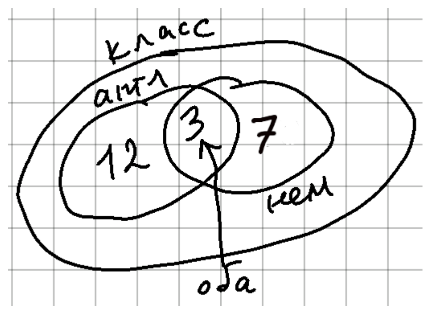
6. Включения и исключения

$\boxed{\textbf{6.1}}$ В классе 10 учеников знают немецкий, 15 знают английский, и среди них есть трое, которые знают оба языка (они входят в 10 и 15, о которых шла речь). Сколько учеников знает хотя бы один из этих двух языков?

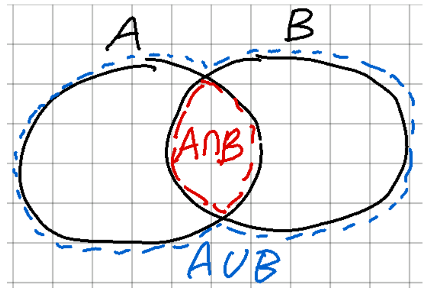
▷ Если мы сложим 10 и 25, то тех, кто знает оба языка, мы посчитаем дважды, а нужно только один раз. Значит, их надо вычесть: получаем, что $10 + 15 - 3 = 22$ человек знают хотя бы один из двух языков. ◁



• Можно объяснить иначе. Ученики в классе делятся на четыре непересекающиеся группы:

- знающие оба языка;
- знающие только немецкий;
- знающие только английский;
- не знающие ни немецкого, ни английского.

К первой группе по условию относятся 3 человека. Вместе со второй группой получаем всех знающих немецкий, их 10, так что во второй группе 7. Первая с третьей — знающие английский, их 15, значит в третьей 12. Ответ — сумма первой, второй и третьей группы, то есть $3+7+12 = 22$. (Заметим, что из условий задачи нельзя определить, сколько человек в четвёртой группе.)



На более математическом языке (мы уже упоминали это в разделе 1) говорят так. Есть два конечных множества A и B (тех, кто знает немецкий, и тех, кто знает английский). Тогда можно рассмотреть их *пересечение* $A \cap B$ (тех, кто знает оба языка) и *объединение* $A \cup B$ (тех, кто знает хотя бы один язык). Тогда

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

где через $|X|$ обозначено число элементов в (конечном) множестве X. В нашем примере было $22 = 15 + 10 - 3$.

• Аналогичное утверждение можно сформулировать для площадей: площадь объединения фигур A и B на рисунке равна сумме площадей фигур минус площадь их пересечения.

Можно задать такой же вопрос для большего числа множеств — известны размеры множеств и размеры всех пересечений, надо найти размер объединения. Можно было бы задать такой вопрос про английский, немецкий и французский, но чтобы не запутывать дело, будем сразу говорить о множествах.

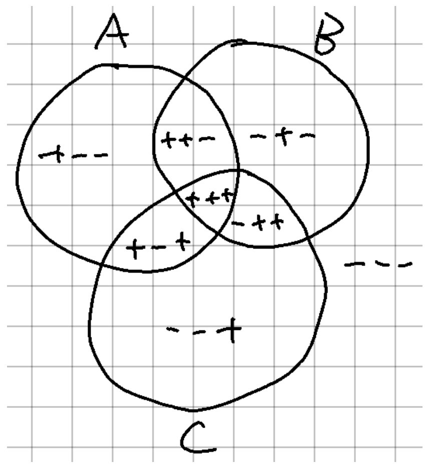
6.2 Докажите, что

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

для любых трёх конечных множеств A, B, C.

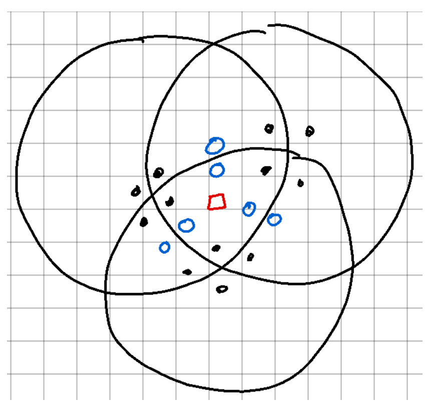
Словами: чтобы найти число людей, знающий хотя бы один язык из трёх, надо сложить числа людей, знающих немецкий, знающих французский и знающих английский, затем вычесть числа людей, знающих одновременно английский и немецкий, английский и французский, немецкий и французский, и, наконец, прибавить число людей, знающих все три языка.

