной форме, не является существенным, так как в этом случае номер позиции каждого элемента в перестановке указывается явно в верхней строке. Например, перестановка (3,2,4,1) из четырех элементов может быть записана по-разному

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 4 & 1 \end{pmatrix}, \begin{pmatrix} 3 & 1 & 4 & 2 \\ 4 & 3 & 1 & 2 \end{pmatrix}, \begin{pmatrix} 2 & 1 & 4 & 3 \\ 2 & 3 & 1 & 4 \end{pmatrix}$$

Пример. Задача о ладьях. Сколькими способами можно расположить на шахматной доске 8 ладей, чтобы они «не били» друг друга?

Решение. Условие «не могли бить» означает, что на каждой горизонтали и вертикали может стоять лишь одна ладья. Ввиду этого, каждому расположению ладей на доске соответствует перестановка

$$\pi = \begin{pmatrix} 1 & 2 & \dots & 8 \\ \pi_1 & \pi_2 & \dots & \pi_8 \end{pmatrix}$$

Верхняя строка перестановки — это номер горизонталей, нижняя — вертикалей, пересечение которых определяет положение ладей на доске. Следовательно, число расстановок ладей равно числу перестановок $P_8 = 8!$ из 8 элементов.

■

Размещения без повторений. Можно рассматривать перестановки, когда объектов больше, чем мест для их размещения. Предположим, например, что в организации – 20 человек и из них требуется выбрать президента, вице-президента, секретаря и казначея. Имеются 20 вариантов выбора президента, 19 вариантов выбора вице-президента, 18 способов выбора секретаря и 17 – казначея. Таким образом, получаем 20 x 19 x 18 x 17 способов выбора должностных лиц. Отметим, что порядок все еще остается существенным.

Предположим, имеется n человек. Требуется выбрать r из них и расположить в определенном порядке. Существует n способов выбрать первого человека, n – 1 способов выбора второго, n−2 способов выбора третьего, n−j+1 способов выбора j – го и n−r+1 способов выбора r – го человека. Следовательно, существует

$$(n)(n-1)(n-2)(n-3)\cdots(n-i+1)\cdots(n-r+1)$$

способов выбрать r человек из n. Но

$$(n)(n-1)(n-2)(n-3)\cdots(n-i+1)\cdots(n-r+1)=\frac{n!}{(n-r)!}.$$

Сформулируем вышесказанное в виде следующей теоремы.

Теорема. Количество способов выбрать r объектов с учетом порядка из n объектов равно

$$(n)(n-1)(n-2)(n-3)\cdots(n-i+1)\cdots(n-r+1)=\frac{n!}{(n-r)!}.$$

Число $A_n^r=\frac{n!}{(n-r)!}$ называется числом размещений r элементов из n элементов или размещением без повторений.

Заметим, что если выбирать все n объектов и размещать их в определенном порядке, то $r = n$ и, поскольку $0! = 1$, имеем

$$A_n^n=\frac{n!}{(n-n)!}=\frac{n!}{0!}=n!$$

Приходим к уже известному результату для перестановок n элементов.

Пример. В турнире «мисс факультет» участвуют 17 девушек. Разыгрываются призы за первое, второе и третье места. Сколькими способами могут быть распределены места?

Решение. 17 девушек претендуют на 3 места. Тогда тройку призеров можно выбрать $A_{17}^3=17\cdot16\cdot15=4080$ способами.

Пример. Сколько различных четырехзначных чисел можно образовать из цифр 1, 2, 3, ..., 9, если все цифры в каждом четырехзначном числе различны?

Для формирования каждого четырехзначного числа выбираем четыре цифры из девяти, поэтому существует

$$A_9^4=\frac{9!}{(9-4)!}=\frac{9!}{5!}=3024$$

таких различных чисел.

Пример. Сколькими способами можно рассадить 10 человек за круглым столом, если имеет значение только порядок соседей.

Существует несколько вариантов решения данной задачи. Первым делом, отметим, что вращение людей вокруг стола не меняет их взаимного расположения, поскольку соседи справа и слева остаются прежними. Предположим, что место за столом уникально. Тогда существует 10! способов рассадить людей за столом. Считаем, что при вращении места остаются теми же, так как соседи не меняются. Существует десять таких вращений, поэтому делим 10! на 10, что дает 9! способов расположить людей за столом, если имеет значение только порядок соседей.

Иной подход к решению задачи состоит в том, чтобы сначала усадить одного человека. Этим исключается вращение, а оставшихся 9 человек можно рассадить 9! способами.

■

Сочетания. В тех случаях, когда нас не интересует порядок элементов в размещении, а интересует лишь ее состав, то говорят о сочетаниях. Т.о. сейчас нас интересуют k – элементные подмножества исходного n – элементного множества. Их называют сочетаниями.

Сочетаниями из n различных элементов по k называют все возможные размещения длины k, образованные из этих элементов и отличающиеся друг от друга составом, но не порядком элементов. Общее число сочетаний обозначают через C_n^k или $\binom{n}{k}$.

Составим все сочетания из n по k. Затем переставим в каждом сочетании элементы всеми возможными способами. Теперь мы получили все размещения. Они отличаются либо составом, либо порядком, т.е. это все размещения без повторений из n по k. Их число равно A_n^k. Учитывая, что каждое сочетание дает k! перестановок, то по правилу произведения можно записать $C_n^k \times k! = A_n^k$. Тогда

$$C_n^k = \frac{A_n^k}{k!} = \frac{n(n-1)\ldots(n-k+1)}{k!} = \frac{n!}{k!(n-k)!}$$

Отсюда также видно, что $C_n^k = C_n^{n-k}$. Т.о. получаем

теорему: *количество способов выбора r объектов из n объектов без учета порядка равно* $C_n^r = \dfrac{n!}{r!(n-r)!}$.

Число C_n^r называется числом сочетаний из n объектов по r. Обратите внимание, что в случае сочетаний, как и в случае перестановок и размещений, при выборе r объектов каждый объект может быть выбран не более одного раза.

Пример. Если множество содержит десять элементов, то сколько оно имеет трехэлементных подмножеств? Поскольку множество не упорядочено, выбираются три элемента из десяти, поэтому всего имеется $C_{10}^3 = \dfrac{10\,!}{7!\,3!} = 120$ различных подмножеств.

Пример. Сколько строк длины девять содержат ровно 5 единиц и 4 нуля?
В строке имеется девять мест для размещения 1 и 0. Можно выбрать любые пять из девяти мест для размещения единицы. Поэтому имеется $C_9^5 = \dfrac{9!}{5!4!} = 126$ мест для размещения единицы. Как только единицы вставлены, остальные места заполняются нулями.

Пример. Сколько существует вариантов выбора 5 карт из стандартной колоды, содержащей 52 карты? Поскольку порядок карт не имеет значения, речь идет о выборе 5 объектов из 52, поэтому существует $C_{52}^5 = \dfrac{52!}{5!47!} = 2598960$ возможных комбинаций.

Пример. *Задача о прямоугольниках*. Сколько различных прямоугольников можно вырезать из клеток доски, размер которой m x n?

1	2	3	...	*n*
2				
3				
...				
m				

Решение. Прямоугольник однозначно определяется положением его сторон. Горизонтальные стороны могут занимать любое из m+1 положения. Тогда число способов их выбора равно C_{m+1}^2 (порядок сторон – какая верхняя, а какая нижняя несущественен). Вертикальные стороны можно выбрать C_{n+1}^2 способами. По правилу прямого произведения заключаем, что количество прямоугольников равно $C_{m+1}^2 \cdot C_{n+1}^2$.

Пример. Сколько всего партий играется в шахматном турнире с n участниками? Ответ: C_n^2, так как каждая партия однозначно определяется двумя ее участниками.

Пример. Вернемся еще раз к примеру с подсчетом количества путей из вершины ***a*** в вершину ***b*** на прямоугольной сети дорог предыдущего параграфа. Каждому пути можно поставить в соответствие двоичное слово, обозначая единицей перемещение по горизонтали и нулем – по вертикали. Чтобы попасть из вершины ***a*** в вершину ***b*** надо сделать 4 перемещения по горизонтали и 3 по вертикали. Поэтому получающееся двоичные слова будут иметь длину 7 и состоять из 4 единиц и 3 нулей. Наоборот, каждому такому слову соответствует некоторый путь из вершины ***a*** в вершину ***b***. Число таких двоичных слов равно $C_7^4 = 35$ и, следовательно, число всевозможных путей равно 35.

□

Рассмотрим **размещение с повторением** $\overline{A}_m^n$ из n элементов по m. На первое место в таком размещении можно поставить любой из n элементов множества X, а после этот элемент возвращается в X. Поэтому на второе и остальные места до m можно снова поставить любой из n элементов. Тем самым имеем

$$\overline{A}_m^n = \underbrace{n \cdot n \cdot \ldots \cdot n}_{m} = n^m$$

Пример. Корзина содержит 50 шаров с номерами. Из него выбирают шар, номер которого записывают. Шар возвращают в корзину, и процедура повторяется 5 раз. Подсчитаем количество возможных комбинаций получаемых чисел. Для каждого из пяти чисел имеется 50 способов выбора. Следовательно, число различных комбинаций составляет 50^5.

□

Сочетание с повторением. Пусть $\{a_1, a_2, \ldots, a_m\}$ – множество из m элементов. Обычно мы полагаем, что элементы множества не повторяются. Откажемся от этого принципа. Множество, составленное из $a_1, a_2, \ldots, a_m$, в котором элементы могут повторяться называется мультимножеством. Для задания мультимножества надо указать, сколько раз в него входит каждый из элементов, например, следующим образом

$$M = \begin{pmatrix} a_1 & \dots & a_m \\ k_1 & \dots & k_m \end{pmatrix}, k_i \geq 0, i = 1, \dots, m.$$

Рассмотрим, например, мультимножество M={*a, a, a, b, b, c, d, d, d, d*}, в котором содержатся 3 элемента *a*, 2 элемента *b*, 1 элемент *c* и 4 элемента *d*. Мультимножество — это то же самое, что и множество, но в нем могут содержаться одинаковые элементы. Здесь элементы *a,b,c,d* различимы, а, например, три элемента *a* мультимножества – неразличимы между собой.

Такие объекты встречаются повседневно. Например, зайдя в булочную, мы выбрали 3 пончика вида *a*, два пончика вида *b*, один пончик вида *c* и 4 пончика вида *d*. Пончики одного вида для нас неразличимы. Всего мы выбрали 10 пончиков из 4 –х различных видов. Можно поставить вопрос, сколько различных комбинаций выбора десяти пончиков из 4 – х видов существует. Для ответа на этот вопрос имеется

Теорема. Количество различных сочетаний n объектов из k различных типов равно

$$\overline{C}_k^n = C_{n+k-1}^{k-1} = C_{n+k-1}^n = \frac{(n+k-1)!}{n!(k-1)!}$$

Доказательство. Предположим, что n объектов выбираются из k типов и повторение допускается. Обозначим a_i – объект типа *i*. Тогда наш выбор определяется мультимножеством

$$M = \begin{pmatrix} a_1 & \dots & a_k \\ n_1 & \dots & n_k \end{pmatrix}$$

или по – другому

$$a_1a_1 \dots a_1 \mid a_2a_2 \dots a_2 \mid \quad \dots \quad \mid a_ka_k \dots a_k \; ,$$

где в i – ой группе объектов элементы a_i повторяются n_i раз и $n_1 + n_2 + \dots + n_k = n$. Группы завершаются вертикальной чертой за исключением последней серии элементов. Поскольку место расположения каждого типа понятно, то выборку можно записать в виде

$$xxx \cdots x | xxx \cdots x | \cdots | xxx \cdots x.$$

Заметим, что разделителей | на один меньше количества типов. Таким образом, имеем n объектов плюс k−1 разделителей, образующих n+k−1 мест для размещения x или |. Каждое расположение знаков x и | дает новый способ выбора n объектов из k типов. Поскольку существует C_{n+k-1}^n способов выбора места для знака x или C_{n+k-1}^{k-1} способов выбора знака | (что эквивалентно), то существуют $C_{n+k-1}^{k-1} = C_{n+k-1}^n$ различных способов выбора n объектов из k типов объектов с возможностью неограниченного повторения.

□

Отметим, что здесь под символом x мы не имели в виду объект какого – то конкретного типа, т.е. мы могли говорить об n неразличимых объектах, но различимых группах, в которые мы их пометили. Поэтому это модельное число часто называют числом размещений n неразличимых предметов по k ящикам. Однако, обычно, это число называют числом сочетаний из n по k с повторением и обозначают $\overline{C}_k^n$.

Пример. Сколькими способами трое ребят могут разделить между собой четыре яблока?

$$\overline{C}_3^4 = C_{3+4-1}^{3-1} = C_6^2 = \frac{6!}{2!(6-2)!} = \frac{6\cdot 5}{2} = 15$$

Проверим этот результат. Для этого перечислим все возможные способы дележа яблок: (4,0,0), (0,4,0), (0,0,4), (3,1,0), (3,0,1), (0,3,1), (1,3,0), (0,1,3), (1,0,3), (2,2,0), (2,0,2), (0,2,2), (2,1,1), (1,2,1), (1,1,2). Всего 15 способов.

Пример. Вернемся к задаче с пончиками. Сколько существует различных вариантов выбора десяти пончиков из 4 – х видов? Поскольку 10 объектов выбираются из 4 различных типов с повторением, то имеются

$$\overline{C}_4^{10} = C_{4+10-1}^3 = \frac{13!}{3!(13-3)!} = \frac{13!}{3!10!} = \frac{13\cdot 12\cdot 11}{6} = 286.$$

Пример. Найти количество целочисленных решений системы

$$x_1 + x_2 + \ldots + x_n = k,\ k \ge 0,\ x_i \ge 0,\ i = 1, 2, \ldots, n;\ n \ge 1.$$

Решение. Число k можно представить в виде суммы k единиц, которые следует разложить в n корзин, сумма единиц в каждой из которых будет представлять число x_i. Но число размещений k неразличимых предметов (в нашем случае единиц) по n корзинам (числам x_i) равно

$$\overline{C}_n^k = C_{n+k-1}^{n-1} = C_{n+k-1}^k = \frac{(n+k-1)!}{k!(n-1)!}.$$

□

Перестановки с повторениями. Рассмотрим теперь задачу о количестве перестановок букв в слове «КОЛОБОК». Оно состоит из семи букв, которые можно переставить 7! способами. Однако в нем есть три буквы «О» и две буквы «К». Поэтому меняя местами буквы «О» или переставляя буквы «К», мы не получим новых «слов». Фактически мы имеем мультимножество М={*К, К, О, О, О, Л, Б*}, элементы которого относятся к 4-м разным типам (являются разными буквами). Обозначим их через *a, b, c, d*. Тогда М={*a, a, b, b, b, c, d*}. Если бы мы рассматривали все элементы множества М как различные, обозначив их $a_1, a_2, b_1, b_2, b_3, c_1, d_1$ то получили бы 7! перестановок, но после отбрасывания индексов многие из них оказались бы одинаковыми. Фактически каждая перестановка множества М встретилась бы ровно 2!3!1!1! раз, поскольку в любой перестановке М индексы при буквах *a* можно расставить 2! способами, при *b* — 3! способами, при *c* и *b* — одним способом. Поэтому число перестановок элементов мультимножества М равно $\frac{7!}{3!\,2!\,1!\,1!}=420$ и, тем самым, мы получаем, что разных «слов» из слова «КОЛОБОК» можно составить 420.

В применении к общему случаю те же рассуждения показывают, что число перестановок любого мультимножества (перестановки с повторениями) равно

$$P(n_1,n_2,\ldots,n_r)=\frac{n!}{n_1!\,n_2!\ldots n_r!},\ n=n_1+n_2+\ldots+n_r.$$

Величины $\frac{n!}{n_1!\,n_2!\ldots n_r!}$ носят название мультиномиальных коэффициентов. Как будет показано далее, они стоят при произведениях $x_1^{n_1}x_2^{n_2}\ldots x_r^{n_r}$ в разложении выражения $(x_1+x_2+\ldots+x_r)^n$ по степеням $x_1^{n_1}x_2^{n_2}\ldots x_r^{n_r}$.

Получим предыдущую формулу другим способом. Пусть мультимножество состоит из r элементов различного вида и элементов каждого вида имеется неограниченное количество. Определим количество перестановок с повторением $P(n_1,n_2,\ldots,n_r)$ по – другому. Из всех n возможных мест, на которых будут находиться элементы, n_1 место должны занимать предметы 1 – го вида. Выбор мест для них можно сделать $C_n^{n_1}$ способами. Из оставшихся $n-n_1$ мест элементы второго типа занимают n_2 места, которые можно выбрать $C_{n-n_1}^{n_2}$ способами. Те же рассуждения показывают, что элементы r - го вида можно расположить в перестановке $C_{n-n_1-\ldots-n_{r-1}}^{n_k}$ способами. Здесь важны места расположение элементов каждого вида и, когда эти места выбраны, порядок элементов одного вида на выбранных для них местах не имеет значения. В результате, согласно правилу прямого произведения, число перестановок с повторениями равно

$$P(n_1,n_2,\ldots,n_k)=C_n^{n_1}C_{n-n_1}^{n_2}\ldots C_{n-n_1-\ldots-n_{k-1}}^{n_k}=\frac{n!}{n_1!n_2!\ldots n_k!}.$$

где $n=n_1+n_2+\ldots+n_k$ общее число элементов.

Пример. Сколько существует различных перестановок из букв слова СССР. Их легко перечислить: СССР, ССРС, СРСС, РССС – всего четыре. По формуле имеем: $P(3_C,1_Ð)=C_4^3\cdot C_{4-3}^1=4$.

Пример. В студенческой группе, состоящей из 25 человек, при выборе старосты за выдвинутую кандидатуру проголосовали 19 человек, против — 3, воздержались — 3. Сколькими способами могло быть проведено такое голосование?

Решение. Имеем три различные корзины: «за», «против», «воздержались», в которые необходимо разложить 25 неразличимых шаров, соответственно 19 —

в первую, 3 — во вторую, 3 — в третью. Количество таких разложений определяется выражением $C_{25}^{19}\,C_6^3\,C_3^3 = \dfrac{25!}{19!\,3!\,3!}$.

□

Пусть имеется множество M мощности $|M| = n$. Разобьем его на несколько непересекающихся подмножеств, объединение которых даст все M. Порядок подмножеств в разбиении не является существенным. Например, разбиения множества `M={1,2,3,4,5}` вида `{{1,3},{4},{2,5}}`, `{{1,3},{2,5},{4}}`, `{{2,5},{1,3},{4}}` считаются одинаковыми.

Подсчитаем, сколькими способами можно разбить множество M на подмножества, среди которых имеется m_1 одноэлементных, m_2 двухэлементных, m_3 трехэлементных и т.д. m_n n-элементных подмножеств так, что $\sum_{i=1}^{n} i \cdot m_i = n$. Число неупорядоченных разбиений множества M обозначается через $N(m_1, m_2, \ldots, m_n)$.

Например, 17 яблок можно разложить в 6 кучек по 2 яблока и одну кучку из 5-и яблок. Тогда $m_2=6$ и $m_5=1$, а остальные $m_i=0$ $(i = 1\ldots17,\ i \neq 2,\ i \neq 5)$. Число различных способов разложения этих яблок обозначается $N(0_1, 6_2, 0_3, 0_4, 1_5, 0_6, 0_7, \ldots, 0_{17})$

Вначале посчитаем количество упорядоченных перестановок множества. Для этого воспользуемся интерпретацией упорядоченных перестановок как разложения n шаров по различным $m_1 + m_2 + \ldots + m_n$ корзинам так, что в каждую из m_i корзин кладут i шаров. Т.е. мы имеем m_1 одноэлементных «различимых» корзин, m_2 – двухэлементных и т.д.

$$\lfloor 1 \rfloor \lfloor 1 \rfloor \ldots \lfloor 1 \rfloor \; \lfloor 1,1 \rfloor \lfloor 1,1 \rfloor \ldots \lfloor 1,1 \rfloor \; \lfloor 1,1,1 \rfloor \ldots \lfloor 1,1,1 \rfloor \ldots \lfloor 1,1,\ldots,1 \rfloor$$

Тогда всего возможных вариантов разбиения будет

$$P\left(\underbrace{1,1,\ldots,1}_{m_1}, \underbrace{2,2,\ldots,2}_{m_2}, \underbrace{3,\ldots,3}_{m_3}, \ldots, \underbrace{n}_{m_n}\right) = \frac{n!}{\underbrace{1!\ldots1!}_{b_1}\,\underbrace{2!\ldots2!}_{b_2}\,\underbrace{3!\ldots3!}_{b_3}\ldots\underbrace{n!}_{m_n}} = \frac{n!}{(1!)^{m_1}(2!)^{m_2}(3!)^{m_3}\cdot\ldots\cdot(n!)^{m_n}}$$

Теперь откажемся от упорядоченности подмножеств в разбиении, т.е. начнем считать корзины с одинаковым количеством элементов неразличимыми. Если все корзины имеют различное число шаров, то их можно рассматривать как различные (они отличаются числом шаров). В этом случае упорядоченные и неупорядоченные разложения шаров совпадают. Пусть теперь в разложении существуют m_i корзин с одинаковым количеством шаров. При упорядоченном разложении такие корзины рассматриваются как различные. Однако при неупорядоченном разложении обмен шарами таких корзин можно рассматривать как соответствующую перестановку указанных корзин, что не приводит к новым разложениям. Если количество корзин с одинаковым числом шаров равно m_i, то неупорядоченных разложений будет в $m_i!$ меньше, чем упорядоченных. Общее число неупорядоченных разбиений будет в $m_1!\,m_2!\ldots m_n!$ раз меньше, чем упорядоченных. Следовательно,

$$N(m_1, m_2, \ldots, m_n) = \frac{n!}{(1!)^{m_1}(2!)^{m_2}\ldots(n!)^{m_n}\, m_1! m_2! \ldots m_n!}$$

Заметим еще раз, что если выполнено упорядоченное разбиение числа n на подмножества различной длины (мощности), то они совпадают с неупорядоченными разбиениями. В этом случае все $m_i = 0$ *или* 1.

Пример. Сколькими способами 17 яблок можно разложить в 6 кучек по 2 яблока и одну кучку из 5-и яблок?

Решение. Требуется разбить множество из 17 яблок на непересекающиеся и неупорядоченные кучки. Откуда искомое число равно

$$N(0_1, 6_2, 0_3, 0_4, 1_5, 0_6, 0_7, \ldots, 0_{17}) = \frac{17!}{(2!)^6 (5!)^1 6! 1!}$$

Пример. Сколькими способами можно разделить колоду из 36 карт пополам так, чтобы в каждой пачке было по два туза?

Решение. Для разделения 4-х тузов на две группы есть $\frac{4!}{(2!)^2 2!} = 3$ способа.

Поясним это. Если бы нам надо было определить количество способов выбрать 2 туза из четырех, то мы бы имели $C_4^2 = \frac{4!}{2!\,2!} = 6$ способов. Перечислим эти варианты в левом столбце следующей таблицы, указав масти тузов, и в правом столбце укажем масти двух оставшихся тузов

♦ ♥	♣ ♠
♦ ♣	♥ ♠
♦ ♠	♥ ♣
♥ ♠	♦ ♣
♥ ♣	♦ ♠
♣ ♠	♦ ♥

В нашей задаче, например, первая и последняя строки представляют одно и то же разбиение, т.к. сейчас для нас не имеет значения в какую из двух групп попали тузы ♦ ♥ и ♣ ♠. Поэтому различных способов разделить 4 туза пополам в два раза меньше, т.е. всего 3.

Далее. Каждая половина любого из этих трех разбиений тузов играет роль различных двух «корзин», куда необходимо разложить пополам оставшиеся 32 карты. Разложение 32 оставшихся карт уже будет упорядоченным, так как «корзины» различные, число разложений равно $\frac{32!}{16!16!}$. Согласно правилу прямого произведения, общее число вариантов разделить колоду пополам равно $3 \times \frac{32!}{16!16!}$.

6.3 Бином Ньютона

Бином Ньютона. Рассмотрим разложение

$$(a+b)^5 = (a+b)(a+b)(a+b)(a+b)(a+b).$$

Каждое слагаемое в разложении является результатом выбора a или b в каждом сомножителе $(a+b)$ и последовательного их перемножения. Например, a^5 получено путем выбора a из каждого сомножителя. Если из первого сомножителя выбрано a и b выбрано из остальных сомножителей, то в результате получим ab^4. Предположим, что требуется найти коэффициент при a^3b^2. Слагаемое a^3b^2 получается при выборе трех a и двух b из пяти сомножителей. Поскольку существует C_5^3 способов выбора трех a, то коэффициент при a^3b^2 равен C_5^3. Обобщая результат, получаем следующую теорему.

Биномиальная **теорема**. Для произвольного положительного целого числа n справедливы равенства

$$(a+b)^n = \sum_{r=0}^{n} C_n^r a^r b^{n-r} = \sum_{r=0}^{n} C_n^r a^{n-r} b^r$$

$$(a+b)^n = C_n^0 a^n + C_n^1 a^{n-1} b + C_n^2 a^{n-2} b^2 + \ldots + C_n^r a^{n-r} br + \ldots + C_n^n b^n.$$

Доказательство. Поскольку $a^r b^{n-r}$ получено в результате r-кратного выбора a и n-r-кратного выбора b из n сомножителей в выражении $(a + b)^n$, то коэффициент при $a^r b^{n-r}$ равен числу способов r-кратного выбора a из n сомножителей C_n^r. Второе равенство следует из того факта, что $C_n^r = C_n^{n-r}$.

□

Последняя формула известна под названием «бином Ньютона». Поскольку C_n^r - это коэффициенты в биноме Ньютона, то числа C_n^r еще называют биномиальными коэффициентами.

Примерами бинома Ньютона при n=1,2,3 будут

$(a+b)^1 = a+b = C_1^0 a^1 + C_1^1 b^1$ при n=1

$$(a+b)^2 = a^2 + 2ab + b^2 = C_2^0 a^2 + C_2^1 ab + C_2^2 b^2 \quad \text{при } n = 2,$$

$$(a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3 = C_3^0 a^3 + C_3^1 a^2 b + C_3^2 ab^2 + C_3^0 b^3 \quad \text{при } n = 3.$$

Пример. Покажем, что для любого положительного целого числа n

$$2^n = \sum_{r=0}^{n} C_n^r$$

Пусть $a = b = 1$, тогда по биномиальной теореме $(1+1)^n = \sum_{r=0}^{n} C_n^r 1^r 1^{n-r}$ и результат очевиден.

Пример. Количество подмножеств n – элементного множества.

Все подмножества заданного n-элементного множества можно перечислить следующим образом. Вначале отмечаем пустое множество, потом перечисляем одноэлементные подмножества, потом перечисляем двухэлементные и т.д. Например, все подмножества множества $\{a,b,c\}$ можно разместить следующим образом

$$\begin{gathered}\varnothing\\ a,\ b,\ c,\\ \{a,b\},\{a,c\},\{b,c\},\\ \{a,b,c\}\end{gathered}$$

Напомним, что C_n^k представляет количество способов выбрать k-элементное подмножество из n-элементного множества или, по-другому, количество различных k-элементных подмножеств. По правилу суммы количество подмножеств n-элементного множества равно

$$C_n^0 + C_n^1 + C_n^2 + \ldots + C_n^{n-1} + C_n^n.$$

Используя результат предыдущего примера, получаем, что каждое n-элементное множество имеет 2^n разных подмножеств.

Пример. Доказать тождество $\sum_{r=0}^{n} C_n^r (m-1)^{n-r} = m^n$.

Воспользуемся формулой бинома Ньютона, где положим *a=1* и *b=m-1*. □

Полиномиальная формула. Пусть слагаемых не два, а больше. Имеет место формула

$$(x_1 + x_2 + \ldots + x_k)^n = \sum_{n_1+n_2+\ldots+n_k=n} \frac{n!}{n_1!n_2!\ldots n_k!} x_1^{n_1} x_2^{n_2} \ldots x_k^{n_k}$$

Она называется полиномиальной, где суммирование выполняется по всем решениям уравнения $n_1 + n_2 + \ldots + n_k = n$ в целых неотрицательных числах, $n_i \geq 0,\ i = 1, 2, \ldots, k$. Для доказательства выполним умножение

$$\underbrace{(x_1 + x_2 + \ldots + x_k)(x_1 + x_2 + \ldots + x_k)\ldots(x_1 + x_2 + \ldots + x_k)}_{n} = (x_1 + x_2 + \ldots + x_k)^n.$$

Чтобы привести подобные в полученном выражении, необходимо подсчитать количество одночленов вида $x_1^{n_1} x_2^{n_2} \ldots x_k^{n_k}$ каждого разбиения $n_1 + n_2 + \ldots + n_k = n$. Для получения же одночлена $x_1^{n_1} x_2^{n_2} \ldots x_k^{n_k}$ необходимо выбрать x_1 в качестве множителя в n_1 скобках при раскрытии выражения $(x_1 + x_2 + \ldots + x_k)^n$. Это можно сделать $C_n^{n_1}$ способами. Из оставшихся $n - n_1$ не раскрытых скобок необходимо выбрать x_2 в качестве множителя в n_2 скобках. Это можно сделать $C_{n-n_1}^{n_2}$ способами и т. д. Тогда количество одночленов $x_1^{n_1} x_2^{n_2} \ldots x_k^{n_k}$ при раскрытии выражения

$$\underbrace{(x_1 + x_2 + \ldots + x_k)(x_1 + x_2 + \ldots + x_k)\ldots(x_1 + x_2 + \ldots + x_k)}_{n}.$$

будет равно числу $C_n^{n_1} C_{n-n_1}^{n_2} \ldots C_{n-n_1-\ldots-n_{k-1}}^{n_k} = \dfrac{n!}{n_1!n_2!\ldots n_k!}$ перестановок с повторениями. ■

Частный вид полиномиальной формулы, содержащий только два слагаемых, $(a+b)^n = \sum_{r=0}^{n} C_n^r a^r b^{n-r}$ называется биномом Ньютона.

Теорема. (Формула Паскаля). Для всех целых чисел r и n таких, что $1 \le r \le n$ имеет место

$$C_n^r = C_{n-1}^{r-1} + C_{n-1}^r \tag{1}$$

Доказательство. С одной стороны мы имеем

$$(a+b)^n = C_n^0 a^n + C_n^1 a^{n-1} b + \ldots + C_n^r a^{n-r} b^r + \ldots + C_n^n b^n,$$

По-другому

$$\begin{gathered}(a+b)^n = (a+b)^{n-1}(a+b) = \\ = (C_{n-1}^0 a^{n-1} + \ldots + C_{n-1}^{r-1} a^{n-r} b^{r-1} + C_{n-1}^r a^{n-r-1} b^r + \ldots + C_{n-1}^{n-1} b^{n-1})(a+b) = \\ = C_{n-1}^0 a^n + \ldots + C_{n-1}^r a^{n-r} b^r + \ldots + C_{n-1}^{n-1} a b^{n-1} + \\ + C_{n-1}^0 a^{n-1} b + \ldots + C_{n-1}^{r-1} a^{n-r} b^r + \ldots + C_{n-1}^{n-1} b^n.\end{gathered}$$

Приравнивая коэффициенты при $a^{n-r} b^r$ в обеих формулах, убеждаемся в справедливости формулы (1).

□

Формула (1) дает эффективный способ вычисления биномиальных коэффициентов. Запишем биномиальные коэффициенты в таблицу (она называется треугольником Паскаля)

$$\begin{array}{ccccccccc} & & & & C_0^0 & & & & \\ & & & C_1^0 & & C_1^1 & & & \\ & & C_2^0 & & C_2^1 & & C_2^2 & & \\ & C_3^0 & & C_3^1 & & C_3^2 & & C_3^3 & \\ \ldots & & & & & & & & \\ C_n^0 & & C_n^1 & & \ldots & & \ldots & C_n^{n-1} & C_n^n \end{array}$$

Каждая (n+1)-я строка этого треугольника состоит из биномиальных коэффициентов, получающихся при раскрытии скобок в выражении $(a+b)^n$. Так как $C_n^0 = C_n^n = 1$, на внешних сторонах треугольника Паскаля всегда стоят единицы. Симметрия относительно вертикальной высоты треугольника следует из тождества $C_n^k = C_n^{n-k}$. Выпишем 5 строк треугольника Паскаля

$$\begin{array}{ccccccccc} & & & & 1 & & & & \\ & & & 1 & & 1 & & & \\ & & 1 & & 2 & & 1 & & \\ & 1 & & 3 & & 3 & & 1 & \\ 1 & & 4 & & 6 & & 4 & & 1 \end{array}$$

Видим также, что каждое внутреннее число равно сумме двух верхних соседей. Это правило нами сформулировано в виде формулы Паскаля $C_n^r = C_{n-1}^{r-1} + C_{n-1}^r$.

Формулу Паскаля можно доказать комбинаторно. Зафиксируем некоторый элемент n – элементного множества. Затем все r – элементные подмножества разобьем на два типа – к одному типу отнесем те подмножества, которые содержат выделенный элемент, а ко второму отнесем те подмножества, которые этот элемент не содержат. Ясно, что сумма количества первых и вторых подмножеств будет равна количеству r – элементных подмножеств, т.е. C_n^r. Определим количество подмножеств первого типа, которые содержат выделенный элемент. Поскольку элемент уже выбран, то надо выбрать r-1 объект из n-1 элемента множества, т.е. всего способов C_{n-1}^{r-1}. Рассмотрим подмножества, которые не содержат выделенный элемент. По-прежнему требуется выбрать r объектов, но теперь из $n - 1$ объектов, учитывая, что один элемент не может быть выбран. Таким образом, существуют C_{n-1}^{r} способов сделать такой выбор. Складывая количество способов выбора в обоих случаях, получаем $C_n^r = C_{n-1}^{r-1} + C_{n-1}^{r}$.

Заметим, что доказать эту теорему можно также методом математической индукции.

Теорема Вандермонда. Пусть m, n, r – положительные целые числа такие, что $r \le \min(m,n)$. Тогда

$$C_{m+n}^{r} = \sum_{k=0}^{r} C_m^k \cdot C_n^{r-k}$$

Доказательство. C_{m+n}^{r} – это число способов выбрать r предметов из $m+n$ предметов. Предметы можно выбирать в два приема: сначала выбрать k предметов из первых m предметов, а затем выбрать недостающие $r-k$ предметов из оставшихся n предметов. Всего для этого существует $C_m^k \cdot C_n^{r-k}$ способов. Отсюда общее число способов выбрать r предметов составляет $\sum_{k=0}^{r} C_m^k \cdot C_n^{r-k}$.

6.4 Принцип включений – исключений.

Поставим задачу подсчитать количество элементов в объединении нескольких множеств. Для двух множеств мы имели

$$|A_1 \cup A_2| = |A_1| + |A_2| - |A_1 \cap A_2| \qquad (1)$$

Рассмотрим теперь объединение трех множеств. Обозначим $A_1 \cup A_2 = A$ и применим предыдущую формулу

$$|A_1 \cup A_2 \cup A_3| = |A \cup A_3| = |A| + |A_3| - |A \cap A_3| =$$
$$= |A_1 \cup A_2| + |A_3| - |(A_1 \cup A_2) \cap A_3| = |A_1 \cup A_2| + |A_3| - |A_1 \cap A_3 \cup A_2 \cap A_3|$$

Применяя к $|A_1 \cup A_2|$ и $|A_1 \cap A_3 \cup A_2 \cap A_3|$ формулу для двух множеств, приходим к следующему соотношению

$$|A_1 \cup A_2 \cup A_3| = |A_1| + |A_2| + |A_3| - |A_1 \cap A_2| - |A_1 \cap A_3| - |A_2 \cap A_3| + |A_1 \cap A_2 \cap A_3|$$

Переходя к n ($n>2$) множествам $A_1, A_2, \ldots, A_n$ с использованием индуктивных соображений можно получить формулу

$$\left|\bigcup_{i=1}^{n}A_i\right|=\sum_{i=1}^{n}|A_i|-\sum_{1\le i<j\le n}|A_i\cap A_j|+\sum_{1\le i<j<k\le n}|A_i\cap A_j\cap A_k|-\ldots+$$
$$+(-1)^{p+1}\sum_{1\le i_1<i_2<\ldots<i_p\le n}|A_{i_1}\cap A_{i_2}\cap\ldots\cap A_{i_p}|+\ldots+(-1)^{n+1}|A_1\cap A_2\cap\ldots\cap A_n|, \quad (2)$$

которая в словесной формулировке выглядит так:
чтобы найти количество элементов в объединении множеств, нужно сложить количества элементов в каждом множестве, затем вычесть количество элементов во всевозможных попарных пересечениях, прибавить количество элементов во всевозможных пересечениях по три и т.д.

Для доказательства формулы (2) необходимо показать, что в правой части каждый элемент множества $A_1\cup A_2\cup\ldots\cup A_n$ подсчитан и учтен ровно один раз. Предположим, что элемент a принадлежит в точности p множествам из набора $A_1,A_2,A_3,\ldots,A_n$. В сумме $\sum_{i=1}^{n}|A_i|$ элемент a учтен p раз. В сумме $\sum_{1\le i<j\le n}|A_i\cap A_j|$ элемент a учтен всякий раз, когда выбраны два множества, содержащие элемент a. Существуют C_p^2 способов выбрать такие два множества. Следовательно, в сумме $\sum_{1\le i<j\le n}|A_i\cap A_j|$ элемент a учтен C_p^2 раз. В сумме $\sum_{1\le i<j<k\le n}|A_i\cap A_j\cap A_k|$ элемент a учтен всякий раз, когда выбраны три множества, содержащие элемент a. Существуют C_p^3способов выбрать такие три множества. Следовательно, в сумме $\sum_{1\le i<j<k\le n}|A_i\cap A_j\cap A_k|$ элемент a учтен C_p^3 раз. В сумме элементов всех возможных пересечений r множеств, где $r\le p$, элемент a учтен C_p^r раз. Следовательно, в правой части формулы (2) элемент a учтен $p-C_p^2+C_p^3-\ldots+(-1)^{i+1}C_p^i+\ldots+(-1)^{p+1}$ раз. Но

$$1-C_p^1+C_p^2-C_p^3-\ldots+(-1)^iC_p^i+\ldots+(-1)^pC_p^p=(1-1)^p=0$$

Поэтому

$$1=C_p^1-C_p^2+C_p^3-\ldots+(-1)^{i+1}C_p^i+\ldots+(-1)^{p+1}$$

и получаем, что элемент a учтен в точности один раз.

□

На основании этой теоремы докажем утверждение, известное как

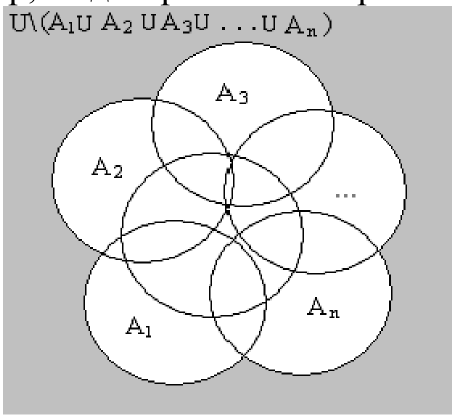
Теорема о включении – исключении. Пусть $A_1,A_2,A_3,\ldots,A_n$ — набор конечных множеств. Количество элементов множества $\overline{A_1}\cap\overline{A_2}\cap\ldots\cap\overline{A_n}$ определяется формулой

$$|\overline{A_1}\cap\overline{A_2}\cap\ldots\cap\overline{A_n}| = |U| - \sum_i |A_i| + \sum_{i<j}|A_i\cap A_j| - \\ - \sum_{i<j<k}|A_i\cap A_j\cap A_k| + \ldots + (-1)^n |A_1\cap A_2\cap\ldots\cap A_n| . \quad (3)$$

Доказательство. Из теории множеств известно, что

$$\overline{A_1}\cap\overline{A_2}\cap\ldots\cap\overline{A_n} = \overline{A_1\cup A_2\cup\ldots\cup A_n} = U\setminus(A_1\cup A_2\cup\ldots\cup A_n)$$

Это легко видеть, например, из диаграммы Эйлера



так что имеем соотношение

$$\left|\overline{A_1}\cap\overline{A_2}\cap\ldots\cap\overline{A_n}\right| = |U| - |A_1\cup A_2\cup\ldots\cup A_n|$$

Заменяя в этом равенстве число $|A_1\cup A_2\cup\ldots\cup A_n|$ его представлением по формуле (2), получаем требуемый результат.

□

Название этой теоремы подчеркивает использование последовательных включений и исключений элементов подмножеств. Формулы (2) или (3) называются формулами включения – исключения, а метод решения комбинаторных задач с их использованием называется методом включения – исключения. Этот метод обобщает правило суммы, которое было рассмотрено нами вначале темы «комбинаторика»

Пример. Рассмотрим слова длины n в алфавите {0,1,2}. Сколько имеется слов, в которых встречаются все три цифры?

Обозначим A_i множество всех слов длины n, в которых не встречается цифра *i*, *i=0,1,2*. Тогда $|A_0| = |A_1| = |A_2| = 2^n$. Кроме того, $|A_0\cap A_1| = |A_0\cap A_2| = |A_1\cap A_2| = 1$. Наконец, $|A_0\cap A_1\cap A_3| = 0$. В множество $A_0\cup A_1\cup A_3$ входят слова, в которых отсутствует хотя бы одна цифра. По принципу включений – исключений (формула (2))

$$|A_0 + A_1 + A_2| = 2^n + 2^n + 2^n - 1 - 1 - 1 + 0 = 3\ (2^n - 1)$$

Следовательно, число слов, в которых присутствуют все три цифры, равно $3^n - 3\cdot(2^n - 1)$.

□

Иногда удобно формулу из теоремы включений и исключений трактовать в другом виде. Пусть подмножество X_i определяется наличием некоторого свойства P_i у элементов множества U. Тогда подмножество $X = X_1 \bigcup X_2 \bigcup \ldots \bigcup X_m$ объединяет элементы из U, которые обладают хотя бы одним из свойств P_i. Дополнение $\overline{X}$ составляют элементы, которые не обладают ни одним из свойств $P_i, i=1,2,\ldots,m$. Пересечения вида $X = X_{i_1} \bigcap X_{i_2} \bigcap \ldots \bigcap X_{i_k}$ объединяют элементы, обладающие одновременно свойствами $P_{i_1}, P_{i_2}, \ldots, P_{i_k}$. Если обозначить число элементов в U через $|U| = n$, число элементов, обладающих одновременно набором свойств $P_{i_1}, P_{i_2}, \ldots, P_{i_k}$, через $N(P_{i_1}, P_{i_2}, \ldots, P_{i_k})$, то для числа элементов, не обладающих ни одним из свойств $P_i, i=1,2,\ldots,m$, имеем формулу

$$N(\overline{P}_1, \overline{P}_2, \ldots, \overline{P}_m) =$$
$$= n - \sum_i N(P_i) + \sum_i \sum_j N(P_i, P_j) - \ldots + (-1)^m N(P_1, P_2, \ldots, P_m).$$

Пример. Сколько натуральных чисел из первых 100 не делятся одновременно на 2, 3 и 5? Обозначим число натуральных чисел из первых 100, делящихся на два через $N(P_2)$. Это число легко найти: $N(P_2) = [100/2] = 50$. Здесь квадратные скобки обозначают наибольшее целое число, не превосходящее данное. Аналогично $N(P_3) = [100/3] = 33$, $N(P_5) = [100/5] = 20$. Далее число натуральных чисел из первых 100, делящихся одновременно на 2 и 3, т. е. на наименьшее их кратное 6, равно $N(P_2, P_3) = [100/6] = 16$. Аналогично получим $N(P_2, P_5) = [100/10] = 10$, $N(P_3, P_5) = [100/15] = 6$. Наконец, число всех натуральных чисел из первых 100, делящихся одновременно на 2, 3 и 5, равно $N(P_2, P_3, P_5) = [100/30] = 3$. Теперь, применяя формулу включений и исключений, получим

$$N(\overline{P}_2, \overline{P}_3, \overline{P}_5) = n - N(P_2) - N(P_3) - N(P_5) + N(P_2, P_3) +$$
$$+ N(P_2, P_5) + N(P_3, P_5) - N(P_2, P_3, P_5) =$$
$$= 100 - 50 - 33 - 20 + 16 + 10 + 6 - 3 = 26.$$

Пример. В группе 23 студента. Из них 18 знают английский язык, 9 — немецкий и 6 — оба языка. Сколько студентов в группе не знают ни одного языка? Сколько студентов знают только один язык?

Решение. Пусть S— множество всех студентов, $|S| = 23$. Назначим свойства элементам $s \in S$: P_1 — знание английского языка, P_2 — знание немецкого языка. $N(0)$ — количество студентов, не знающих языков (не обладают свойствами), $N(1)$ — количество студентов, знающих только один язык. Тогда

$$N(0) = 23 - N(P_1) - N(P_2) + N(P_1 \cdot P_2) = 23 - 18 - 9 + 6 = 2$$
$$N(1) = (N(P_1) - N(P_1 \cdot P_2)) + (N(P_2) - N(P_1 \cdot P_2)) = 18 - 6 + 9 - 6 = 15$$

Пример (задача о беспорядках). Перестановка $a_1 a_2 \ldots a_n$ чисел $1,2,\ldots,n$ называется беспорядком, если $a_i \neq i$ для всех $i = 1, 2, \ldots, n$. Таким образом, беспорядок – это перестановка, в которой ни один элемент не занимает «своего» места. Задача о беспорядках состоит в том, чтобы подсчитать число D_n

перестановок – беспорядков. Например, D_2=1 – два числа {1,2} образуют только один беспорядок (2,1). D_3=2, поскольку беспорядками из трех чисел {1,2,3} являются (3,1,2) и (2,3,1). Перечислим беспорядки из четырех элементов:

2143; 2341; 2413; 3142; 3412; 3421; 4123; 4312; 4321.