▷ Как и раньше, можно посмотреть, правильно ли будут учтены элементы всех сортов. Теперь сортов 8 (для каждого множества есть два варианта: принадлежит или нет; каждый язык можно знать или не знать). Можно нарисовать аналогичную диаграмму (и проверить, что она «представительна» в том смысле, что представлены все восемь вариантов)



Здесь плюсами и минусами отмечена принадлежность и непринадлежность трём множествам (слева направо: A, B, C), видно, что для любой комбинации из трёх плюсов или минусов есть своя область (ровно одна).

Теперь надо проверить, что во всех частях учёт (согласно правой части доказываемой формулы) проводится правильно.



Отметим точками части, учтённые в $|A| + |B| + |C|$: видно, что знающие ровно один язык учтены правильно, знающие ровно два языка учтены дважды, а знающие три языка учтены трижды. После того как мы вычтем попарные пересечения (кружочки), знающие ровно два языка будут вычтены один раз (они попадут только в одно попарное пересечение), а знающие три языка будут вычтены три раза. Наконец, при добавлении тройного пересечения (квадратик) всё будет учтено правильно ($2-1$ и $3-3+1$ равны единице). ◁

6.3 Как изменятся слагаемые в правой части, если из множества A удалить элемент, который не входил ни в B, ни в C? Если удалить элемент, который входил в A и B, но не входил в C? Если удалить элемент, который входил во все три множества?

▷ Если удалённый элемент входил только в A, то изменится только $|A|$ (уменьшится на единицу). Если удалённый элемент входил в A и B (но не в C), то уменьшатся на единицу A, B и $A \cap B$ (и последнее изменение частично скомпенсирует два первых: выражение в целом уменьшится на единицу). Наконец, если элемент входил во все три множества, то все слагаемые уменьшатся на единицу (и сумма тоже уменьшится на единицу, так как слагаемых с плюсом на одно больше). ◁

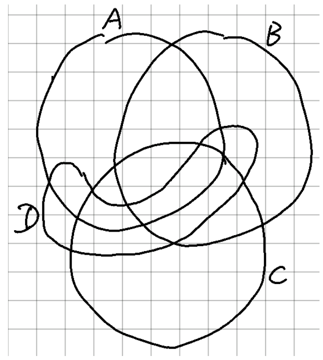
• Эта задача показывает, что во всех трёх случаях правая часть уменьшается на единицу — синхронно с левой. Если по очереди удалить все элементы, то она станет равной нулю — значит, изначально она была равна числу элементов в $A \cup B \cup C$, как и утверждает формула включений и исключений.

Теперь уже можно догадаться, как будет выглядеть формула для четырёх множеств: чтобы найти число элементов в их объединении, надо

- сложить числа элементов в этих множествах;
- вычесть все шесть попарных пересечений;
- добавить четыре тройных пересечения;
- наконец, вычесть пересечение всех четырёх множеств.

(Понятно, почему попарных пересечений шесть? потому что $C_4^2 = 6$.)

▷ Чтобы доказать эту формулу, тоже можно нарисовать «представительную» картинку для четырёх множеств, где были бы все 16 вариантов принадлежности и непринадлежности. Это уже не так просто, но если не настаивать на симметрии, то можно:



По этой картинке тоже можно посчитать, сколько раз вошла и вышла каждая часть. Но нам хотелось бы доказать аналогичную формулу для любого числа множеств, так что надо искать другое доказательство. ◁

6.4* Докажите общую *формулу включений и исключений*:

$$|A_1\cup\ldots\cup A_n| = \sum_i |A_i| - \sum_{i<j} |A_i\cap A_j| + \sum_{i<j<k} |A_i\cap A_j\cap A_k| - \ldots - (-1)^n |A_1\cap\ldots\cap A_n|.$$

Здесь написано, что для подсчёта размера объединения надо сложить размеры множеств, вычесть размеры попарных пересечений, прибавить размеры тройных пересечений и так далее (последний знак будет плюс при нечётном n и минус при чётном n, поэтому там и стоит коэффициент $-(-1)^n$).