Значит, $D_4 = 9$.

Если перестановке назначить свойство – один предмет остается на своем месте, то таких перестановок будет $C_n^1(n-1)!$, поскольку предмет может находиться на любом из C_n^1 месте, а остальные элементы можно переставить $(n-1)!$ способами. Аналогично, перестановок в которых два предмета остаются на своем месте будет $C_n^2(n-2)!$ и т.д. Подставляя это в (3) приходим к

$$D_n = n! - C_n^1(n-1)! + C_n^2(n-2)! - \ldots + (-1)^n C_n^n$$

Последнее равенство можно переписать следующим образом:

$$\frac{D_n}{n!} = 1 - \frac{1}{1!} + \frac{1}{2!} + \ldots + (-1)^r \frac{1}{r!} + \ldots + (-1)^n \frac{1}{n!}.$$

Также ясно, что количество перестановок n различных предметов, при которых ровно k предметов стоят на своих первоначальных местах, выражается числом $D_{n,k} = C_n^k \cdot D_{n-k}$.

6.5 Метод рекуррентных соотношений

Последовательность $a_0, a_1, a_2, \ldots$ называют рекуррентной, если указана зависимость общего члена последовательности от предыдущих и заданы значения необходимого числа начальных членов.

Примерами рекуррентных последовательностей могут служить арифметические и геометрические прогрессии.

Члены геометрической прогрессии $a_0,\ a_1,\ a_2, \ldots$ со знаменателем q по определению связаны рекуррентным соотношением

$$a_{n+1} = q\, a_n.$$

и для однозначного ее определения надо задать начальное значение a_0.

Члены произвольной арифметической прогрессии $a_0,\ a_1,\ a_2, \ldots$ связаны рекуррентным соотношением $a_{n+1} = a_n + \Delta$ или, например, соотношением

$$a_{n+2} = 2a_{n+1} - a_n.$$

Действительно

$$a_n = a_0 + (n-1)\Delta$$

$$a_{n+1} = a_0 + n \cdot \Delta$$

$$a_{n+2} = a_0 + (n+1)\Delta = 2(a_0 + n \cdot \Delta) - (a_0 + (n-1)\Delta) = 2 \cdot a_{n+1} - a_n$$

Последовательность факториалов 1,2,6,…,n!,… определяется рекуррентным соотношением

$$a_{n+1} = (n+1)a_n$$

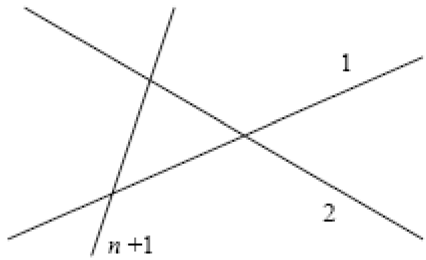
с заданием начального значения $a_0 = 1$.

Как рекуррентость может трактоваться формула

$$C_{n+1}^k = C_n^{k-1} + C_n^k,$$

связывающая биномиальные коэффициенты.

Пример. Рассмотрим задачу о разбиении плоскости прямыми. Пусть D_n – число областей, на которые разбивают плоскость n прямых общего положения (таких, что никакие три из них не пересекаются в общей точке и никакие две прямые не параллельны). Ясно, что $D_0 = 1$, $D_1 = 2$. Предположим, что на плоскости уже проведено n прямых, и посмотрим, сколько новых областей добавляется при проведении «новой» $n+1$-й прямой.



Каждую область, по которой проходит эта прямая, она рассекает на две. Таким образом, общее число областей увеличится на число областей, через которые проходит $n+1$-я прямая. Двигаясь по $n+1$-й прямой в одном направлении, мы пересечем границы областей n раз по числу «старых» прямых. Значит, $n+1$-я прямая пройдет через $n+1$ область (в последовательности область – граница – … – область – граница – область, число областей на единицу больше, чем число границ). В результате получаем рекуррентное соотношение

$$D_{n+1}=D_n+(n+1).$$

Чтобы найти замкнутое выражение для членов последовательности D_n, просуммируем следующие равенства:

$$D_1 = D_0 + 1;$$
$$D_2 = D_1 + 2;$$
....
$$D_n = D_{n-1} + n.$$

После сокращений получаем

$$D_n = D_0 + 1 + 2 + \ldots + n.$$

Следовательно,

$$D_n = 1 + \frac{n(n+1)}{2}$$

□

Рекуррентная последовательность $u_0, u_1, u_2, \ldots$ называется линейной однородной, если ее члены связаны соотношением вида

$$u_{n+r} = a_1 u_{n+r-1} + a_2 u_{n+r-2} + \ldots + a_r u_n, \qquad (1)$$

где a_i, $i=1,2,\ldots,r$, – постоянные, не зависящие от n. Соотношение (1) называется линейным однородным рекуррентным уравнением порядка r.

Пример. Уравнение первого порядка с постоянными коэффициентами имеет вид

$$y_n - a\,y_{n-1} = 0 \text{ и задается начальное значение } y_0 = y^0. \qquad (2)$$

Последовательные вычисления дают

$$y_n = a\cdot y_{n-1} = a\cdot a\cdot y_{n-2} = \ldots = a^n y_0 = a^n y^0.$$

Т.о. геометрическая прогрессия $y^0\cdot a^n$ является решением линейного однородного рекуррентного уравнения первого порядка.

Пример. Уравнение первого порядка с переменными коэффициентами имеет вид $y_{n+1} - a_n y_n = 0$ с начальным значением $y_0 = y^0$. Тогда

$$y_1 = a_0\,y_0$$
$$y_2 = a_1\,y_1 = a_1 a_0 y_0$$
$$y_3 = a_2\,y_2 = a_2 a_1 a_0 y_0$$
$$\ldots$$
$$y_n = a_{n-1}\cdots a_2 a_1 a_0 y_0 = y_0 \prod_{k=0}^{n-1} a_k$$

где использовано обозначение $\prod_{k=0}^{n} a_k = a_0\cdot a_1\cdot a_2\cdot\ldots\cdot a_n$.

Пример. Линейное неоднородное уравнение 1-го порядка имеет вид $y_{n+1} - a\,y_n = f_n$, для которого надо задать начальное значение $y_0 = y^0$ и последовательность значений f_n. Имеем

$$y_1 = a\cdot y_0 + f_1$$
$$y_2 = a\cdot y_1 + f_2 = a\left(a\cdot y_0 + f_1\right) + f_2 = a^2 y_0 + \left(af_1 + f_2\right)$$
$$y_3 = a\cdot y_2 + f_3 = a\left(a^2 y_0 + \left(af_1 + f_2\right)\right) + f_3 = a^3 y_0 + \left(a^2 f_1 + af_2 + f_3\right)$$

Ясно, что

$$y_n = a^n y_0 + \sum_{k=1}^{n} a^{n-k} f_k \qquad (3)$$

Здесь мы имеем сумму решения однородного уравнения (2) и некоторого частного решения неоднородного уравнения $z_n = \sum_{k=1}^{n} a^{n-k} f_k$. Действительно

$$z_n - a\cdot z_{n-1} = \left(a^{n-1} f_1 + a^{n-2} f_2 + \ldots + af_{n-1} + f_n\right) -$$
$$- a\left(a^{n-2} f_1 + a^{n-3} f_2 + \ldots + af_{n-2} + f_{n-1}\right) = f_n$$

Пример. Требуется подсчитать количество двоичных слов длины n, в которых единицы не могут стоять на соседних местах. Будем называть такие слова правильными и обозначим через A_n число правильных слов длины n.

Разобьем множество правильных слов длины n на два класса: слова оканчивающиеся на ноль и слова, оканчивающиеся на единицу. Количество слов в этих классах обозначим $A_n^{(0)}$ и $A_n^{(1)}$ соответственно. Очевидно, что $A_n = A_n^{(0)} + A_n^{(1)}$. У слова, оканчивающегося на ноль первые n-1 символов образуют правильное слово длины n-1. Следовательно $A_n^{(0)} = A_{n-1}$. Если правильное слово длины n оканчивается на единицу, то предыдущий символ этого слова должен быть нулем, а первые n-2 символа будут образовывать правильное слово длины n-2. Поэтому $A_n^{(1)} = A_{n-2}$. В результате мы имеем соотношение

$$A_n = A_{n-1} + A_{n-2}$$

Как видим, решение нашей задачи свелось к рекуррентному уравнению. В переводе – возвратному, так как для подсчета интересующей нас величины для некоторого n нужно возвратиться к предыдущим значениям этой величины. Здесь легко видеть, что $A_1 = 2$ и $A_2 = 2$. Поэтому $A_3 = 2 + 2 = 4$, $A_4 = 2 + 4 = 6$ и т.д.

□

В общем случае рекуррентное соотношение имеет вид

$$a_n = F(a_{n-1}, a_{n-2}, \ldots, a_{n-k}) \tag{4}$$

Как правило, мы будем иметь дело с уравнением вида

$$a_n = c_1 a_{n-1} + c_2 a_{n-2} + \ldots + c_k a_{n-k} \tag{5}$$

где $c_1, c_2, \ldots, c_k$ - заданные числа. Такое соотношение называют линейным рекуррентным уравнением k – го порядка с постоянными коэффициентами. Соотношение (4) или (5) мы будем рассматривать как уравнение относительно неизвестной функции $A(n)=A_n$ и каждую последовательность

$$\tilde{A} = (A_0, A_1, \ldots, A_n, \ldots)$$

для которой выполняется соотношение (4) или (5) будем называть решением рекуррентного соотношения.

Лемма 1. Если последовательность $\tilde{A} = (A_0, A_1, \ldots, A_n, \ldots)$ является решением рекуррентного соотношения (5) и C любое число, то и последовательность $C\tilde{A} = (CA_0, CA_1, \ldots, CA_n, \ldots)$ также является решением рекуррентного соотношения (5).

Доказательство. Подставив вторую последовательность в (5) и поделив на C , убеждаемся, что и вторая последовательность является решением. Если же C=0, то последовательность $C\,\tilde{A}$ состоит из одних нулей и, очевидно, также удовлетворяет соотношению (5).

Лемма 2. Если $\tilde{A} = (A_0, A_1, \ldots, A_n, \ldots)$ $\tilde{B} = (B_0, B_1, \ldots, B_n, \ldots)$ являются решениями рекуррентного соотношения (5), то последовательность

$$\tilde{C} = \tilde{A} + \tilde{B} = (A_0 + B_0, A_1 + B_1, \ldots, A_n + B_n, \ldots)$$

также является решением рекуррентного соотношения (5).

Доказательство. Так как $\tilde{A}$ и $\tilde{B}$ являются решениями, то

$$A_n = c_1 A_{n-1} + c_2 A_{n-2} + \ldots + c_k A_{n-k}$$
$$B_n = c_1 B_{n-1} + c_2 B_{n-2} + \ldots + c_k B_{n-k}$$

Сложив эти соотношения, получаем

$$A_n + B_n = c_1(A_{n-1} + B_{n-1}) + c_2(A_{n-2} + B_{n-2}) + \ldots + c_k(A_{n-k} + B_{n-k}),$$

т.е.

$$C_n = c_1 C_{n-1} + c_2 C_{n-2} + \ldots + c_k C_{n-k}$$

□

Рекуррентные уравнения для нас важны, поскольку часто решение одной комбинаторной задачи удается свести к решению аналогичных задач меньшей размерности. Тем самым решение сложной задачи можно получить, последовательно находя решения более легких задач, и далее, пересчитывая по рекуррентным соотношениям, находить решение трудной задачи. Формула (4) является рекуррентным соотношением между элементами последовательности чисел $a_{n-k},\ldots,a_n$. Т.о. рекуррентное соотношение позволяет по известным значениям $a_{n-k},\ldots,a_n$ вычислить значение a_n., но первые несколько значений $a_0,\ldots,a_{k-1}$ нужно знать заранее. Как мы видели в примерах, иногда удается получить из рекуррентного соотношения общую формулу для вычисления a_n по номеру n. Тогда можно сразу вычислить окончательный результат без вычисления всех предыдущих результатов.

Пример. Вернемся к рекуррентному соотношению $a_n=a_{n-1}+ a_{n-2.}$ Оно задает так называемые числа Фибоначчи F_0, F_1, F_2, F_3, … , если положить F_0 =0 и F_1=1. Тогда последовательность решений будет иметь вид 0, 1, 1, 2, 3, 5, 8, 13, 21, …Найдем общую формулу для чисел Фибоначчи. Будем искать ее в виде $a_n = \lambda^n$. Подстановка в рекуррентное соотношение дает $\lambda^n = \lambda^{n-1} + \lambda^{n-2}$. Разделим обе части равенства на $\lambda^{n-2} \neq 0$. Получим квадратное уравнение $\lambda^2 - \lambda - 1 = 0$. Найдем его корни

$$\lambda_{1,2} = \frac{1 \pm \sqrt{5}}{2}.$$

В соответствии с леммой 1 решением являются $a_n = C_1\lambda_1^n$ и $a_n = C_1\lambda_2^n$, где C_1 и C_2 произвольные постоянные. В соответствии с леммой 2 решением будет их сумма. Т.о. общим решением будет линейная комбинация

$$a_n = C_1{\lambda_1}^n + C_2{\lambda_2}^n = C_1\left(\frac{1+\sqrt{5}}{2}\right)^n + C_2\left(\frac{1-\sqrt{5}}{2}\right)^n.$$

Произвольные постоянные найдем из начальных условий a_0=0, a_1=1. Имеем

$$a_0 = 0 = C_1 + C_2 \Rightarrow C_2 = -C_1$$

и

$$a_1 = 1 = C_1\frac{1+\sqrt{5}}{2} + C_2\frac{1-\sqrt{5}}{2} = C_1\left(\frac{1+\sqrt{5}}{2} - \frac{1-\sqrt{5}}{2}\right) = C_1\sqrt{5}$$

Тогда $C_1 = \frac{1}{\sqrt{5}}, C_2 = -\frac{1}{\sqrt{5}}$. Окончательно общее решение рекуррентного соотношения будет иметь вид

$$a_n = \frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right),$$

который называется формулой Бине.

□

Заметим, что деление отрезка такое, что отношение длин полученных частей равняется $\frac{-1+\sqrt{5}}{2}$ называют золотым сечением. Поэтому это число часто называют отношением золотого сечения.

Рассмотренный пример позволяет сформулировать общий прием решения линейных рекуррентных уравнений.

Теорема. Пусть $a_n + c_1 a_{n-1} + \ldots + c_k a_{n-k} = 0$ – линейное однородное рекуррентное уравнение с постоянными коэффициентами $c_i = Const$ $(i = 1,2,..k)$, и пусть λ — корень характеристического уравнения $\lambda^k + c_1 \lambda^{k-1} + \ldots + c_k = 0$. Тогда:

1) последовательность с общим членом $C\lambda^n$, где C — произвольная константа, удовлетворяет исходному однородному рекуррентному уравнению;
2) если $\lambda_1, \lambda_2, \ldots, \lambda_k$ — простые корни характеристического уравнения, то общее решение рекуррентного соотношения имеет вид

$$a_n = C_1\lambda_1^n + C_2\lambda_2^n + \ldots + C_k\lambda_k^n,$$

где C_i, *i= 1,2, ..., k,* — произвольные постоянные;
3) если λ_i — корень кратности r_i, $i = 1, \ldots, s$, характеристического уравнения, то общее решение рекуррентного соотношения имеет вид

$$a_n = \sum_{i=1}^{s}(C_{i1} + C_{i2}n + \ldots + C_{ir_i}n^{r_i-1})\lambda_i^n,$$

где $C_{i\,j}$ — произвольные постоянные.

Пример. Рассмотрим соотношение $a_n = 3\cdot a_{n-1} - 2\cdot a_{n-2}$, $a_1 = 1, a_2 = 3$. Оно является линейным рекуррентным соотношением с постоянными коэффициентами $a_n - 3\cdot a_{n-1} + 2\cdot a_{n-2} = 0$. Составим характеристическое уравнение $\lambda^2 - 3\lambda + 2 = 0$. Его корни $\lambda_{1,2} = 1{,}2$. Поэтому общее решение имеет вид $a_n = C_1 + C_2 2^n$. С учетом начальных условий имеем

$$\begin{cases} C_1 + 2\,C_2 = 1 \\ C_1 + 4\,C_2 = 3 \end{cases}$$

Тем самым $C_1 = -1$, $C_2 = 1$. Окончательное решение будет $a_n = 2^n - 1$.

Пример. Рассмотрим уравнение $a_n - 4\cdot a_{n-1} + 4\cdot a_{n-2} = 0$. Составим характеристическое уравнение $\lambda^2 - 4\lambda + 4 = 0$ или $(\lambda - 2)^2 = 0$. Его корень $\lambda = 2$ имеет вторую кратность. Поэтому общее решение имеет вид $a_n = C_1 2^n + C_2 n\, 2^n$.

■

Для неоднородных линейных рекуррентных соотношений имеет место

Теорема. Пусть $a_n + c_1 a_{n-1} + \ldots + c_k a_{n-k} = f_n$ – неоднородное линейное рекуррентное уравнение, Его общее решение представляется в виде суммы общего решения соответствующего однородного уравнения и некоторого частного решения неоднородного уравнения $a_n = a_n^0 + a_n^H$.

Доказательство. Мы имеем

$$a_n^0 + c_1 a_{n-1}^0 + \ldots + c_k a_{n-k}^0 = 0 \quad \text{(однородное уравнение)}$$

$$a_n^H + c_1 a_{n-1}^H + \ldots + c_k a_{n-k}^H = f_n \quad \text{(неоднородное уравнение)}$$

Сложим эти соотношения

$$\left(a_n^0 + a_n^H\right) + c_1\left(a_{n-1}^0 + a_{n-1}^H\right) + \ldots + c_k\left(a_{n-k}^0 + a_{n-k}^H\right) = f_n$$

Т.о. $a_n = a_n^0 + a_n^H$ является решение неоднородного уравнения

■

Пример. Рассмотрим уравнение $a_n = 3 \cdot a_{n-1} - 2 \cdot a_{n-2} + 2 \cdot 3^n$, $a_1 = 30, a_2 = 86$. Оно является неоднородным рекуррентным уравнением 2 – го порядка

$$a_n - 3 \cdot a_{n-1} + 2 \cdot a_{n-2} = 2 \cdot 3^n$$

Соответствующее ему однородное уравнение было рассмотрено в предыдущем примере и имело общее решение $a_n^0 = C_1 + C_2 2^n$. Теперь найдем частное решение неоднородного уравнения. Будем искать его в виде $a_n^H = A \cdot 3^n$, где коэффициент A будем подбирать. Подстановка в исходное неоднородное уравнение дает

$$A \cdot 3^n - 3 \cdot A \cdot 3^{n-1} + 2 \cdot A \cdot 3^{n-2} = 2 \cdot 3^n \text{ или } 2 \cdot A \cdot 3^{n-2} = 2 \cdot 3^n$$

Сокращая на 3^{n-2}, получаем A=9. Т.о. общее решение неоднородного уравнения будет иметь вид $a_n = C_1 + C_2 2^n + 9 \cdot 3^n$. Из начальных условий находим

$$30 = C_1 + C_2 2^1 + 9 \cdot 3^1$$

$$86 = C_1 + C_2 2^2 + 9 \cdot 3^2$$

Решение системы уравнений дает C_1=1, C_2=1. Окончательно, решение имеет вид $a_n = 1 + 2^n + 9 \cdot 3^n$.

Пример. Найти частное решение уравнения $u_n - 3 \cdot u_{n-1} + 2 \cdot a_{n-2} = n$.

Решение будем искать в виде $u_n = An^2 + Bn + C$. Подставим это выражение в уравнение.

$$A(n+2)^2 + B(n+2) + C - 3 \cdot \left(A(n+1)^2 + B(n+1) + C\right) + 2 \cdot \left(An^2 + Bn + C\right) = n.$$

После упрощений получаем

$$A \cdot (-2n + 1) + B \cdot (-1) = n.$$

Коэффициент C сокращается, и его можно взять любым, например, $C = 0$. Последнее соотношение является тождеством и оно должно выполняться для любого n. Поэтому должны выполняться равенства

$$\begin{cases} A \cdot (-2n) = n \\ A - B = 0 \end{cases}$$

Отсюда получаем $A=-\frac{1}{2}, B=-\frac{1}{2}$. Т.о. частное решение имеет вид $u_n^{частное}=-\frac{1}{2}n^2-\frac{1}{2}n=-\frac{1}{2}n(n+1)$.

7. Элементы теории графов.

7.1 Введение

Линейность или последовательность представления информации является характерной чертой большинства информационных систем, в том числе человеческого языка и языков программирования. Однако линейное (последовательное) представление информации не всегда является наиболее простым способом и с точки зрения человеческого восприятия. Например, график функции часто информативнее, чем формула. Графы это еще один способ представления информации, который в некоторых случаях является более удобным, чем другие способы.

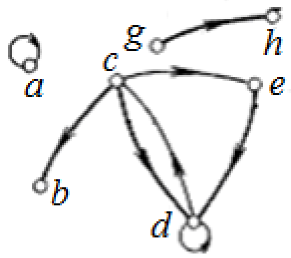
Если необходимо изобразить в наглядной форме систему взаимосвязанных объектов, то прибегают к такому построению: на плоскости рисуют точки, представляющие объекты, и некоторые из них соединяют линиями, представляющими связи. Рисунок (графический объект), который получается в результате такого построения, называется графом. Примерами графов являются блок – схемы алгоритмов, соединений элементов электрической сети, отношение родителей и потомков на множестве людей, сеть дорог на карте и т.д.

При построении графов, как и во всех современных математических теориях, пользуются сокращенными обозначениями, дающими значительную экономию мышления и делающими орудие исследования более гибким и эффективным. *Обозначим через X непустое конечное множество точек плоскости, называемых вершинами и через U конечное множество соединяющих их линий, называемых ребрами (или дугами, если на ребре задано направление). Тогда математически граф можно определить как пару множеств X и U*:

$$G=(X, U).$$

Одну и ту же систему объектов и связей между ними можно изобразить по – разному. Поэтому существуют различные виды графов и понятий, с ними связанных. Неотъемлемыми атрибутами графов являются вершины и соединяющие их ребра или дуги. Каждое ребро (дуга) связывает (соединяет) пару вершин. *Если на всех ребрах графа задана ориентация, то его называют ориентированным графом (орграфом), иначе – неориентированным*.

На следующем рисунке изображен ориентированный граф, вершинами которого являются точки *a, b, c, d, e, g, h,* а дугами – отрезки (*a, a*), (*c, b*), (*c, d*), (*d, c*), (*d,d*), (*c,e*), (*e, d*), (*g, h*).



Общий вид графа

Иногда удобно дать орграфу другое определение. Можно считать, что множество направленных дуг U, соединяющих элементы множества X, отображает это множество само в себя. Поэтому можно считать орграф заданным, если даны множество его вершин X и способ отображения Г множества X в X. Таким образом орграф G есть пара (X, Г), состоящая из множества X и отображения Г, заданного на этом множестве, т.е.

$$G=(X, \Gamma)$$

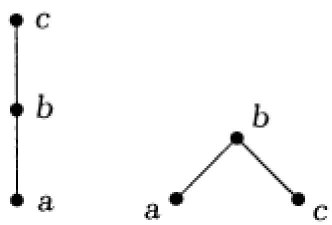
Для орграфа, изображенного на предыдущем рисунке, отображение Г определяется следующим образом:

$$\Gamma a = a; \quad \Gamma b = \varnothing; \quad \Gamma c = \{b,\ d,\ e\};$$
$$\Gamma d = \{d,\ c\}; \quad \Gamma e = d; \quad \Gamma g = h; \quad \Gamma h = \varnothing.$$

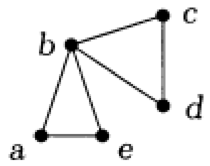
Нетрудно видеть, что такое определение графа полностью совпадает с определением отношения на множестве, которое мы в теме 3 дали следующим образом: всякое подмножество $R \subset A^2$ декартового квадрата A^2 непустого множества A называется (бинарным) отношением, заданным на множестве A. Мы говорили, что элемент *a* из A находится в отношении R к элементу b из A, и писали *a* R b, если пара $(a,b) \in R$.

Обычно конечный неориентированный граф изображают в виде диаграммы, на которой вершины обозначаются точками, а ребра, соединяющие их,— линиями между этими точками. Вершины графа часто обозначают буквами, например *a* и *b*, а ребра, их соединяющие, парой букв $\{a,\ b\}$. Наоборот, если $\{a,\ b\}$ — ребро, тогда вершины *a* и *b* называются концами ребра. Если на ребре задано направление, то его (дугу) мы будем обозначать (a,b), используя круглые скобки.

Пример. Граф с множеством вершин V = $\{a,b,c\}$ и множеством ребер E=$\{\{a,\ b\},\ \{b,\ c\}\}$ может быть изображен, как показано на следующем рисунке



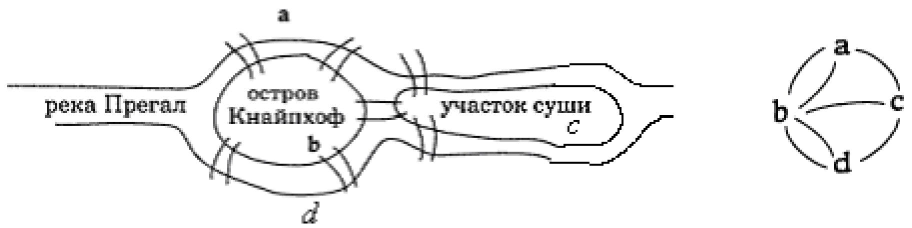
Пример. Граф с множеством вершин V = $\{a,b,c,d,e\}$ и множеством ребер E = $\{\{a,\ b\},\ \{a,\ e\},\ \{b,\ e\},\ \{b,\ d\},\ \{b,\ c\},\ \{c,\ d\}\}$ может быть изображен диаграммой, показанной на следующем рисунке.



Графы, определенные нами в двух последних примерах, *называют простыми графами, поскольку имеется ограничение на существование только одного ребра между двумя вершинами*. Это дает возможность представлять любое ребро как множество из двух элементов – вершин ребра.

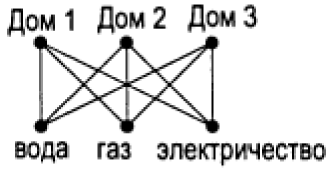
Ребро, которое соединяет вершину саму с собой, называется петлей. *Если в графе допускается наличие петель, то он называется графом с петлями. Если в графе допускается наличие более одного ребра между двумя вершинами, то он называется мультиграфом. Если же допускается как наличие петель, так и существование более одного ребра между двумя вершинами, то мы будем называть такой объект псевдографом*.

Теория графов берет начало с решения знаменитым математиком Эйлером задачи о кенигсбергских мостах в 1736 году. Задача возникла в прусском городе Кенигсберг. Его жителям нравилось гулять по дорожке, которая включала семь мостов через реку. Люди интересовались, могут ли они, начав путь с одного участка суши, обойти все мосты, проходя по каждому лишь один раз, вернуться в точку начала пути. Этот район и семь мостов показаны на следующем рисунке слева.

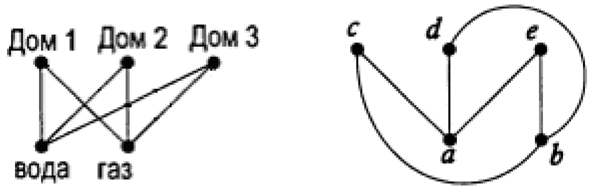


Эйлер построил мультиграф, показанный на предыдущем рисунке справа. В этом мультиграфе участки суши Эйлер изобразил как вершины, а дорожки через мосты — как ребра. В таком случае задача приобретает следующую формулировку: начиная с любой вершины, проходя по каждому ребру только один раз, вернуться в исходную вершину. Далее будет показано, как на основе понятия мультиграфа Эйлер решил эту задачу. Приведенный пример иллюстрирует полезность математических абстракций, по крайней мере, в двух аспектах. Во-первых, абстракция устраняет в содержании проблемы несущественные детали и концентрирует внимание на тех понятиях, которые действительно важны. Во-вторых, абстракция часто выявляет кажущееся различие задач, совпадающих по сути.

Примером другой проблемы, которую можно промоделировать на основе теории графов, является задача о снабжении трех домов тремя видами коммунальных услуг. Согласно условию задачи к каждому из трех домов необходимо подключить три вида коммунальных услуг, например, воду, газ и электричество, посредством подземных линий труб и кабелей. Вопрос состоит в том, можно ли обеспечить эти три дома коммунальными услугами без пересечения линий снабжения. Граф, моделирующий данную задачу, показан на следующем.

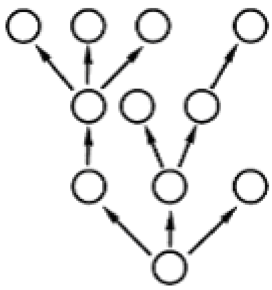


Если бы проблема заключалась в предоставлении двух видов коммунальных услуг трем домам или трех видов коммунальных услуг двум домам, то ее можно было бы решить, используя в каждом случае граф, изображенный на следующем рисунке, в котором линии снабжения не будут пересекаться.

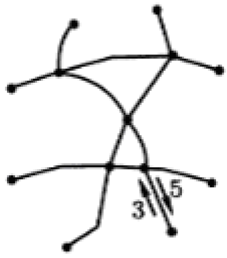


Аналогичная, но более практичная задача возникает при создании микросхем. В них пересечение проводов на каждом уровне является недопустимым.

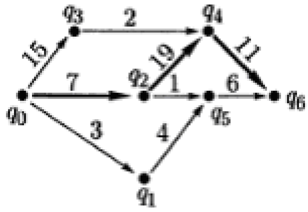
На следующем рисунке изображен граф, который имеет определенную внутреннюю структуру и называется деревом. Он будет генеалогическим деревом, если в кружках написать имена людей, наследующих графский (дворянский) титул. В каждую вершину, кроме первой, входит только одна дуга. В дереве нет циклов.



Еще одним примером графа является транспортная сеть. Это, например, сеть дорог, трубопроводная, железнодорожная, информационная и т.д. Вершинами графа могут быть города, аэропорты, железнодорожные станции, телефонные станции и т.д. Дугами графа – односторонние дороги, трубопроводы, линии электропередач и т.д. На дугах задают нагрузки, которыми могут быть пропускная способность дороги, стоимость проезда, протяженность, количество перевозимого груза (грузооборот) и т.д.



Еще одним примером графа является сетевой график. Граф, описывающий некоторый технологический процесс (проект создания какой-либо системы), называется сетевым. Вершины графа – главные события процесса.



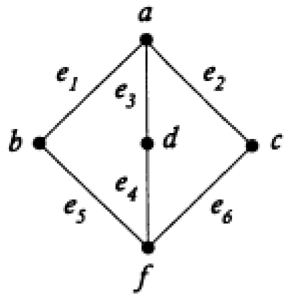
Здесь событие q_0 – начало выполнения проекта. Из вершины q_0 дуги только выходят. Вершина графа q_6 – событие завершения проекта. В вершину q_6 дуги только входят. Каждая дуга (q_iq_j) соответствует некоторой операции (работе). Нагрузка на дуге означает длительность по времени данной операции. Жирной стрелкой выделен критический путь в сетевом графике от q_0 до q_6. Среди всех путей из q_0 в q_6 он самый длительный по времени. Его длительность называется критическим временем. Это есть время выполнения всего проекта, т. е. минимальное время, за которое можно выполнить весь проект.

7.2 Основные понятия теории графов

Пусть $\{a, b\}$ *— ребро, тогда вершины* a *и* b *называются концами ребра* $\{a,b\}$. *Ребро* $\{a, b\}$ *называют также инцидентным к вершинам* a *и* b. *Обратно, говорят, что вершины* a *и* b *инцидентны к ребру* $\{a, b\}$. *Две вершины называются смежными, если они являются концами ребра, или, что-то же самое, если они инцидентны к одному ребру. Два ребра называются смежными, если они инцидентны к общей вершине.*

Степенью $\deg(v)$ *вершины* v *называется количество ребер, инцидентных этой вершине. Вершина степени 0 называется изолированной.*

Пример. В графе, показанном на следующем рисунке, вершины a и c – смежные и e_1, e_2, e_3 - смежные ребра.



Однако, вершины a и f смежными не являются, а e_2 и e_5 не являются смежными ребрами. Вершины b, c и d имеют степень 2, в то время как вершины a и f имеют степень 3.

Теорема. Сумма степеней вершин графа всегда четная.

Доказательство. Поскольку каждое ребро графа имеет два конца, степень каждого конца увеличивается на 1 за счет одного ребра. Таким образом, в сумму степеней всех вершин каждое ребро вносит 2 единицы, поэтому сумма должна в два раза превышать количество ребер. Поэтому сумма будет четным числом.

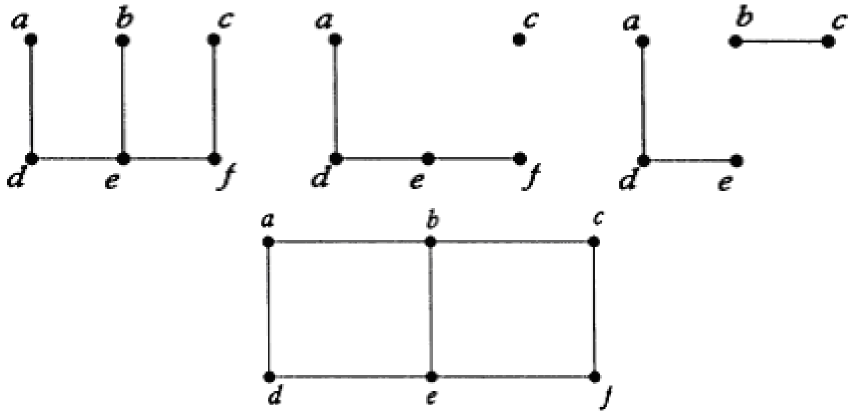
Теорема. В любом графе количество вершин нечетной степени четно.

Доказательство. Доказательство проведем методом от противного. Если теорема не верна, то имеется нечетное количество вершин, степени которых нечетны. Но сумма степеней вершин с четными степенями четна. Сумма степеней всех вершин есть сумма степеней вершин с нечетными степенями плюс сумма степеней вершин с четными степенями, т.е. будет нечетным числом. Но это противоречит предыдущей теореме.

□

Граф $G'(V',E')$ называется подграфом графа $G(V,E)$, обозначается $G'(V',E') \le G(V,E)$, если $V' \subseteq V$ и $E' \subseteq E$. Таким образом, каждая вершина в G' является вершиной в G, и каждое ребро в G' является ребром в G.

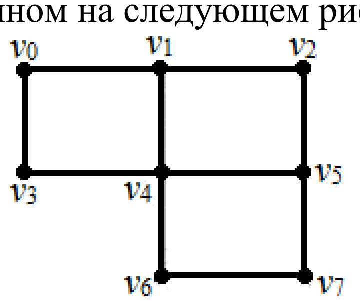
Пример. Графы, изображенные на следующем рисунке сверху, являются подграфами графа на нижнем рисунке.



Путь (маршрут) в графе — это совокупность ребер, которые объединены вместе вершинами так, что вдоль них можно двигаться по графу. В определении, приведенном ниже, свойства пути очерчены более формально.

Пусть $G=G(V,E)$ — граф с вершинами $v_0, v_1, v_2, \ldots, v_k \in V$ и ребрами $e_1, e_2, \ldots, e_k \in E$. Путем длины k из v_0 в v_k называется последовательность $v_0 e_1 v_1 e_2 v_2 \ldots v_{k-1} e_k v_k$ такая, что $e_i = (v_{i-1}, v_i)$. Таким образом, путь длины k имеет k ребер. Для краткости мы будем обозначать путь, указывая только вершины, например, $v_0 v_1 v_2 \ldots v_{k-1} v_k$. Каждые два последовательных ребра пути имеют общую вершину, поэтому являются смежными. *Простым путем из v_0 в v_k называется путь, в котором нет повторяющихся вершин*. Для неориентированного графа вместо термина путь часто используют термин цепь.

Пример 1. В графе, приведенном на следующем рисунке



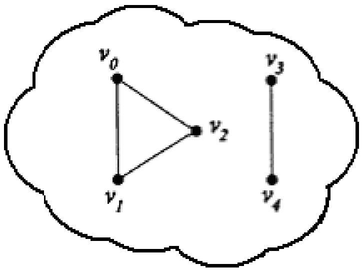
из v_0 в v_7 ведут пути $v_0v_1v_2v_5v_7$, $v_0v_1v_2v_5v_4v_1v_2v_5v_7$, $v_0v_1v_4v_5v_4v_5v_7$ и $v_0v_3v_4v_6v_7$ длины 4, 8, 6 и 4 соответственно. Пути $v_0v_1v_2v_5v_7$ и $v_0v_3v_4v_6v_7$ являются простыми.

Некоторые пути, приведенные в этом примере, можно сократить. Например, путь $v_0v_1v_2v_5v_4v_1v_2v_5v_7$ можно сократить до $v_0v_1v_2v_5v_7$. Поскольку вершина v_1 повторялась, необходимо удалить часть пути между двумя появлениями вершины v_1 и после первого появления вершины v_1 переходить сразу к v_2, стоящему после второго появления вершины v_1. Аналогично, путь $v_0v_1v_4v_5v_4v_5v_7$ можно сократить до $v_0v_1v_4v_5v_7$. Таким образом, если путь включает какую-либо вершину v_i более чем один раз, его можно сократить, удалив v_i и вершины, лежащие на пути между двумя появлениями вершины v_i. Можно продолжать действовать таким образом, пока не будет исключено повторение любой из вершин. Проведенные рассуждения дают возможность сформулировать следующую теорему.

Теорема. Пусть $G = G(V,E)$ — граф. Если существует путь из вершины v_i в вершину v_j, тогда существует простой путь из вершины v_i в вершину v_j.

Граф G называется связным, если имеется путь между любыми двумя его различными вершинами.

Пример. Граф, приведенный на следующем рисунке не связный



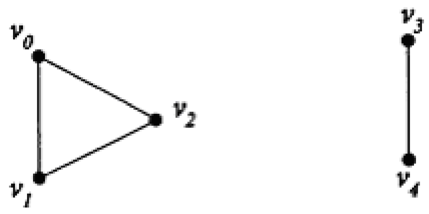
Например, нет пути из v_0 в v_3 и между v_2 и v_4.

□

Комбинируя определение простого пути и предыдущую теорему, приходим к следствию: граф G является связным тогда и только тогда, когда между любыми двумя его вершинами существует простой путь.

Пусть G=G(V,E) — граф. Подграф G' графа G называется компонентой графа G, если он является максимальным связным подграфом графа G. Это значит, что. если G' — непустой связный граф и G" какой – то связный подграф графа G для которого G' < G", то G' = G".

Пример. Два графа, изображенные на следующем рисунке, являются компонентами графа, представленного на предыдущем рисунке.

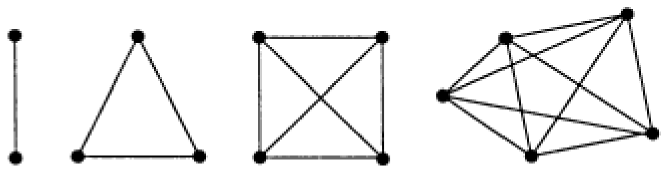


Пусть G=(V,E) — граф. *Циклом называется путь ненулевой длины, соединяющий вершину саму с собой и не содержащий повторяющихся ребер.* Простым циклом называется цикл, соединяющий вершину v саму с собой и не содержащий повторяющихся вершин, кроме v. Цикл называется n-циклом, если он содержит n ребер и n различных вершин. Для ориентированного графа вместо термина *цикл* часто используется термин *контур*.

Пример. Если опять рассмотреть граф из примера 1, то в нем пути $v_0v_1v_4v_3v_0$, $v_0v_1v_2v_5v_7v_6v_4v_3v_0$, $v_1v_2v_5v_7v_6v_4v_1$ и $v_0v_1v_4v_5v_7v_6v_4v_3v_0$ являются циклами. При этом все циклы, кроме последнего, являются простыми.

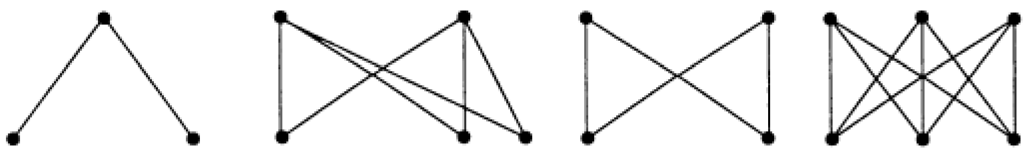
Граф называется полным, если любые две его вершины соединены ребром. Полный граф с n вершинами обозначается через K_n.

Пример. На следующем рисунке показаны графы K_2, K_3, K_4 и K_5 .



Граф G=(V,E) называется двудольным, если V можно представить как объединение непересекающихся множеств $V = A \bigcup B$, так что каждое ребро имеет вид $\{a,b\}$, где $a \in A, b \in B$. Таким образом, каждое ребро связывает вершину из A с вершиной из B, но никакие две вершины из A или две вершины из B не являются связанными. Двудольный граф называется полным двудольным графом $K_{m,n}$, если A содержит m вершин, B содержит n вершин и для каждого $a \in A, b \in B$ имеется связывающее их ребро.

Пример. Графы $K_{1,2}$, $K_{2,3}$, $K_{2,2}$, $K_{3,3}$ приведены по порядку на следующем рисунке

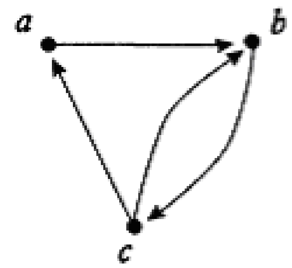


Во многих случаях необходимы графы, у которых ребра представляют собой «улицу с односторонним движением». Это означает, что если рассматривается ребро, выходящее из вершины a в вершину b, то его нельзя рассматривать выходящим из вершины b в вершину a. Например, если граф моделирует поток нефти в трубопроводе или отношение родителей и потомков, то на ребрах между вершинами графа следует указывать ориентацию. Для трубопровода направление потока нефти существенно, также как важно в отношении отцовства – материнства явно указывать родителя и потомка. Такие графы принято называть ориентированными.

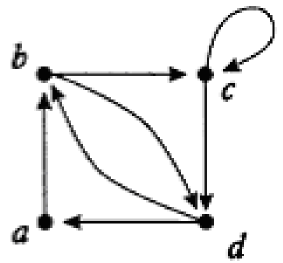
Ориентированный граф или орграф G обозначается через *G(V, E)* и состоит из множества V вершин и множества E упорядоченных пар элементов из V, называемого множеством ориентированных ребер. Элемент множества E называется ориентированным ребром или дугой. Если $(a,b) \in E$, то a называется начальной вершиной ребра (a,b), а b называется конечной вершиной. Понятие ориентированного графа допускает наличие петель. Кроме того, как говорилось ранее, орграф можно интерпретировать как отношение на множестве.

Ребро (a, b) ориентированного графа обозначается на диаграмме стрелкой, направленной из a в b. Отметим, что в неориентированном графе ребро представляется двухэлементным подмножеством $\{a,b\}$, чтобы подчеркнуть, что отношение симметрично, в то время как в ориентированном графе ребро представлено упорядоченной парой (a,b), чтобы акцентировать важность порядка.

Пример. Орграф, у которого V = {a, b, c} и E = {(a, b), (b, c), (c, b), (c, a)}, изображен на следующем рисунке.



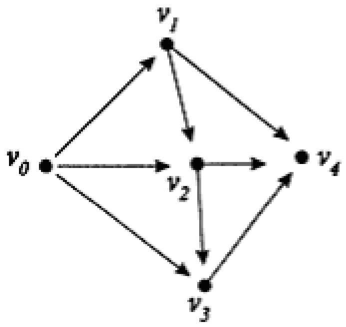
Пример. Орграф, у которого V = {a, b, c, d} и E = {(a, b), (b, c), (c, c), (b, d), (d, b), (c, d), (d, a)}, изображен на следующем рисунке



Определение смежности и инцидентности вершин и дуг для орграфов такое же, как для простых графов.

В орграфе степенью выхода вершины v называется количество ребер, для которых v является начальной вершиной, обозначается outdeg(v). Степенью входа вершины v называется количество ребер, для которых v является конечной вершиной, обозначается indeg(v). Если indeg(v) = 0 и outdeg(v)>0, то вершина v называется источником. Если outdeg(v) = 0 и indeg(v)>0, то вершина v называется стоком.

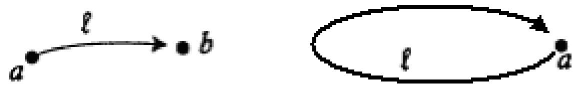
Пример. В орграфе приведенном на следующем рисунке indeg(v_0) = 0, indeg(v_1) = 1, indeg(v_2) = 2, indeg(v_3) = 2, Также outdeg(v_0) = 3, outdeg(v_1) = 2, outdeg(v_2) = 2, outdeg(v_3) = 1 и outdeg(v_4) = 0. Таким образом, вершина v_0 — источник, а вершина v_4 — сток.