▷ Есть разные способы это доказать, но раз уж мы изучали свойства биномиальных коэффициентов, то можно их использовать. Пусть у нас есть какой-то элемент, который принадлежит s множествам среди $A_1, \ldots, A_n$, а остальным $(n-s)$ множествам не принадлежит. Надо понять, сколько раз он будет учтён на каждом шаге, и проверить, что вместе получится правильно (однократный учёт).

На первом шаге, раз он принадлежит s множествам, он будет учтён (с плюсом) s раз. На втором шаге он будет вычтен столько раз, скольким попарным пересечениям он принадлежит. Эти пересечения — пересечения пар, выбранных среди s множеств, поэтому их C_s^2. На третьем шаге он будет добавлен столько раз, сколько есть троек среди s множеств, то есть C_s^3. И так далее — последний раз он будет добавлен или вычтен один раз, это соответствует $C_s^s = 1$ (а знак зависит от чётности s). Можно ещё заметить, что число s на первом шаге есть C_s^1. Общее число раз будет, таким образом,

$$C_s^1 - C_s^2 + C_s^3 - \ldots \pm C_s^s.$$

Это знакопеременная сумма биномиальных коэффициентов из одной строки треугольника Паскаля, только не всех, а кроме первого ($C_s^0 = 1$). Если бы первый добавить с минусом, то получилась бы знакопеременная сумма всех, равная нулю (задача 5.5), так что без него эта сумма равна единице, как и требуется. ◁

6.5* Докажите, что выражение

$$\sum_i |A_i| - 2\sum_{i<j} |A_i \cap A_j| + 4 \sum_{i<j<k} |A_i \cap A_j \cap A_k| - \ldots - (-2)^{n-1}|A_1 \cap \ldots \cap A_n|.$$

(добавлены степени двойки) тоже имеет комбинаторный смысл: оно равно числу элементов, которые входят в *нечётное* число множеств $A_1, \ldots, A_n$. (Если множества A_i не пересекаются, то по-прежнему получается сумма.)

▷ Можно использовать тот же метод и подсчитать, с каким коэффициентом будет учтён некоторый элемент x, принадлежащий ровно s множествам. Сначала он будет учтён $s = C_s^1$ раз, потом будет вычтен C_s^2 раз с коэффициентом 2, потом добавлен C_s^3 раз с коэффициентом 4, и так далее. Суммарный коэффициент будет

$$A = C_s^1 - 2C_s^2 + 4C_s^3 - \ldots + (-2)^{s-1}C_s^s$$

Это почти бином Ньютона для $(1-2)^s$:

$$(1-2)^s = 1 - 2C_s^1 + 4C_s^2 - 8C_s^3 + \ldots + (-2)^s C_s^s = 1 - 2A$$

Поэтому при чётных s это выражение равно 1 и $A = 0$, а при нечётных s это выражение равно -1 и $A = 1$, так что мы учитываем по разу все элементы, входящие в нечётное число множеств A_i (а входящие в чётное число — не учитываем). ◁

Приведём два примера, где можно использовать формулу включений и исключений: (1) перестановки без неподвижных точек и (2) слова, использующие все буквы (сюръекции).

6.6* Слово состоит из n различных букв. Сколько разных слов можно составить, переставляя эти буквы, если дополнительно требуется, чтобы ни одна буква не осталась на своём прежнем месте?

▷ Можно было бы спросить, сколькими способами n писем для n адресатов можно разложить по n надписанным конвертам, если нужно, чтобы каждый адресат получил бы неправильное письмо. (А математики сказали бы коротко «найти число перестановок без неподвижных точек».) ◁

▷ Пусть A_i — те варианты, когда i-я буква остаётся на своём месте. Тогда объединение всех A_i будет состоять из слов, которые нам не годятся (где есть буква, остающаяся на месте). Если мы будем знать, сколько слов в этом объединении, то вычтя это число из $n!$, получим число годных слов.