Ориентированный граф с более чем одним ребром из одной вершины в другую называется мультиграфом, или, более точно, ориентированным мультиграфом. Для таких графов имеет смысл помечать ребра, идущие из одной вершины в другую, поскольку не имеет смысл иметь два ребра между вершинами, если нет возможности их различать. Если каждое ребро помечено, будем говорить, что это размеченный ориентированный, или просто размеченный граф, с пониманием того, что это ориентированный граф. Формально размеченный граф можно определить следующим образом.

Размеченный граф $G = G(V,L,E)$ представляет собой множество вершин V, множество меток L и множество дуг E и каждая дуга $e \in E$ графа G имеет вид (a, l, b), где l — метка, а a, b — вершины.

Графически ребро $e=(a,l,b)$ размеченного графа обозначается так, как на следующем рисунке слева (значком возле изображающей дугу линии) или как на рисунке справа, если ребро — петля.

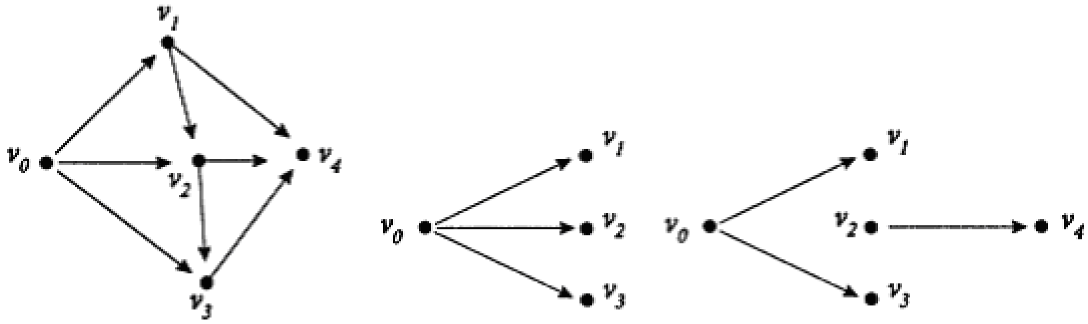


Понятие ориентированного пути в ориентированном графе вводится по аналогии с понятием пути в графе. Разница лишь в том, что, перемещаясь вдоль ориентированного пути, двигаться нужно в направлении, задаваемом ориентацией ребер. Т.о путем в орграфе называется такая последовательность дуг $\mu=(u_1,u_2,...,u_k)$, в которой конец каждой предыдущей дуги совпадает с началом последующей. Длиной пути $\mu=(u_1,u_2,...,u_k)$ называется число $l(\mu)=k$, равное числу дуг, составляющих путь μ. Путь, в котором никакая дуга не встречается дважды, называется простым. Путь, в котором никакая вершина не встречается дважды, называется элементарным. Контур – это конечный путь $\mu=(x_1,x_2,...,x_k)$, у которого начальная вершина x_1 совпадает с конечной x_k. При этом контур называется элементарным, если все его вершины различны (за исключением начальной и конечной, которые совпадают). Контур единичной длины вида (a,a) называется петлей.

Ориентированный путь из a в b задается последовательностью вершин $v_0v_1...v_n$. Длиной ориентированного пути называется количество ориентированных ребер, входящих в путь.

Понятие подграфа для ориентированных графов определяется также как и ранее.

Пример. Для графа G, изображенного на следующем рисунке слева, графы изображенные правее являются подграфами.



Ориентированные пути в G включают $v_0v_1v_2v_4$, $v_1v_2v_4$ и $v_0v_3v_4$.

□

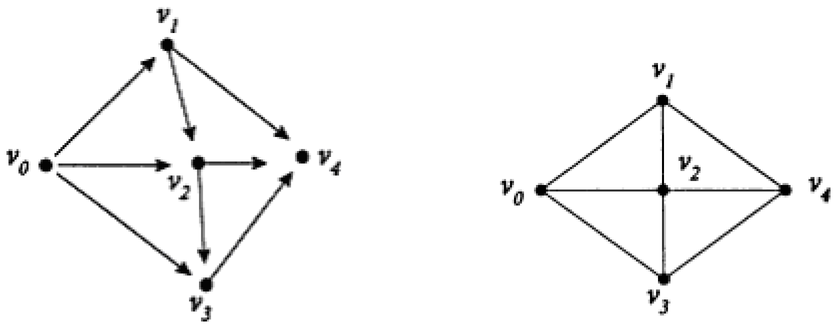
Теперь для заданного ориентированного графа G построим неориентированный граф G^s такой, что каждое ориентированное ребро G (исключая петли) станет неориентированным ребром графа G^s.

Пусть для каждого ориентированного графа $G(V, E)$ $E' = E - \{(v, v) : v \in V\}$, так что $G'(V,E')$ — ориентированный подграф графа G(V, E), в котором удалены петли. Пусть R — симметричное замыкание множества E', так что если $(a, b) \in E'$, то $(b, a) \in R$, а E^s — множество ребер, представляющих отношение R. В таком случае граф $G^s(V,E^s)$ называется соотнесенным графом ориентированного графа $G(V, E)$.

Еще раз. Неориентированный граф $G^s(V,E^s)$ называется соотнесенным графом ориентированного графа $G(V,E)$ если множество его ребер E^s можно определить таким образом: $\{a,b\} \in E^s$ тогда и только тогда, когда для различных вершин a и b ребро $(a,b) \in G$ или $(b,a) \in G$. Проще говоря, соотнесенный граф к ориентированному графу это та же диаграмма, но без петель и без стрелок.

Ориентированный граф $G(V, E)$ называется связным, если его соотнесенный граф является связным. Иными словами, граф связен, если любые две его вершины можно соединить цепью. Ориентированный граф называется сильно связным, если для любой пары вершин $a,b \in V$ существует ориентированный путь из a в b.

Пример. Для графа, приведенного на следующем рисунке слева, соотнесенным является граф, показанный справа.



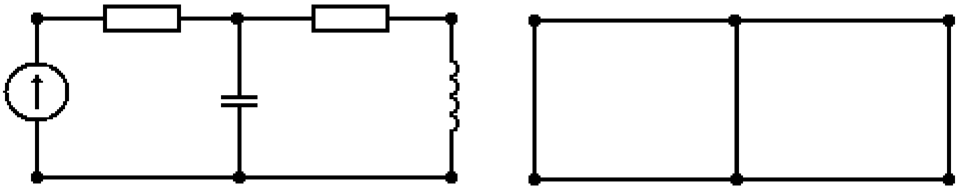
При этом, ориентированный граф связен, но не сильно связен, т.к. из вершины из v_1 в v_0 не существует ориентированного пути.

Решение многих технических задач методами теории графов сводится к определению тех или иных характеристик графов. Приведем здесь определения некоторых из них.

Цикломатическое число. Пусть G – неориентированный граф, имеющий n вершин и m ребер и r компонент связности. Цикломатическим числом графа G называется число

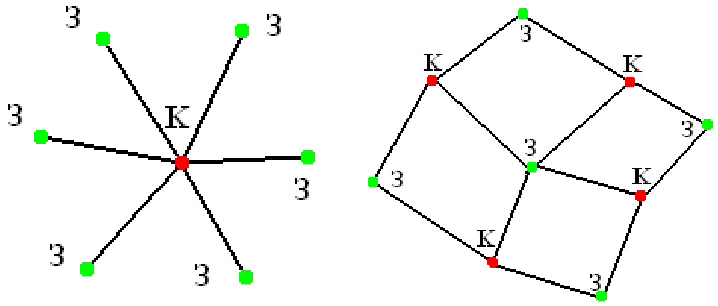
$$\nu(G) = m - n + r$$

Его смысл состоит в том, что оно равно наибольшему числу независимых циклов в графе. Например, при расчете электрических цепей цикломатическим числом можно пользоваться для определения числа независимых контуров. Например, на следующем рисунке слева изображена электрическая схема, граф которой показан справа.



У графа 6 вершин, 7 ребер и одна компонента связности. Следовательно, его цикломатическое число равно $\nu(G) = 7 - 6 + 1 = 2$. Это и есть число независимых контуров электрической схемы, приведенной слева. Второй закон Кирхгоффа гласит, что сумма падений напряжений на участках замкнутого контура равна нулю. Но для составления системы уравнений, описывающей электрическую цепь, следует использовать только независимые контуры, иначе уравнения будут зависимыми.

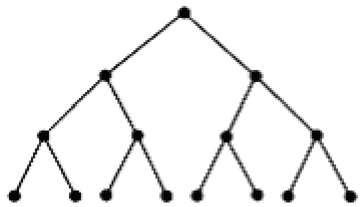
Хроматическое число. Пусть p – натуральное число. Граф G называется p – хроматическим, если его вершины можно раскрасить p различными цветами так, чтобы никакие две смежные вершины не были раскрашены одинаково. Наименьшее число p, при котором граф является p – хроматическим, называется хроматическим числом графа и обозначается $\gamma(G)$. Если $\gamma(G) = 2$, то граф называется бихроматическим. Необходимым и достаточным условием того, чтобы граф был бихроматическим, является отсутствие в нем циклов нечетной длины. На следующем рисунке приведены примеры бихроматических графов, в которых вершины имеют красный цвет (К) и зеленый (З) цвета.



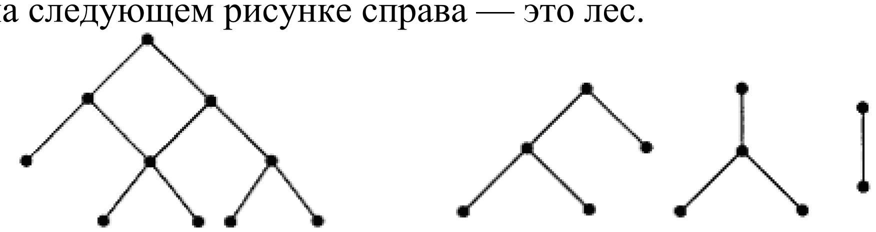
Хроматическое число играет важную роль при решении задачи наиболее экономичного использования ячеек памяти при программировании. Однако его определение, за исключением случая бихроматического графа, представляет собой довольно трудную задачу.

7.3 Деревья

Дерево — это связный граф без циклов. Лес — это граф, компоненты которого являются деревьями. Дерево и названо деревом, поскольку, будучи нарисованным, выглядит как дерево, только перевернутое «вверх ногами». Граф, изображенный на следующем рисунке является примером дерева.

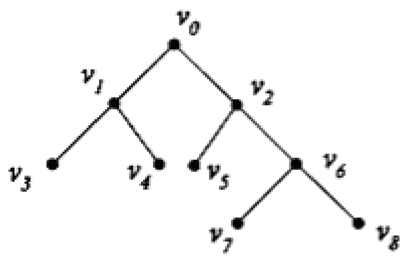


Граф на следующем рисунке слева не является деревом, поскольку содержит цикл. Граф на следующем рисунке справа — это лес.



Ориентированное дерево T представляет собой свободный от петель ориентированный граф, соотнесенный граф которого является деревом. Если существует путь из вершины *a* к вершине *b*, то он единственный, иначе можно было бы из двух таких путей образовать цикл. Аналогично, если в ориентированном дереве имеется ребро (*a, b*), тогда не существует ребро (*b, a*), в противном случае путь *aba* был бы циклом, и путь из *a* в *b* не был бы единственным.

В дереве (неориентированном или ориентированном) путь называется максимальным, если его нельзя продолжить. Максимальный путь не обязательно совпадает с самым длинным путем в дереве. Максимальных путей различной длины может быть много. Например, в дереве на следующем рисунке пути $v_0v_2v_5$, $v_0v_1v_3$, $v_0v_1v_4$, $v_0v_2v_6v_7$, $v_0v_2v_6v_8$ являются максимальными. Из них первые три имеют длину 2, а последние два – длину 3. Путь $v_3v_1v_0v_2v_6v_8$ имеет длину 5.



Если неориентированное дерево имеет хотя бы одно ребро, то оно имеет хотя бы две вершины со степенью 1. Чтобы убедиться в этом, рассмотрим все возможные пути. Если у дерева *n* вершин, тогда длина пути не может превысить *n* - 1, так как каждую вершину можно использовать только один раз. В противном случае получились бы циклы, а у дерева нет циклов. Поэтому существует максимальный путь, который нельзя продлить, чтобы получить более длинный путь. Предположим, что он начинается в вершине *a* и заканчивается в вершине b. Как a, так и b должны иметь степень 1, поскольку в противном случае путь можно было бы продолжить. Но это противоречит тому, что путь максимальный.

Вершины степени 1 называются листьями. На предыдущем рисунке вершины v_3, v_4, v_5, v_8, v_7 представляют собой листья. Другие вершины называются внутренними вершинами.

Теорема. Для любых двух вершин a и b дерева T существует единственный путь из a в b.

Доказательство. Предположим, что для некоторых вершин a и b дерева T путь из a в b не является единственным, и покажем, что в таком случае T не будет деревом. Допустим, что существуют два различных пути: $v_0v_1v_2...v_n$ длины n и $v'_0v'_1v'_2...v'_m$ длины m, где $a=v_0=v'_0$ и $b=v_n=v'_m$. В каждом пути должна существовать первая вершина, начиная с которой соответствующие вершины не совпадают, скажем $v_i \neq v'_i$, и в каждом из путей должна существовать точка, начиная с которой вершины опять одни и те же, скажем, $v_j=v'_k$ (это могут быть только первая и последняя вершины). Тогда $v_{i-1}v_iv_{i+1}v_{i+2}\ldots\genfrac{}{}{0pt}{}{v_j}{v'_k}v'_{k-1}v'_{k-2}\ldots v'_iv_{i-1}$ является циклом. В этой записи пути обозначения вершин, стоящие одна над другой, представляют одну и ту же вершину. Т.о. граф не является деревом.

□

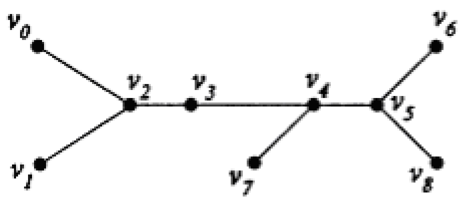
Последнее свойство иногда используют для определения понятия «дерево». Дерево можно определить как граф, в котором между любыми двумя его вершинами имеется единственный путь.

Теорема. Если для любых двух вершин графа G существует единственный путь из вершины a в вершину b, тогда G — дерево.

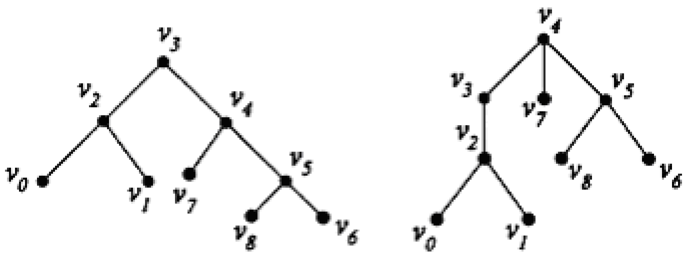
Доказательство. Предположим, что G не является деревом. Тогда либо G не является связным, либо содержит цикл. Если граф G не связный, то существуют вершины $a,b \in G$, для которых не существует пути из a в b. В таком случае, очевидно, не существует и единственного пути из a в b. Если G содержит цикл $v_0v_1v_2v_3v_4\cdots v_{k-1}v_k$, то как $v_2v_3\cdots v_{k-1}v_kv_0$ так и $v_2v_1v_0$ являются путями из v_2 в v_0. Положив $a = v_2$ и $b = v_0$, видим, что путь между вершинами a и b не является единственным.

□

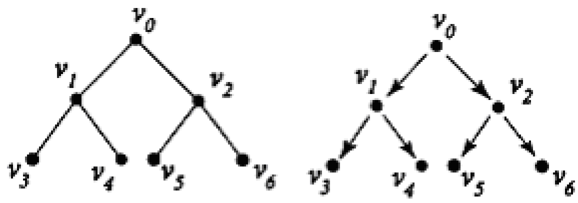
Предположим, что дерево представляет собой физический объект, подвижный в вершинах, и подвесим дерево за одну из его вершин так, что остальная его часть повиснет ниже этой вершины. Пусть задано следующее дерево



Если подвесить его за вершину v_3, получим дерево, представленное на следующем рисунке слева. Если подвесим дерево за вершину v_4, оно будет выглядеть так, как показано справа



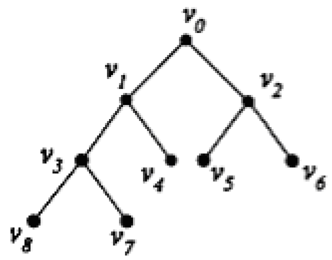
Вершина в самой верхней части каждого из изображений называется корнем дерева. Если корень дерева определен, дерево называется корневым деревом. При необходимости можно заменить корневое дерево T на ориентированное T". Например, дерево изображенное на следующем рисунке слева примет вид, приведенный справа



Такое дерево называется корневым ориентированным деревом, порожденным неориентированным корневым деревом. При этом следует помнить, что это ориентированное дерево отличается от неориентированного дерева и что вид ориентированного дерева зависит от выбора корня.

Если корень выбран, уровень вершины v определяется длиной единственного пути из корня в вершину. Высотой дерева называется длина самого длинного пути от корня дерева до листа. Для корневого ориентированного дерева вершина u называется родителем вершины v, а v называется сыном вершины u, если существует ориентированное ребро из u в v. Если u — родитель v и v', тогда v и v' называются братьями. Если существует ориентированный путь из вершины u в вершину v, тогда u называется предком вершины v, а v называется потомком вершины u. Если наибольшая из степеней выхода для вершин дерева равна m, тогда дерево называется m-арным деревом. В частном случае, когда $m = 2$, дерево называется бинарным деревом. В каждом бинарном дереве каждый сын родителя обозначается либо как левый сын, либо как правый сын.

Пример. Граф, приведенный на следующем рисунке, является бинарным деревом. Уровень вершины v_6 равен 2, уровень вершины v_8 равен 3. Высота дерева — 3, поскольку длина пути $v_0v_1v_3v_8$ равна 3 и не существует более длинного пути от корня к листу. Вершина v_1 является родителем для v_3 и v_4.



Вершины v_3 и v_4 — братья. Таковыми же являются вершины v_1 и v_2, v_5 и v_6 и v_7 и v_8. Вершина v_1— предок вершин v_3, v_7, v_8, а v_3, v_7 и v_8 — потомки вершины v_1. Вершина v_8 — левый сын вершины v_3, а v_4 — правый сын вершины v_1.

□

Теперь докажем, что в каждом дереве число вершин на единицу больше числа ребер. Предположим, что имеется дерево T. Ранее уже показано, что любое дерево можно представить как корневое дерево, и это никоим образом не меняет ни числа ребер, ни числа вершин. Рассмотрим теперь ориентированное дерево T', порожденное деревом T. У каждого ребра T одна и только одна конечная вершина. Следовательно, число ребер и вершин одно и то же, исключая корневую вершину. Если учесть корневую вершину, получим, что вершин на одну больше, чем ребер. Таким образом, нами доказана следующая теорема.

Теорема. Если у дерева T имеется *e* ребер и *v* вершин, тогда $v = e + 1$.

Справедлива также и обратная теорема.

Теорема. Если в связном графе G, содержащем *e* ребер и *v* вершин, имеем $v = e + 1$, тогда G — дерево.

Доказательство. Если G содержит цикл, то для ребра $\{v_i,v_j\}$ входящего в цикл, существуют два пути из v_i в v_j. Таким образом, ребро $\{v_i,v_j\}$ можно из цикла удалить, а путь из вершины v_i в вершину v_j будет существовать. Пусть a и b — любые точки в G. Поскольку граф G — связный, существует путь из a в b. Если путь проходил через удаленное ребро цикла $\{v_i,v_j\}$, то его (ребро) можно заменить альтернативным путем из v_i в v_j. Удалим ребро $\{v_i,v_j\}$ из G и, если оставшийся граф все еще содержит цикл, удалим другое ребро, используя ту же процедуру. Будем продолжать, пока все циклы не будут удалены. В результате получим связный граф, скажем, G', без циклов. Поэтому G' является деревом, и его число вершин $v = e' + 1$, где e' — число ребер графа G'. Поскольку ни одна из вершин не была удалена, число вершин остается таким, которое было раньше. Если было удалено n ребер, тогда $e = e' + n$. Но поскольку $v = e + 1$ и $v = e' + 1$, то $e = e'$ и $n = 0$. Следовательно, ни одно ребро не было удалено, поэтому G — дерево.

□

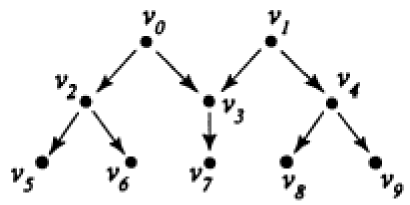
Дерево С, построенное из G в процессе приведенного выше доказательства, называется остовным (или каркасным) деревом графа G. Дадим более формальное определение.

Дерево T называется остовным деревом графа G, если T – подграф графа G и каждая вершина в G является вершиной в T. Итак, доказана следующая

Теорема. У каждого связного графа существует подграф, который является остовным деревом.

□

До сих пор рассматривались ориентированные деревья, образованные из корневых (неориентированных) деревьев. Ориентированное дерево называется корневым ориентированным деревом, если существует единственная вершина v_0 такая, что indeg(v_0) = 0, и существует путь из v_0 в каждую другую вершину дерева. Заметим, что ориентированные деревья, которые были рассмотрены до сих пор, в действительности соответствуют этому определению. Рассмотрим, однако, ориентированное дерево, представленное на следующем рисунке



Это ориентированное дерево, но оно не является корневым ориентированным деревом.

7.4 Матрицы инцидентности и смежности

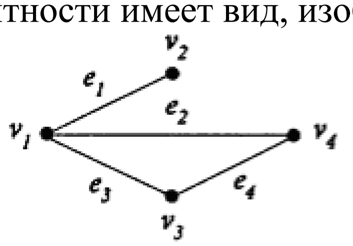
Здесь мы определим две матрицы, связанные с графами: матрицы инцидентности и матрицы смежности. Задание любой из этих матриц дает возможность восстановить граф.

Пусть G – граф и I – матрица, строки которой обозначены вершинами графа, а столбцы обозначены ребрами графа. Будем считать, что вершины и ребра графа пронумерованы. Элемент i-ой строки и j-го столбца матрицы I, обозначаемый s_{ij}, равен 1, если i-я вершина инцидентна j-у ребру, и равен 0 в противном случае. Матрица I называется матрицей инцидентности графа G.

Таким образом, квадратная матрица $I = [s_{ij}]$ порядка $n \times m$ называется матрицей инцидентности графа, если

$$s_{ij} = \begin{cases} 1, \text{ если } e_j \text{ инцидентна } v_i \\ 0, \text{ если } e_j \text{ неинцидентна } v_i \end{cases}$$

Пример. Пусть G – граф, изображенный на следующем рисунке слева. Тогда его матрица инцидентности имеет вид, изображенный на рисунке справа

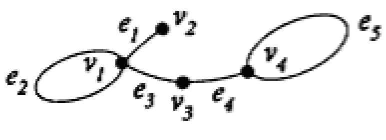


$$\begin{array}{c} \\ v_1 \\ v_2 \\ v_3 \\ v_4 \end{array} \begin{array}{c} \begin{array}{cccc} e_1 & e_2 & e_3 & e_4 \end{array} \\ \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \end{array}$$

Легко видеть, что степень вершины равна сумме элементов строки, обозначенной этой вершиной, так как каждая единица в этой строке представляет инцидентность этой вершины ребру. При этом в каждом столбце будут ровно две единицы, так как каждое ребро инцидентно двум вершинам.

Можно также включить в рассмотрение матрицы инцидентности для графов с петлями. Вид матрицы инцидентности непосредственно показывает, является ли данное ребро петлей, так как ребро представляет собой петлю тогда и только тогда, когда соответствующий столбец содержит только одну единицу. В матрице инцидентности для графа с петлями сумма элементов строки, соответствующей данной вершине, не представляет собой степень вершины, если в ней имеются петли.

Пример. Пусть G – граф, изображенный на следующем рисунке слева. Его матрица инцидентности изображена на рисунке справа.



$$\begin{array}{c} \\ v_1 \\ v_2 \\ v_3 \\ v_4 \end{array} \begin{array}{c} \begin{array}{ccccc} e_1 & e_2 & e_3 & e_4 & e_5 \end{array} \\ \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \end{array}$$

Обратите внимание, что наличие петель e_2 и e_5 приводит к тому, что в столбцах, обозначенных этими ребрами, содержится только по одной единице.

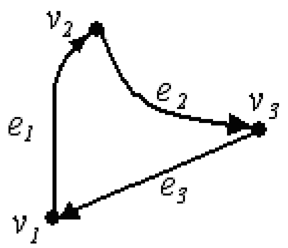
Матрицы инцидентности не имеют большого значения при рассмотрении ориентированных графов, т.к. они не содержат информации о том, как ребро ориентировано. Поэтому, используя матрицу инцидентности, нельзя восстановить ориентированный граф.

Если граф не содержит петель, то матрицу инциденций можно построить так, чтобы она содержала информацию об ориентации ребер. В самом деле обозначим через $v_1, v_2, \ldots, v_n$ вершины графа, а через $e_1, e_2, \ldots, e_m$ – его дуги. Введем числа

$$s_{ij} = \begin{cases} +1, \textit{если } e_j \textit{ исходит из } v_i \\ -1, \textit{если } e_j \textit{ заходит в } v_i \\ \;0\;, \textit{ если } e_j \textit{ неинцидентна } v_i \end{cases}$$

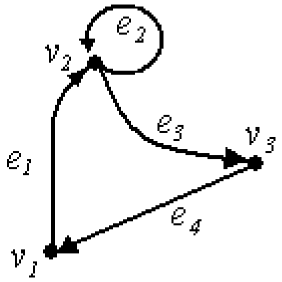
Матрица $S = [s_{ij}]$ порядка $n \times m$ называется матрицей инциденций ориентированного графа. В таком виде она определима только для графов без петель.

Пример. Для ориентированного графа, изображенного на следующем рисунке слева, матрица инциденций показана справа



$$\begin{array}{c} v_1 \\ v_2 \\ v_3 \end{array} \begin{bmatrix} +1 & 0 & -1 \\ -1 & +1 & 0 \\ 0 & -1 & +1 \end{bmatrix}$$

При наличии петель, матрицу инциденций следует расчленить на две матрицы: положительную и отрицательную. Тогда для вершины с петлей в одной матрице будет стоять +1, а в другой -1 на одинаковом месте. Например, для следующего графа, положительной и отрицательной матрицами, связывающие выходящие и входящие ребра с вершинами, будут



$$\begin{array}{c} v_1 \\ v_2 \\ v_3 \end{array}\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 0 & 0 & 0 & -1 \\ -1 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{bmatrix}$$

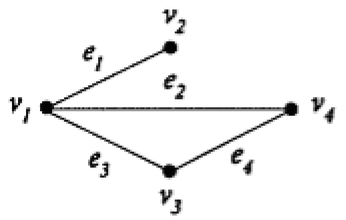
Здесь элемент 2-2 отличен от нуля в обеих подматрицах.

Для работы с ориентированными графами удобно использовать также другой тип матриц – матрицы смежности, определение которых будет сейчас дано.

Пусть G — граф (ориентированный граф) и пусть S — матрица, строки которой обозначены вершинами графа и столбцы обозначены теми же вершинами в том же самом порядке. Элемент i-ой строки и j-го столбца матрицы S равен 1, если имеется ребро (или ориентированное ребро) из i-ой вершины в j-ю вершину, и равен 0 в противном случае. Таким образом, квадратная матрица $S=[s_{ij}]$ порядка $n\times n$ называется матрицей смежности графа, если

$$s_{ij}=\begin{cases} 1, \textit{есть дуга, соединяющая вершину i с вершиной j} \\ 0, \textit{если такой дуги нет} \end{cases}$$

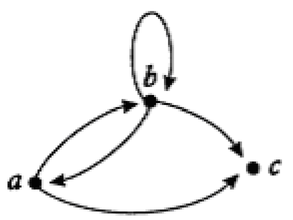
Пример. Пусть G –граф (неориентированный), изображенный на следующем рисунке слева. Его матрица смежности приведена на рисунке справа.



$$\begin{array}{c|cccc} & v_1 & v_2 & v_3 & v_4 \\ \hline v_1 & 0 & 1 & 1 & 1 \\ v_2 & 1 & 0 & 0 & 0 \\ v_3 & 1 & 0 & 0 & 1 \\ v_4 & 1 & 0 & 1 & 0 \end{array}$$

Поскольку петли отсутствуют, все элементы главной диагонали матрицы равны 0. Матрица смежности (неориентированного графа) симметрична, так как граф представляет симметричное отношение.

Пример. Пусть G – ориентированный граф, изображенный на следующем рисунке слева. Его матрица смежности приведена на рисунке справа.



$$\begin{array}{c|ccc} & a & b & c \\ \hline a & 0 & 1 & 1 \\ b & 1 & 1 & 1 \\ c & 0 & 0 & 0 \end{array}$$

Как правило, обозначения вершин несущественны. В таких случаях матрицы приводятся без обозначений строк и столбцов. Например, матрица

$$\begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

является матрицей смежности для ориентированного графа, у которого четыре вершины и восемь ребер.

7.5 Задача Эйлера

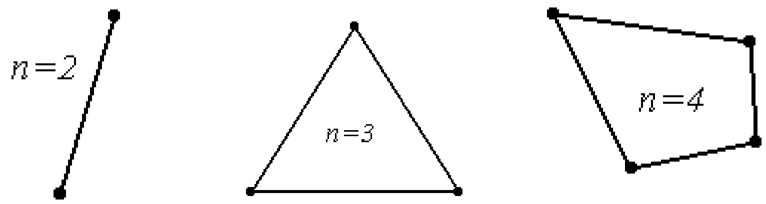
Пусть G = (V, E) — граф. Цикл, который включает все ребра и вершины графа G, называется эйлеровым циклом. Если это условие выполняется, говорят, что граф G имеет эйлеров цикл.

Если теперь вернуться к задаче о кенигсбергских мостах, то ясно, что она сводится к попытке определить, содержит ли граф, изображающий задачу, эйлеров цикл. Для этой цели нам потребуется одна теорема. Она справедлива также для мультиграфов и псевдографов, исключая тот случай, когда псевдограф имеет только одну вершину. Для ясности будем использовать термин граф, понимая, что наши утверждения справедливы для мультиграфов и псевдографов.

Теорема. Граф (и мультиграф) с более чем одной вершиной имеет эйлеров цикл тогда и только тогда, когда он связный и каждая его вершина имеет четную степень.

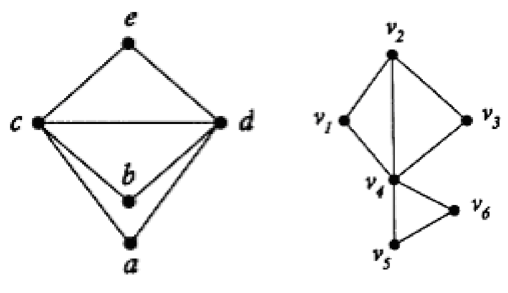
Доказательство теоремы в одну сторону почти очевидно. Предположим, что граф G имеет эйлеров цикл. Тогда граф является связным, т.к. каждая вершина принадлежит циклу. Для всякой вершины v графа G каждый раз, когда эйлеров цикл проходит через v, он вносит 2 в степень v. Поэтому степень v четная. Таким образом, если эйлеров цикл существует, то каждая его вершина имеет четную степень.

Обратно, нужно показать, что каждый связный граф, у которого степени вершин четные, имеет эйлеров цикл. Доказательство этого факта можно провести, используя индукцию по числу вершин. Например, легко видеть, что теорема справедлива при n = 3 и n = 4 (см. рисунок).



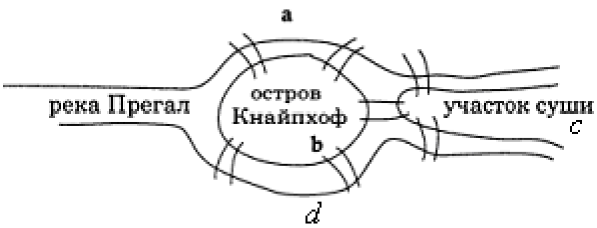
Строгое доказательство мы здесь не приводим; читатель может найти его в литературе по теории графов.

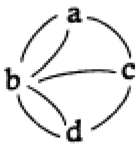
Пример. Граф, показанный на следующем рисунке слева, имеет эйлеров цикл, поскольку степень каждой его вершины четная.



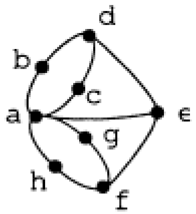
Пример. Граф, показанный на предыдущем рисунке справа, не имеет эйлерова цикла, поскольку степени вершин v_2 и v_4 — нечетные.

Вернемся к задаче о кенигсбергских мостах. Мультиграф, построенный Эйлером, и показанный на следующем рисунке справа, имеет нечетные степени всех его вершин.





Это значит, что он (мультиграф) не имеет эйлерова цикла, поэтому невозможно пройти каждый мост по одному разу и вернуться в исходную точку пути. Хотя задача была решена с использованием мультиграфа, ее можно решить, используя простой граф. Когда имеется более одного ребра между вершинами, можно к исходному мультиграфу добавить вершины для представления середин каждого моста. Тогда получится простой граф, показанный на следующем рисунке, который также описывает задачу о кенигсбергских мостах.



Относительно этих мостов, можно также задать вопрос: возможно ли пройти каждый мост по одному разу, но не возвращаться обязательно в исходную точку маршрута? Это предположение приводит к следующему определению и теореме.

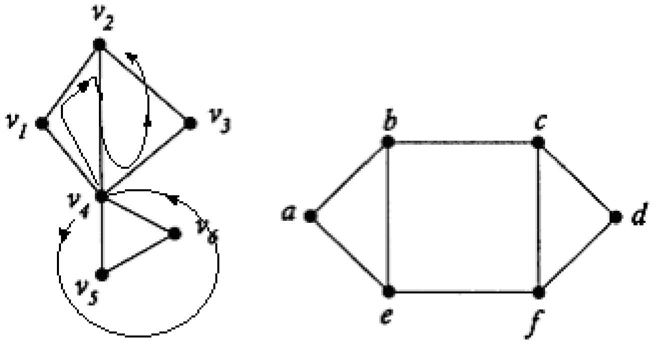
Пусть G = (V,E) — граф. Путь, который включает каждое ребро графа G только один раз называется эйлеровым путем. В этом случае говорят, что граф G имеет эйлеров путь.

Сейчас нас будет интересовать случай, когда эйлеров путь не является эйлеровым циклом. Такой путь будем называть собственным эйлеровым путем. Следующую теорему приведем без доказательства.

Теорема. Граф (мультиграф или псевдограф) имеет собственный эйлеров путь тогда и только тогда, когда он связный и ровно две его вершины имеют нечетную степень.

Поскольку граф для кенигсбергских мостов имеет четыре вершины с нечетными степенями, можно сделать вывод о невозможности пройти каждый мост по одному разу, даже если не нужно возвращаться в исходную точку маршрута.

Пример. Граф, показанный на следующем рисунке слева, имеет собственный эйлеров путь, т.к. ровно две его вершины имеют нечетную степень. Путь начинается в вершине v_4 и заканчивается в вершине v_2.



Пример. Граф, показанный на предыдущем рисунке справа, не имеет собственного эйлерова пути, т.к. четыре его вершины имеют нечетную степень.

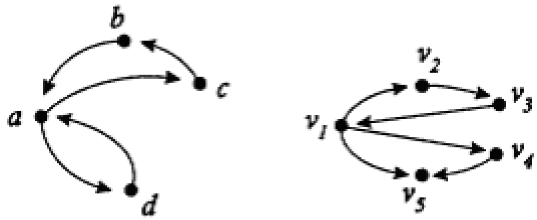
Пусть G = (V,E) — ориентированный граф. Ориентированным циклом называется ориентированный путь ненулевой длины из вершины в ту же вершину без повторения ребер.

Пусть G = (V,E) — ориентированный граф. Ориентированный цикл, который включает все ребра и вершины графа G, называется эйлеровым циклом. В этом случае говорят, что ориентированный граф G имеет эйлеров цикл.

Следующую теорему приведем без доказательства.

Теорема. Ориентированный граф имеет эйлеров цикл тогда и только тогда, когда он связный и степень входа каждой вершины равна ее степени выхода.

Пример. Граф, показанный на следующем рисунке слева, имеет эйлеров цикл, так как степень входа каждой вершины равна степени выхода.



Пример. Граф, показанный на предыдущем рисунке справа, не имеет эйлерова цикла, так как степень входа вершины v_1 не равна ее степени выхода.

7.6 Задача о кратчайшем пути

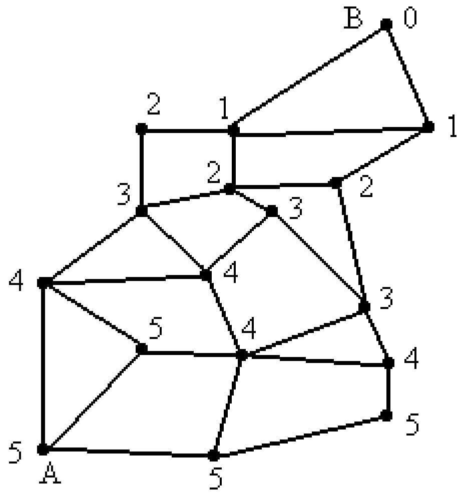
Пусть дан неориентированный граф $G = (X, U)$. Каждому его ребру припишем некоторое число $l(u) \geq 0$, называемое длиной ребра. В частности это может быть расстояние между вершинами, временем или стоимостью проезда по этому ребру и т.п. Любая цепь μ будет характеризоваться длиной $\sum_{u \in \mu} l(u)$. Для двух произвольных вершин *a* и *b* графа G требуется найти путь μ_{ab} такой, что его полная длина будет наименьшей.

Здесь мы приведем метод решения этой задачи частного вида, когда длина каждого ребра равна единице. Алгоритм решения этой задачи состоит в приписывании каждой вершине x_i индекса λ_i, равного длине кратчайшего пути из данной вершины в конечную. Приписывание индексов для графа с ребрами единичной длины производится в следующем порядке:

- конечной вершине x_0 приписывается индекс 0;
- всем вершинам, из которых идет ребро в конечную вершину, приписывается индекс 1;
- вершинам еще не имеющих индексов, но для которых есть ребро ведущее в вершину уже имеющую индекс λ_i, приписываем индекс $\lambda_i + 1$ и продолжаем процесс приписывания индексов, до тех пор, пока не будет помечена начальная вершина.

По окончании разметки индекс у начальной вершины будет равен длине кратчайшего пути. Сам кратчайший путь найдем, если будем двигаться из начальной вершины в направлении убывания индексов.

На следующем рисунке показан пример графа с проставленными индексами вершин



В результате кратчайший путь из A в B состоит из 5 ребер.

8. Элементы теории кодирования

8.1 Введение

Десятичная позиционная система счисления – это способ кодирования натуральных чисел, двоичная система счисления – это другой способ кодирования этих чисел. Римские цифры – еще один способ кодирования. Декартовы координаты – способ кодирования геометрических объектов. Замена имен номерами является кодированием, запись текстового файла на съемный носитель – кодирование, представление фотографии в цифровом фотоаппарате в виде файла – кодирование.

Упрощая, можно сказать, что кодирование – это способ представления объектов одной природы (точек плоскости, чисел, слов какого либо языка и т.д.) конечными последовательностями объектов другого множества (обычно конечного). Поскольку элементы этого второго конечного множества можно перенумеровать, то его всегда можно считать набором чисел, однако, это необязательно.

Наборы символов, которыми пользуются при кодировании, называются алфавитами. Последовательности символов алфавита называются словами. Письменность – это алфавитное кодирование. Например, можем закодировать слово «дай» последовательностью нулей и единиц. Для этого можем заменить буквы их порядковыми номерами в алфавите, и затем номера записать в двоичном виде.

Теория кодирования представляет собой раздел дискретной математики, в котором исследуются отображения конечных или счетных множеств объектов произвольной природы в множества последовательностей из чисел 0, 1, …, q-1, где q – некоторое натуральное число, часто q=2. Такие отображения называют кодированием. Многие задачи теории кодирования сводятся к отысканию кодирования, обладающего определенными свойствами и оптимального в некотором смысле. Критерий оптимальности может быть разным. Это может быть, например, минимизация длины образов объектов. Свойства кодирования также могут быть разными. В частности такими свойствами могут быть существование однозначного обратного отображения (декодирования), возможность исправления ошибок при декодировании и т.д.

Ранее кодирование не рассматривалось как отдельный предмет математического изучения. Но с появлением компьютеров ситуация радикально изменилась, поскольку кодирование буквально пронизывает информационные технологии и является центральным вопросом при решении многих задач программирования:

- представление данных (чисел, текста, графики) в памяти компьютера;
- защита информации;
- сжатие информации.

Методы решения каждой из указанных выше задач разрослись в целые науки, некоторым из которых даны даже свои названия. Так, когда речь идет о кодах, часто имеются в виду секретные коды. Проблема передачи информации, которая должна быть понятна только тому, для кого она предназначена, весьма важна для военных и коммерческих целей, для обеспечения конфиденциальности переписки. Параллельно с разработкой кодов идет работа по их взламыванию. Раздел теории кодов, посвященный этим вопросам, называется криптологией.

В кибернетике особое место занимает проблема надежности передачи информации. Это связано с тем, что разнообразные дискретные устройства имеют многочисленные каналы передачи информации. В этих каналах происходит искажение сигналов. Кодирование позволяет надежно передать требуемое сообщение, несмотря на наличие помех в канале.

Итак, формализуем постановку задачи теории кодирования. Пусть задано конечное множество $A=\{a_1,\dots,a_n\}$, которое называется алфавитом. Элементы алфавита называются буквами. Последовательность букв называется словом. Множество слов в алфавите A обозначается A^*. Количество букв в слове называется длиной слова $|\alpha|=|a_1\dots a_k|=k$. Задача кодирования формулируется следующим образом. Пусть заданы алфавиты $A=\{a_1,\dots,a_n\}$, $B=\{b_1,\dots,b_m\}$ и функция $F: S \to B^*$, где S – некоторое множество слов в алфавите A, $S \subset A^*$. Функция F называется кодированием, элементы множество S – сообщениями, а элементы $\beta = F(\alpha), \alpha \in S, \beta \in B^*$ - кодами сообщений. Обратная функция F^{-1}, если она существует, называется декодированием. Наиболее часто рассматривается случай, когда $B = \{0,1\}$, и F называется двоичным кодированием.

Типичная задача теории кодирования формулируется следующим образом: при заданных алфавитах А и В, и множестве сообщений S, найти такое кодирование F, которое удовлетворяет заданным ограничениям и в некотором смысле оптимально. Оптимальность часто понимается как минимизация длины кодов. Свойства, которые требуются от функции кодирования F, бывают разными:

- существование декодирования, т.е. обратной функции F^{-1};
- помехоустойчивость: декодирование F^{-1} обладает свойством $F^{-1}(\beta) = F^{-1}(\beta')$, если β' близко к β в некотором смысле;
- простота алгоритма кодирования или декодирования.

Наиболее важное свойство кода состоит в том, что когда сообщение кодируется как двоичная строка, состоящая из конкатенации элементов кода, эта конкатенация однозначна. При декодировании сообщения не должно возникать проблем с тем, какую букву представляет элемент кода. Такой код назовем однозначно декодируемым кодом. Существует несколько способов достижения этой цели. Самым простым является *блоковое кодирование*, когда все символы кодируются двоичными строками одной длины. Соответствующий код называется блоковым. Например, если для кодирования каждого символа используется 8 бит, то известно, что каждые восемь бит представляют один символ передаваемого сообщения.

Другим способом построения однозначно декодируемого кода является использование *префиксного кода*, который будет рассмотрен ниже. Согласно определению, код является префиксным, если обладает тем свойством, что элемент кода не может быть начальной строкой другого элемента кода. Таким образом, при чтении строки из единиц и нулей, изображающей символ, всегда известен момент завершения строки. И хотя коды символов имеют разную длину, конец кода символа всегда известен.

Часто необходимо сжимать данные, чтобы минимизировать объем памяти для их хранения или время для передачи данных. Что касается минимизации объема памяти, то наиболее эффективным является код Хаффмана. Он тоже является префиксным кодом. Кодирование при помощи кода Хаффмана будет описано позже.

Разновидностью префиксного кода является кома-код. При его использовании каждый символ кодируется строкой из единиц, в конце которой стоит нуль. Значит, множество строк кода имеет вид $\{0,10,110,11110,111110,...\}$. Этот код имеет явный недостаток: элементы кода могут быть очень длинными и занимать большой объем памяти.

В процессе передачи данных могут возникать ошибки. Все, что может стать причиной ошибок, называется неопределенным термином "шум". Например, данные, полученные от отдаленного космического корабля наверняка подвержены различного рода шумам. В некоторых случаях интерес представляет только определение наличия ошибки. Коды, обладающие свойством определения наличия ошибок, называются *кодами, обнаруживающими ошибки*.

В другом случае, когда данные не могут быть переданы еще раз (например, данные от удаленного космического корабля), требуется дополнительная информация о данных с целью не только обнаружения, но и исправления ошибки. Коды, позволяющие исправлять ошибки, называются *кодами, исправляющими ошибки*. Может показаться разумным всегда использовать коды с исправлением ошибок. Проблема использования кодов с исправлением ошибок и кодов с обнаружением ошибок состоит в том, что они должны включать в себя дополнительную информацию, поэтому они являются менее эффективными в отношении минимизации объема памяти. К тому же, использование кодов с исправлением ошибок и кодов с обнаружением ошибок не дает абсолютной гарантии того, что ошибка будет исправлена или обнаружена. Проблема заключается в возможности многих ошибок. Несомненно, ошибку можно исправить или обнаружить, если она одна. Все, что мы можем сделать, — это уменьшить вероятность того, что ошибки останутся необнаруженными и неисправленными. Но проблема, опять-таки, в том, что чем в большей степени будет уменьшена эта вероятность, тем больше информации потребуется переслать, и тем менее эффективным будет наш код.

В качестве первого метода обнаружения ошибок рассмотрим бит контроля четности. Продемонстрируем этот метод на примере кода ASCII. ASCII-код является блоковым кодом, который использует 7 битов (основная кодировка), поэтому любой закодированный символ изображается строкой из семи символов 1 и 0. Восьмой бит добавляется таким образом, чтобы количество единиц было четным. Поэтому, если код переданной строки получен с единственной ошибкой, то количество единиц будет нечетным, и получатель информации поймет, что произошла ошибка. К сожалению, если произошло две ошибки, их нельзя будет обнаружить, поскольку количество единиц опять будет четным. Предположим, что вероятность ошибки при передаче равна 0.01 как для изменения 1 на 0, так и для изменения 0 на 1. Будем также считать, что вероятность ошибки не зависит от расположения ошибки и от того, что является ошибкой: изменение 1 на 0 или наоборот. Предположим также, что появление одной ошибки не влияет на вероятность появления другой. Из формул теории вероятности можно вычислить, что вероятность появления точно одной ошибки равна $C_8^1 \cdot 0.01 \cdot (0.99)^7$ что приблизительно равно 0.07. Вероятность же появления в точности двух ошибок равна $C_8^2 \cdot (0.01)^2 \cdot (0.99)^6$, что приблизительно равно 0.002 и является существенно меньшей величиной.

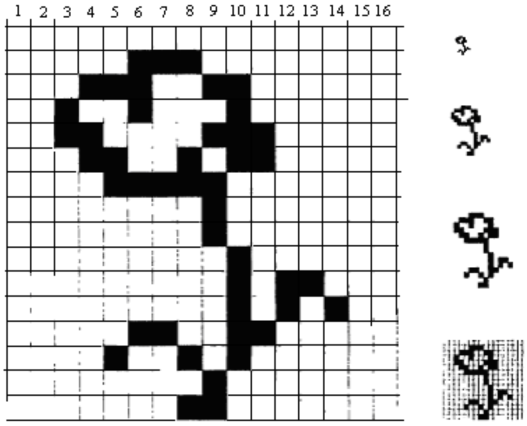
Поскольку 3 ошибки будут обнаружены, то вероятность не обнаружить более двух ошибок меньше, чем вероятность наличия четырех или более ошибок, т.к. любое нечетное количество ошибок будет обнаружено. Эта вероятность пренебрежимо мала по сравнению с вероятностью одной ошибки.

Рассмотрим другой способ кодирования с контролем ошибок, в котором для кодирования используется повторение кодируемой строки заданное число раз. Например, если при кодировании каждую строку кода нужно повторить один раз, то 10110 будет закодировано как 10110 10110. Если при кодировании каждую строку кода нужно повторить дважды, то 10110 будет закодировано как 10110 10110 10110. Если при кодировании каждая строка повторена один раз, то в результате получаем код с обнаружением ошибок. Так, если произошла ошибка, то соответствующие позиции (единицы или нули в этих позициях) не будут совпадать. Исправить ошибки мы не можем, поскольку не знаем, в какой из копий (первой или второй) присутствует ошибка. Если кодируемая строка повторяется дважды, то можно выявить ошибку. Если имеются три копии строки, то можно исправить код при наличии единственной ошибки. Например, если получим код 10110 10110 11110, то видим, что первые два кода совпадают, а ошибка находится в третьей группе символов. Если имеется отличие в битах в соответствующих позициях строк, то выбирается значение, которое встречается дважды. Или, например, если строка имеет длину 4 и нам передано 1101 1001 1101, то во второй позиции каждой четверки символов мы два раза получим 1 и один раз 0. Таким образом, предполагаем, что правильное значение равно 1, и правильным вариантом закодированной строки является 1101. Конечно, если ошибка появляется в одной и той же позиции строки более одного раза, возникает проблема. Если строка повторена так, что имеется n ее копий, то исправление ошибки дает правильный результат, если при повторении ошибка встречается в одной и той же позиции менее чем $[n/2]$ раз.

8.2 Примеры кодирования

Кодирование изображений. Еще одна очень важная трактовка набора из нулей и единиц — это кодирование геометрического изображения. Двухцветная картинка (будем по традиции говорить о черно-белой картинке – черный рисунок на белом фоне) может трактоваться как растр – совокупность отдельных точек, расставленных на прямоугольной решетке. Сопоставляя черным точкам единицы, а белым – нули, мы и закодируем картинку в виде таблицы из нулей и единиц. Затем таблицу можно "вытянуть" в одну линию, соединяя ее строки или столбцы.

Пример. Рассмотрим решетку 16 х 16 и картинку на ней



Кодируя столбцы двоичными цифрами, имеем

Столбец	Двоичный код	Шестнадцатеричный код
1	0000 0000 0000 0000	0000
2	0000 0000 0000 0000	0000
3	0001 1000 0000 0000	1800
4	0010 1100 0000 0000	2C00
5	0010 0110 0000 0100	2604
6	0111 0010 0000 1000	7208
7	0100 0010 0000 1000	4208
8	0100 0110 0000 0101	4605
9	0010 1011 1000 0011	2B83
10	0011 1100 0111 1100	3C7C
11	0000 1100 0000 1000	0C08
12	0000 0000 0011 0000	0030
13	0000 0000 0010 0000	0020
14	0000 0000 0001 0000	0010
15	0000 0000 0000 0000	0000
16	0000 0000 0000 0000	0000

Таким образом, мы получаем вектор, который можно представить в шестнадцатеричной системе:

0000, 0000, 1800, 2C0O, 2604, 7208, 4208, 4605,
2B83, 3C7C, 0C08, 0030, 0020, 0010, 0000, 0000.

Упражнение 1. Нарисуйте сами и закодируйте какой-либо символ 8x8 (букву латинского или русского алфавита).

Упражнение 2. Молодая художница подарила нам картинку

0000, 0018, 0124, 01E2, 03C4, 04B8, 0250, 03U,
027C, 5650, 6F4C, 0F92, 00A2, 0154, 01C8, 0000.

Что на ней изображено?

Возможность точечного рисования картинок широко используется в вычислительной технике и в обработке изображений. Например, экран дисплея рассматривается как растр. Растром (от латинского `rastrum` — грабли) в полиграфии называют прямоугольную сетку, через которую фотографировали тоновое изображение, разбивая его на точки. Символы, изображаемые на дисплее или на печатающем устройстве, также имеют растровое представление. При этом растр, в зависимости от конкретного устройства, трактуется как последовательность строк или как последовательность столбцов.

`Кодирование чисел.` Целые положительные числа в памяти компьютера можно представлять в двоичной форме. При этом число двоичных разрядов для записи чисел фиксировано и обычно равно 8, 16, 32 или 64. Это количество двоичных разрядов называется разрядностью или длиной разрядной сетки компьютера. Например, в первых персональных компьютерах IBM целые числа занимали 16 двоичных разрядов. При таком кодировании старшие разряды небольших положительных чисел заполняются нулями. Например, число 45_{10}= 1011012 в памяти такого компьютера, занимая 2 байта, имело вид 00000000 00101101.

Представление целых чисел любого знака требует новых решений. Из многих мыслимых вариантов устойчиво закрепились два: кодирование со смещением и дополнительный код.

При *кодировании со смещением* к каждому числу `r` прибавляется константа, выбранная так, чтобы при любом `r` сумма `r+D` была неотрицательной; получающаяся сумма и кодируется вместо `r`. Саму константу нетрудно выбрать. Действительно, если для представления чисел выбрано `k` двоичных разрядов, то можно закодировать $2^к$ различных чисел. Естественно поделить эти возможности поровну между положительными и отрицательными числами и приписать ноль к положительным числам. Тогда константа смещения равна D = $2^{к-1}$. Например, если `k = 5`, то D = 2^4 = 16 и к кодируемому числу надо просто прибавить 16. Например, code(0) = 16, code(11) = 27, code(-7) = 9, а code(17) и code(-19) не определены, так как эти числа выходят за границы интервала определения кода -16: 15.

В современных компьютерах для представления отрицательных чисел используется метод двоичного дополнения. Он состоит в следующем. Если дано положительное целое число, представленное с использованием `n` бит, то соответствующее ему отрицательное число находят путем вычитания этого числа из 2^n, выраженного в двоичной форме.

Пример. Найдем представление числа -56 в 8-битовом двоичном дополнительном коде. Сначала число 56 представляется как 00111000. Затем выполняем вычитание 2^8-56, только в двоичном представлении

$$-\begin{array}{r}100000000\\ 00111000\\ \hline 11001000\end{array}$$

Фактически этот алгоритм выполняется следующим образом. Берем модуль отрицательного целого числа (т.е. положительное число), записываем его в двоичном виде, например 00111000. Затем каждый ноль заменяем единицей, а единицу – нулем и к полученному двоичному числу, в нашем примере 11000111, прибавляем единицу.

$$
\begin{array}{r}
+\begin{array}{r}11000111\\ 1\end{array}\\
\hline
11001000
\end{array}
$$

Целое положительное число всегда будет начинаться с 0, а целое отрицательное число будет начинаться с 1. Поэтому при восьмиразрядном представлении число 11001000 будет представлять целое отрицательное число. Поскольку при таком способе записи наибольшее возможное целое положительное число есть 01111111, то всего имеется 127 целых положительных чисел. Отрицательные двоичные дополнения изменяются от 11111111 (= – 1_{10}) до 10000000 (= – 128_{10}), так что их имеется 128. Таким образом, если использовать 8 двоичных разрядов, можно представить все числа от –128 до 127. Если имеется n двоичных разрядов, то можно представить целые числа от -2^{n-1} до $2^{n-1}-1$.

Кодирование со смещением применяется при задании чисел с плавающей точкой. В нем число x представляется в виде $x = s \cdot \alpha \cdot 2^{\beta}$, где s = ±1, α – мантисса, неотрицательное число из диапазона, разного для разных случаев, которое с некоторой условностью можно считать целым, β— порядок, целое число. Двоичный вектор, отведенный для записи числа, разбивается на три поля: для знака числа, для мантиссы и для порядка. Знак числа имеет два возможных значения и ему достаточно одного разряда. Мантисса записывается в дополнительном коде, а вот для записи порядка принято использовать код со смещением.

Штриховые коды. Во многих практических информационных системах – в магазинах, в почтовой службе и в других местах, где требуется быстро и просто считывать в компьютер небольшую числовую или текстовую информацию, используют так называемые штриховые коды (barcodes). Мы встречаем их на упаковках товаров. Имеется много разновидностей этих кодов; пример одной из них, так называемого кода 3 из 9, представлен на следующем рисунке.

Каждый символ кодируется 9 полосками, поочередно черными и белыми, в конце добавляется еще одна белая полоска. Три полоски из основных девяти имеют увеличенную ширину. Сопоставляя каждой широкой полоске 1, а узкой полоске 0, получаем 10-битовые последовательности, в которых последний бит всегда равен 0, так что существенными для информации оказываются 9 битов. Три широких полоски можно разместить на любом из 9 мест, т.е. имеется возможность закодировать $C_9^6 = 84$ символов. В реальности кодируются 44 различных алфавитных, цифровых и специальных символов. Кодируемыми символами являются цифры (0-9), буквы верхнего регистра (A-Z) и восемь других символов (пробел ("_"), "*", "$", "%", "+", "-", "." и "/"). Символ звездочки (*) применяется исключительно в качестве кода начала или конца; звездочки внутри набора данных не допускаются. Другое название этой символики – код 39.

Кодирование текста. При кодировании текстов обычно принимается какой-нибудь стандартный размер кодовых последовательностей. Например, в кодировке ASCII каждый символ занимает один байт — 8 битов, а в Unicode два байта. В первых моделях телеграфа появилась и долго использовалась пятибитовая кодировка. Каждая из таких кодировок имеет ограниченное число кодовых возможностей, хотя в Unicode это ограничение слабо чувствуется.

В языках, использующих латиницу, завоевал всемирное признание код ASCII — American Standard Code for Information Interchange, первоначально разработанный и стандартизованный в США. В системе кодирования ASCII первые 32 кода (шестнадцатеричные числа от 00 до 1F) являются управляющими, с помощью которых системные или прикладные программы реализуют свои рабочие функции. Остальные 96 (от 20 до 7F) используются для кодирования знаков пунктуации, знаков арифметических операций, некоторых специальных символов, цифр, и конечно букв латинского алфавита. Для кодирования символов ASCII используется 7 битов, что позволяет закодировать 128 символов (коды от 00 до 7F). Этого не всегда достаточно, поэтому задействуют еще один – восьмой бит, что позволяет расширить таблицу в 2 раза. Символы, помещаемые в расширение основной таблицы ASCII, имеют шестнадцатеричные коды от 80 до FF. Именно в расширение таблицы ASCII и помещают набор букв кириллицы. К сожалению, единой общепринятой кодировки для букв кириллицы нет. Наиболее популярна альтернативная кодировка, она же кодировка cp866. Зарубежные программные средства, работающие, как правило, с латинским алфавитом, используют в основном расширение таблицы ASCII, соответствующее стандартному знакогенератору IBM PC, в котором кириллицы нет. При этом из расширения таблицы ASCII привлекаются главным образом символы псевдографики.

Система кодов ASCII обычно представляется в виде таблицы, называемой основной таблицей ASCII. Ее вид приведен ниже. Мы в нее не вписали две первые строки, в которых содержатся управляющие коды, и последнюю строку, в которой находятся редко используемые символы.

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
2		!	"	#	$	%	&	'	(	)	*	+	,	–	.	/
3	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
6		a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7	p	q	r	s	t	u	v	w	x	y	z	{	|	}	~	
8	А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П
9	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я
A	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п
B				│	┤	╡	╢	╖	╕	╣	║	╗	╝	╜	╛	┐
C	└	┴	┬	├	─	┼	╞	╟	╚	╔	╩	╦	╠	═	╬	╧
D	╨	╤	╥	╙	╘	╒	╓	╫	╪	┘	┌	█	▄	▌	▐	▀
E	р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я

В настоящее время получил распространение Unicode способ кодирования. Его поддерживают Internet и операционная система Windows, а язык программирования Java использует его как основной. Он включает ASCII как подмножество (с кодами от 0000 до 007F). Греческие символы в нем закодированы кодами с шестнадцатеричными номерами 0391 – 03F7, а буквы кириллицы с кодами 0410 – 044F. Применение этого стандарта позволяет закодировать очень большое число символов разных языков: в документах Unicode могут соседствовать китайские иероглифы, математические символы, буквы греческого алфавита, латиницы и кириллицы, при этом становится ненужным переключение кодовых страниц.

Имеется еще UTF-8 представление Юникода. Оно обеспечивает наилучшую совместимость со старыми системами, использовавшими 8-битные символы. Текст, состоящий только из символов с номером меньше 128, при записи в UTF-8 превращается в обычный текст ASCII. Т.о. в тексте UTF-8 любой байт со значением меньше 128 изображает символ ASCII с тем же кодом. Остальные символы Юникода изображаются последовательностями длиной от 2 до 6 байт, в которых первый байт всегда имеет вид 11xxxxxx, а остальные — 10xxxxxx. Кодировка UTF-8 относится к классу кодировок с расширением. Расширение дает простой выход из положения, но при массовом использовании и нестандартных наборах символов оно не всегда является эффективным.

Кодирование multimedia. В связи с развитием multimedia — всевозможных периферийных аудио- и видеоустройств — появились многочисленные форматы кодирования для звуковых файлов и видео (MP3, AVI и др.).

8.3 Сжатие информации. Код Шеннона-Фано и алгоритм Хаффмена.

Еще об одной группе форматов нужно сказать особо — о форматах сжатия. В связи с желанием экономить место появились многочисленные способы сжатия информации – более компактной ее записи. Сейчас эти способы хорошо известны по так называемым архиваторам — сервисным программам, сжимающим файлы. Их основу представляет идея К. Шеннон и Р. Фано, которые предложили в наиболее чистом виде конструкцию кода переменной длины. В этом коде у каждого символа своя длина кодовой последовательности, как в азбуке Морзе. Но в отличие от азбуки Морзе, где конец кодовой последовательности определяется "третьим символом" — паузой, здесь нужно побеспокоиться о том, как определять завершение кода отдельного символа. Предлагается такое ограничение на код: никакая кодовая последовательность не является началом другой кодовой последовательности. Это свойство называется свойством префикса, а код, обладающий таким свойством, называется префиксным кодом.

Поясним смысл префиксного кода. Вначале напомним определения. Рассмотрим конечное множество X ($|X|=m$) и его декартовую степень $X \times X \times \ldots \times X = X^k$. В таком случае множество X называется алфавитом, его элементы – буквами, а элементы декартова произведения X^k, т.е.

упорядоченные пары из k букв ($x_1,x_2,\ldots,x_k$) называют словами в алфавите X. Число k называется длиной слова. В теме КОМБИНАТОРИКА мы доказали, что *число слов длины k в алфавите из m букв равно* m^k т.е. $|X^k| = |X|^k$.

Множество слов $\alpha = a_1 a_2 .. a_k$ в алфавите X обозначается X^*. Для слова $\alpha = \alpha_1\alpha_2$, составленного из двух слов $\alpha_1 = a_1..a_{k_1}$ и $\alpha_2 = c_1..c_{k_2}$, слово α_1 называется префиксом (началом), а слово α_2 - постфиксом (окончанием). Если при этом $\alpha_1 \neq \Lambda$ (пустому слову Λ), то α_1 называют собственным началом слова α. Соответственно, если $\alpha_2 \neq \Lambda$, то α_2 называют собственным окончанием слова α.

Пусть заданы алфавиты A=$\{a_1, a_2,\ldots,a_n\}$, B=$\{b_1, b_2,\ldots, b_m\}$ и функция $F : S \to B^*$ отображающая некоторое множество слов S в алфавите A $\left(S \subset A^*\right)$ в слова в алфавите B. Тогда функция F называется кодированием, элементы множества S – сообщениями, а элементы $\beta = F(\alpha), \alpha \in S, \beta \in B^*$ - кодами сообщений. Обратная функция, если она существует, называется декодированием.

Кодирование может сопоставлять код всему сообщению как единому целому или же строить код сообщения из кодов его частей. Элементарной часть сообщения является одна буква исходного алфавита A. Если кодирование сообщения выполняется путем кодирования каждой его буквы, то оно называется *алфавитным кодированием*. Алфавитное (или побуквенное) кодирование задается схемой (или таблицей) кодов σ : $\sigma = \{a_1 \to \beta_1,\ldots,a_n \to \beta_n\}$, где $a_i \in A$ - буквы алфавита A и $\beta_i \in B^*$ - некоторые слова, составленные из букв алфавита B. Множество кодов $V = \{\beta_i\}$ (слов β_i) называется множеством элементарных кодов. Если F – кодирование с таблицей σ, то для любого слова $\alpha = a_{i_1}..a_{i_k} \in A^*$ его кодом будет $F(\alpha) = F(a_{i_1})..F(a_{i_k}) = \beta_{i_1}...\beta_{i_k}$, т.е. каждая буква a_{i_j} слова α заменяется соответствующим ему словом β_{i_j} в кодовой таблице.

Пример 1. Рассмотрим алфавиты $A = \{0,1,2,3,4,5,6,7,8,9\}$, $B = \{0,1\}$ и схему

$$\sigma = \begin{Bmatrix} 0 \to 0, 1 \to 1, 2 \to 10, 3 \to 11, 4 \to 100, 5 \to 101, \\ 6 \to 110, 7 \to 111, 8 \to 1000, 9 \to 1001 \end{Bmatrix}$$

Кодирование не является взаимно-однозначным, т.к. $F(333) = 111111$ и $F(77) = 111111$. Это значит, что невозможно декодирование.

Пример 2. С другой стороны, схема

$$\sigma = \begin{Bmatrix} 0 \to 0000, 1 \to 0001, 2 \to 0010, 3 \to 0011, 4 \to 0100, 5 \to 0101, \\ 6 \to 0110, 7 \to 0111, 8 \to 1000, 9 \to 1001 \end{Bmatrix}$$

известная под названием «двоично-десятичное кодирование», допускает однозначное декодирование.

Рассмотрим схему алфавитного кодирования σ и различные слова, составленные из элементарных кодов. Схема σ называется разделимой, если любое слова, составленное из элементарных кодов β_i единственным образом разлагается на элементарные коды, т.е. если $\beta_{i_1}..\beta_{i_k}=\beta_{j_1}..\beta_{j_l}$, то отсюда следует, что $k=l$ и $\forall\, t=1,...,l$ имеет место $i_t=j_t$.

Алфавитное кодирование с разделимой схемой допускает декодирование. При этом имеется ввиду, что мы не будем рассматривать таблицы элементарных кодов V, в которых есть одинаковые коды, т.е. $\exists\, i\neq j\ \ \beta_i\neq\beta_j$ иначе, очевидно, однозначное декодирование невозможно.

Схема σ называется префиксной, если элементарный код одной буквы не является префиксом (начальной частью) кода другой буквы, т.е. для любой пары элементарных кодов $\beta_i,\beta_j\in V\ (i\neq j)$ не существует слова β такого, что $\beta_i=\beta_j\beta$ или $\beta_j=\beta_i\beta$.

Теорема. *Префиксная схема является разделимой.*

Доказательство от противного. Пусть кодирование с префиксной схемой σ не является разделимым. Тогда существует такое слово β

$$\beta=\beta_{i_1}..\beta_{i_k}=\beta_{j_1}..\beta_{j_l}=b_1..b_n$$

представимое некоторым набором букв алфавита B, который может быть составлен из разных последовательностей элементарных кодов β_i. Тогда до некоторой позиции *t-1* элементарные коды слева и справа совпадают и, начиная с позиции *t* элементарного кода β_{it}, мы будем иметь $\beta_{it}...\beta_{i_k}=\beta_{j_t}..\beta_{j_l}$ совпадение последовательности букв алфавита B, но несовпадение последовательности элементарных кодов. То есть имеет место

$$\underbrace{b_1..b_k}_{\beta_{i_t}}b_{t+1}...=\underbrace{b_1..b_k..b_l}_{\beta_{j_t}}b_{l+1}...\ \text{ или }\ \underbrace{b_1..b_k..b_l}_{\beta_{i_t}}b_{l+1}...=\underbrace{b_1..b_k}_{\beta_{j_t}}b_{k+1}...,$$

в зависимости от того, длина какого из слов (элементарных кодов) β_{i_t} или β_{j_t} короче. Но это значит, что существует слово β' такое, что $\beta_{j_t}=\beta_{i_t}\beta'$ или $\beta_{i_t}=\beta_{j_t}\beta'$. Т.е. либо β_{j_t} оказалось префиксом β_{i_t}, либо β_{i_t} оказалось префиксом β_{j_t}. Но это противоречит тому, что схема префиксная.

Замечание. Свойство быть префиксной схемой является достаточным, но не является необходимым для разделимости схемы.

Пример 3. Например, разделимой, но не префиксной является схема A={*a*, *b*}, B={0, 1}, $\sigma=\{a\to0,b\to01\}$. Если в закодированном слове встречается единица, то ей обязательно должен предшествовать ноль – самостоятельной единице в таблице σ нет. Это значит что декодирование возможно.

Приведем без доказательства следующий результат. Чтобы схема алфавитного кодирования была разделимой, необходимо, чтобы длины элементарных кодов удовлетворяли определенному соотношению, известному как неравенство Макмиллана.

Теорема. Если схема $\sigma = (a_i \to \beta_i)_{i=1}^n$ разделима, то $S = \sum_{i=1}^{n} \frac{1}{2^{s_i}} \le 1$, где $s_i = |\beta_i|$.

Пример 4. Для схемы из предыдущего примера А={a, b}, В={0, 1}, $\sigma = \{a \to 0, b \to 01\}$ имеем $S = \frac{1}{2^1} + \frac{1}{2^2} = \frac{3}{4} < 1$.

Пример 5. Для алфавитной схемы кодирования из примера 2

$$\sigma = \begin{Bmatrix} 0 \to 0000, 1 \to 0001, 2 \to 0010, 3 \to 0011, 4 \to 0100, 5 \to 0101, \\ 6 \to 0110, 7 \to 0111, 8 \to 1000, 9 \to 1001 \end{Bmatrix}$$

имеем $S = 10 \cdot \frac{1}{2^4} = \frac{10}{16} < 1$

В обратную сторону теорема формулируется следующим образом.

Теорема. Если числа $s_1, \ldots, s_n$ удовлетворяют неравенству $\sum_{i=1}^{n} \frac{1}{2^{s_i}} \le 1$, то существует разделимая схема алфавитного кодирования $\sigma = (a_i \to \beta_i)_{i=1}^n$, где $\forall i \;\; s_i = |\beta_i|$.

На практике важно, чтобы коды сообщений имели по возможности минимальную длину. Если про множество передаваемых слов $S=A^*$ ничего неизвестно, то сформулировать задачу оптимизации невозможно. Однако часто доступна дополнительная информация. Например, для текстов на разговорных языках известны вероятности (частоты) появления букв в сообщении. Использование этой информации позволяет строго сформулировать и решить задачу построения оптимального алфавитного кодирования.

В предположении, что кодируемые символы появляются в тексте независимо, нужно стремиться уменьшать среднее число битов, приходящееся на один символ. Это значит, что надо уменьшить математическое ожидание (средняя частота появления) длины кодовой комбинации случайно выбранного символа, которое вычисляется по формуле

$$\sigma = \sum_i p_i s_i$$

где p_i — вероятность (частота появления), а s_i – длина кодовой последовательности i-го символа. Величина σ называется ценой (или длиной) кодирования при заданном распределении вероятностей.

Пример 7. Для разделимой схемы А={a, b}, В={0, 1}, $\delta = \{a \to 0, b \to 01\}$ при вероятностях (0.5, 0.5) цена кодирования составляет $0.5 \cdot 1 + 0.5 \cdot 2 = 1.5$, а при вероятностях (0.9, 0.1) она равна $0.9 \cdot 1 + 0.1 \cdot 2 = 1.1$.

Всегда можно построить схему равномерного алфавитного кодирования, т.е такую что $\forall i{:}\, |\beta_i| = L$, где L некоторое фиксированное число, зависящее от числа символов $A = \{a_1, \ldots, a_n\}$, для которой очевидно $\sigma = L$ (предполагается, что $p_1 + \ldots + p_n = 1$). Схемы с переменной длиной кодовых последовательностей будут иметь $1 \le \sigma \le L$. Поскольку таких схем конечное количество, то среди них есть оптимальная, т.е. имеющая минимальное значение σ.

Шеннон и Фано предложили строить код, близкий к оптимальному, следующим способом: разбить все символы на две группы с приблизительно равными частотами появления, коды первой группы начать с 0, а второй группы — с 1; внутри каждой группы делать то же самое, пока в каждой группе не останется только по одному символу.

Пример 8. Коды, построенные Фано для заданного распределения вероятностей (n=7)

p_i	Первый символ	Второй символ	Третий символ	Код	l_i
0.20	0	0		00	2
0.20	p=0.39 0	1	0	010	3
0.19	0	1	1	011	3
0.12	1	0	0	100	3
0.11	1	0	1	101	3
0.09	p=0.41 1	1	0	110	3
0.09	1	1	1	111	3

Здесь вертикальные линии (одинарная и двойная) выделяют подгруппу символов из текущей группы. Символам одной группы назначается одинаковое начальное значение ноль или один. Цена всего кодирования равна

$$\sigma = 2 \cdot 0.2 + 3 \cdot (0.2 + 0.19 + 0.12 + 0.11 + 0.09 + 0.09) = 2.8.$$

В результате в сообщении, записываемом с использованием семи символов с заданными вероятностями, средняя длина кода одного символа может быть равна 2.8 бита.

Элегантный алгоритм для точного решения этой задачи предложил Д.Хаффмен. Алгоритм основывается на нескольких очевидных свойствах оптимального набора $\{s_i\}$.

Лемма 1. Пусть $\{p_i\}$ — набор вероятностей символов и $\{s_i\}$ — длины оптимальных кодовых комбинаций. Если $p_1 \ge p_2 \ge \ldots \ge p_n$, то $s_1 \le s_2 \le \ldots \le s_n$. Доказательство леммы следует из решения следующей задачи.

Задача о минимуме скалярного произведения. Пусть заданы m чисел $x_1, x_2, \ldots, x_m$ и еще m чисел $y_1, y_2, \ldots, y_m$. Составим пары (x, y), включив каждое x_i, и каждое y_i ровно в одну пару. Затем перемножим числа каждой пары и сложим получившиеся произведения. Требуется найти такое разбиение чисел на пары, при котором значение получившейся суммы будет наименьшим.

Теорема. Наименьшее значение суммы попарных произведений достигается при сопоставлении возрастающей последовательности x_i, убывающей последовательности y_i.

Доказательство. Покажем, что если найдутся две пары чисел (x_i, y_i) и (x_j, y_j) такие что $x_i < x_j$ и $y_i < y_j$, то значение суммы попарных произведений можно уменьшить, заменив эти две пары парами (x_i, y_j) и (x_j, y_i). Действительно для суммы попарных произведений имеем $S = ..x_i y_i + ... + x_j y_j + ...$. Но так как $(x_j - x_i)(y_j - y_i) > 0$, то, раскрывая скобки, получим $x_j y_j - x_i y_j - x_j y_i + x_i y_i > 0$ или $x_j y_j + x_i y_i > x_i y_j + x_j y_i$. Это значит, что сумму S можно уменьшить до величины $S' = ..x_i y_j + ... + x_j y_i + ...$. Если в сумме снова найдутся пары такие что $x_i < x_j$ и $y_i < y_j$, то процесс можно продолжить. Поскольку число возможных расположений пар равно m!, т.е. конечное число, то за конечное число шагов мы закончим процесс улучшений на таком расположении, для которого пар, таких что $x_i < x_j$ и $y_i < y_j$ не существует. На нем и достигается минимум.

Доказательство леммы. Достаточно сослаться на теорему о перестановке, минимизирующей скалярное произведение двух векторов и на предположение, что код префиксный.

Лемма 2. В обозначениях и предположении леммы 1 две самые длинные кодовые комбинации имеют одинаковую длину, т.е. $s_{n-1} = s_n$.

Доказательство. Действительно, пусть $s_n > s_{n-1}$ и, следовательно, n-я кодовая комбинация — самая длинная. Никакая кодовая комбинация не является началом никакой другой. Тогда Первые n−1 символ n-й кодовой комбинации не совпадают с (n−1)-й кодовой комбинацией. Это значит, что n-ю кодовую комбинацию можно сократить с конца на один символ, т.е. до длины (n−1)-ой. В результате мы получим снова уникальную кодовую комбинацию и более короткую, чем раньше, что невозможно для оптимальной кодировки. Поэтому $s_n = s_{n-1}$.

Лемма 3. Рассмотрим наравне с исходной задачей P сокращенную задачу P' которая получается объединением двух самых редких символов в один символ, — в предположении леммы 1 это два последних символа с суммарной вероятностью $p'_{n-1} = p_{n-1} + p_n$. Минимальное значение целевой функции в задаче P' отличается от значения в задаче P на p'_{n-1}, а оптимальный кодовый набор для задачи P получается из решения для задачи P' удлинением на один бит кодовой последовательности для объединенного символа.

Доказательство. Действительно, каждому кодовому набору для задачи P' можно так, как сказано в утверждении леммы, сопоставить кодовый набор для задачи P с равными длинами самых редких символов. Это удлинение и приводит к приращению целевой функции.

Теперь можно рассуждать так: в задаче P мы ищем минимум на множестве кодов, которое мы обозначим через C_n а в задаче P' — аналогично на множестве C_{n-1}. В соответствии с леммой 2 в задаче P можно искать минимум на множестве $C`_n$ собственном подмножестве $C_n$, состоящем из таких наборов, в которых две самые длинные кодовые последовательности имеют одинаковую длину. Имеется взаимно-однозначное соответствие между $C_{n-1}$ и $C`_n$, и значения целевых функций на соответствующих элементах двух множеств отличаются на постоянное слагаемое. Следовательно, и минимумы отличаются на это же постоянное слагаемое, так что минимуму в задаче P` соответствует минимум на $C`_n$, а он будет минимумом и для задачи P.

■

Алгоритм, основанный на этих леммах, описывается в несколько строк. Если в алфавите два символа, то нужно закодировать их 0 и 1. Если больше, то соединить два самых редких символа в один новый символ, решить получившуюся задачу и вновь разделить этот новый символ, приписав 0 и 1 к его кодовой последовательности.

Пример 9. Пусть алфавит состоит из пяти символов – *a, b, c, d, e*, вероятности которых равны

$$0.37\,(a), 0.22\,(b), 0.16\,(c), 0.14\,(d), 0.11\,(e).$$

Объединяя *d* и *e* в один символ, имеем

старый список: $0.37\,(a), 0.22\,(b), 0.16\,(c)$;

новый список: $0.25\,(d\,e)$.

На следующем шаге объединяем *b* и *c*:

старый список: $0.37\,(a)$;

новый список: $0.25\,(d\,e),\ 0.38\,(b\,c)$

Затем объединяем *a* и *de*:

старый список: пуст;

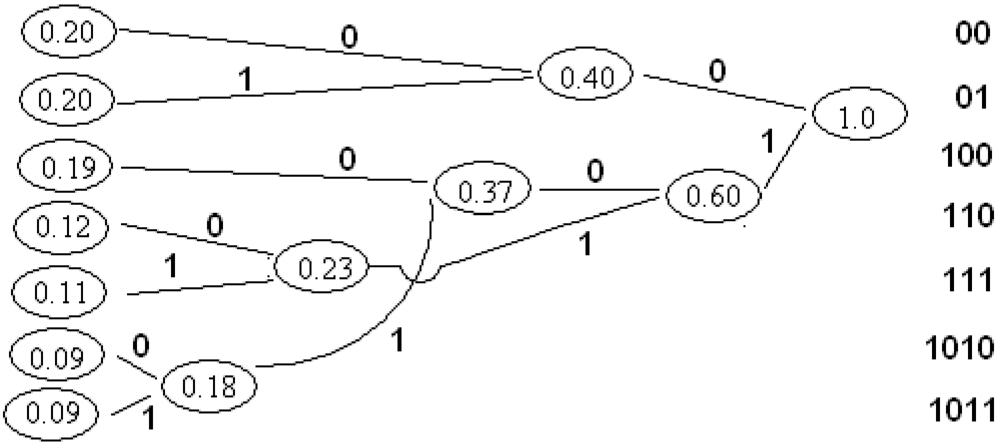
новый список: $0.38\,(b\,c),\ 0.62\,(a\,d\,e)$

Сопоставим *ade* код 0, а *bc* – код 1 и выполним «обратный ход». Расщепим символы *ade* на *a* и *de* с кодами 00 и 01. Затем символ *bc* на *b* и *c* с кодами 10 и 11. Наконец, расщепив *de*, получим для исходного алфавита

$$00\,(a), 10\,(b), 11\,(c), 010\,(d), 011\,(e).$$

Обратите внимание на то, что в кодируемом тексте символы (в том числе составные) могут иметь равные вероятности, а из-за этого может существовать несколько различных оптимальных кодовых деревьев. У всех этих деревьев математическое ожидание числа битов на один символ будет одинаковым.

Пример 10. Пусть алфавит состоит из 7 символов. Процесс кодирования по Хаффману можно представить в виде алгоритма построения «кодового» дерева следующим образом. Объединяем пары символов с наименьшими вероятностями в узел. Этому узлу приписываем вероятность, равную сумме вероятностей исходных символов. На следующем шаге объединяем символы или узлы с минимальными вероятностями, которые образуют следующий узел. Ему присваиваем вероятность, равную сумме вероятностей составляющих. Процесс повторяется до тех пор, пока не сойдется в один узел с единичной суммарной вероятностью. Этот узел будет вершиной дерева. Для нашего примера это дерево показано на следующем рисунке.



Затем, начиная с вершины, приписываем верхним ребрам нули, а нижним – единицы (или наоборот). Значение кода символа формируется в виде последовательности кодов ветвей (нулей и единиц), начиная с вершины дерева. Полученные коды мы привели справа от дерева. Обратите внимание, что более короткие коды не встречаются в виде префикса более длинных кодов.

Цена кодирования в нашем примере будет равна

$$0.2\times2+0.2\times2+0.19\times3+0.12\times3+0.11\times3+0.09\times4+0.09\times4=2.78,$$

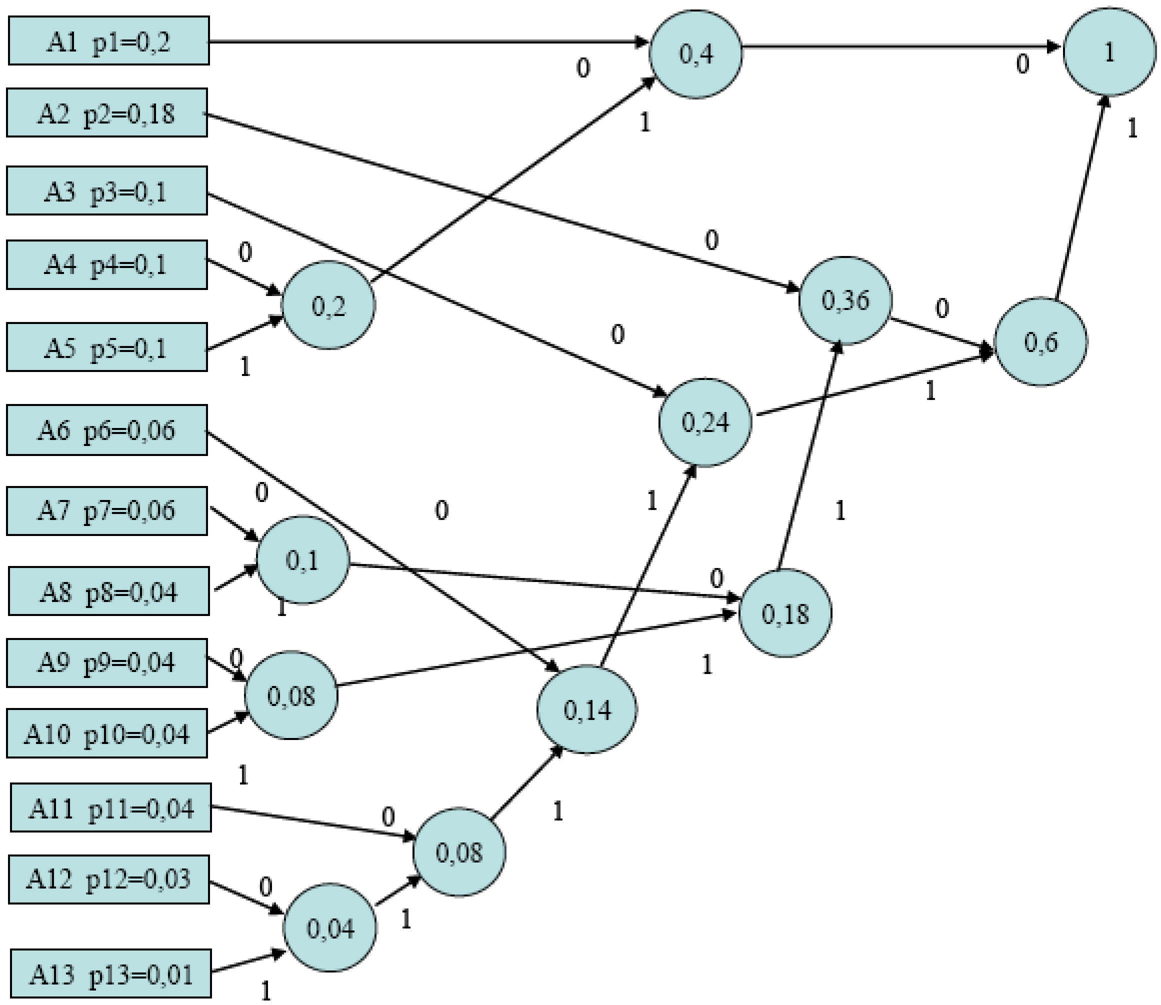
что несколько лучше, чем в кодировании, полученном нами алгоритмом Фано в примере 8.

Метод кодирования Хаффмена относится к группе методов сжатия данных без потерь информации. Он применяется также при записи графических изображений в файлы и является компонентом алгоритмов сжатия данных *JPEG* и *MPEG*. Метод предложен Дэвидом Хаффманом в 1952 г. Особенностью метода является использование кодов переменной длины, при этом наиболее вероятным символам присваиваются наиболее короткие кодовые слова, а менее вероятным – длинные. Благодаря такой стратегии, код Хаффмана при алфавитном кодировании дает минимально возможную среднюю длину кодовой последовательности. Рассмотрим построение кодовой таблицы еще на одном примере.

Пример 11. Пусть кодируется 13 символов упорядоченных по невозрастанию вероятности их появления: $\{A_1, A_2,\ldots,A_{13}\}$. Пусть вероятности (частоты) появления A_i обозначаются *pi* и составляют {0,2; 0,18; 0,1; 0,1; 0,1; 0,06; 0,06; 0,04; 0,04; 0,04; 0,04; 0,03; 0,01} соответственно.

Опишем алгоритм с использованием символики теории графов еще раз. Объединяем два символа с наименьшими вероятностями в узел кодового дерева, этому узлу приписываем вероятность, равную сумме вероятностей появления символов, составляющих узел. В примере минимальную вероятность имеют символы A_{12} и A_{13}, суммарная вероятность которых равна 0.04. Далее объединяем символы или узлы с минимальными вероятностями, образуя следующий узел, которому присваиваем вероятность, равную сумме вероятностей составляющих.

Процесс повторяется до тех пор, пока не сойдется в один узел. После этого левые и правые ветви дерева обозначаются «0» и «1» соответственно (или наоборот, «1» и «0»). Значение кодового слова формируется в виде последовательности кодов ветвей по пути от вершины кодового дерева к узлу. Кодовое дерево для приведенного примера представлено на следующем рисунке.



В результате кодирования по методу Хаффмана получены коды, приведенные в следующей таблице.

Таблица кодов для последовательности {A1,...,A13}

Имя	Код	S_i	p_i	Имя	Код	S_i	p_i
A1	00	2	0,2	A8	10101	5	0,04
A2	100	3	0,18	A9	10110	5	0,04
A3	110	3	0,1	A10	10111	5	0,04
A4	010	3	0,1	A11	11110	5	0,04
A5	011	3	0,1	A12	111110	6	0,03
A6	1110	4	0,06	A13	111111	6	0,01
A7	10100	5	0,06				

Здесь *Si* - длина кодового слова *i*-го символа (количество разрядов). Таблица кодов построена в соответствии с частотой появления символов во входной последовательности.

Код Хаффмана является префиксным кодом, то есть таким, что ни одно кодовое слово не является префиксом (началом) другого кодового слова. Это можно проверить, например, код символа А2, равный 100, не стоит в начале кода никакого другого символа.

Средняя длина кода вычисляется по формуле

$$S_{cp} = \sum_{i=1}^{N} p_i S_i$$

и в рассмотренном примере средняя длина кода составляет 3.42 разряда.

Процедура кодирования выполняется в два этапа. Вначале оцениваются частоты появления символов в последовательности, и по ним строится кодовая таблица. Затем, на втором этапе, производится собственно кодирование.

Для декодирования необходимо передать с потоком кодовую таблицу, содержащую алфавит (все символы от А1 до А13) и соответствующие им коды. Часто кодовую таблицу не передают, а применяется заранее составленная на основании статистических данных кодовая таблица.

Префиксные коды декодируются однозначно, но обратное утверждение не выполняется. Например, код {0, 01, 010, 0101, …,} не является префиксным, хотя декодируется однозначно.

Казалось бы, что код Хаффмана невозможно улучшить – при алфавитном кодировании он дает оптимальную кодовую таблицу. Но понимать под алфавитом можно не только наборы букв реального языка. Можно, например, к кодируемому алфавиту присоединить пары или тройки букв, которые наиболее часто встречаются в конкретном кодируемом тексте. И к полученному набору применить алгоритм Хаффмана. Оказывается, что этот путь является продуктивным и дает лучшее сжатие текстов, чем простое кодирование только одних букв. Алгоритмы, использующие эту идею, получили в настоящее время наибольшее распространение. Один из них называется методом сжатия Зива – Лемпеля и его модифицированный близнец – алгоритм Лемпеля — Зива — Велча (LZW).

Упражнения.

1. Является ли схема алфавитного кодирования

$$\left(a \to 0, b \to 10, c \to 011, d \to 1011, e \to 1111\right)$$

префиксной? разделимой?

2. Построить оптимальное префиксное алфавитное кодирование для алфавита $\{a, b, c, d\}$ со следующими вероятностями появления букв

$$p_a = 1/2,\ p_b = 1/4,\ p_c = 1/8,\ p_d = 1/8.$$