В каждом A_i переставляются произвольным образом все буквы, кроме i-й (остальные буквы могут оставаться на своём месте или нет, для A_i это не важно). Значит, в нём $(n-1)!$ элементов. Попарные пересечения $A_i \cap A_j$ состоят из слов, у которых обе буквы i и j остаются на месте (а остальные как хотят), в таком пересечении $(n-2)!$ элементов, и разных таких пересечений будет C_n^2. Тройных пересечений будет C_n^3, и в каждом $(n-3)!$ элементов. Теперь остаётся написать формулу включений и исключений: негодных слов будет

$$n \cdot (n-1)! - C_n^2(n-2)! + C_n^3(n-3)! + \ldots \pm C_n^{n-1}1! \mp C_n^n 0!$$

(два последних знака противоположны и зависят от чётности n). Чтобы упростить эту формулу, вспомним, что $C_n^k = n!/(k!(n-k)!)$ и сократим $(n-k)!$, а также вынесем из всех слагаемых $n!$, получится

$$n!(1 - 1/2! + 1/3! - 1/4! + \ldots \mp 1/n!).$$

Это число надо вычесть из $n!$, так что окончательный ответ будет

$$n!(1 - 1 + 1/2! - 1/3! + 1/4! - \ldots \pm 1/n!)$$

(первые два члена можно сократить, а можно для единообразия переписать как $1/0! - 1/1!$). ◁

▷ Знатоки математического анализа сразу скажут, что выражение в скобках с ростом n быстро приближается к значению $1/e$, где $e = 2{,}7182818284\ldots$ — так называемое «основание натуральных логарифмов», так что в примерно 37% случаев ни одна буква не стоит на своём месте.

(А знатоки теории вероятностей добавили бы, что в среднем в перестановке n букв есть одна неподвижная точка, так что не очень и удивительно, что около трети слов не имеют её — зато другие имеют больше одной.) ◁

А вот второй пример, где полезна формула включений и исключений.

Мы знаем, что слов длины n в алфавите из k букв (то есть последовательностей из n символов, выбираемых из k-элементного списка) всего k^n (для каждой буквы есть k вариантов). Наложим дополнительное ограничение: *в слове должны встречаться все буквы*. Сколько тогда останется слов?

Вопрос имеет смысл при $n \geqslant k$ (иначе мест для всех букв не хватит). Скажем, можно спросить (и мы сделаем это в следующей задаче), сколько существует десятизначных чисел с нечётными цифрами, в которых каждая из пяти нечётных цифр встречается не менее одного раза. Это соответствует $n = 10$, $k = 5$. Но нет смысла спрашивать, сколько бывает пятизначных чисел, в которых встречаются все 10 цифр — их не бывает.

В случае $k = n$ ответ будет $n!$ (или $k!$), потому что мест столько же, сколько букв, и если каждая буква встречается, то ровно один раз (иначе мест для других не хватит). Но как найти ответ при $n > k$?

6.7* Сколько существует десятизначных чисел, составленных из нечётных цифр, в которых каждая нечётная цифра использована (встречается хоть раз)? (Скажем, таковы числа 1975311111 или 5533997711, но не 1111335777.)

▷ Эту задачу можно решать, используя формулу включений и исключений, а можно написать рекуррентную формулу. Вот как это делается.

(Первое решение) Всего десятизначных чисел с нечётными цифрами 5^{10}. Из них надо вычесть те, где *хотя бы одна из нечётных цифр не встречается*, так что достаточно подсчитать их количество. Интересующее нас множество является объединением пяти множеств N_1, N_3, N_5, N_7 и N_9: числа, где не встречается цифра 1 (множество N_1), числа, где не встречается 3 (множество N_3), и так далее. Каждое из множеств имеет размер 4^{10} (скажем, N_1 состоит из слов длины 10 из букв $3, 5, 7, 9$). Но они пересекаются, поэтому нельзя просто сложить их размеры, чтобы найти размер объединения, надо использовать формулу включений и исключений.

В неё входят пересечения множеств, скажем, $N_1 \cap N_3$. Это пересечение состоит из чисел, в которых нет ни цифры 1, ни цифры 3 — то есть это слова длины 10 из букв $5, 7, 9$, и их будет 3^{10}. А самих этих пересечений будет C_5^2 — надо из пяти множеств выбрать два и их пересечь. Аналогично пересечений по три будет C_5^3, а каждое из них содержит 2^{10} элементов, и так далее (до пересечения всех пяти множеств, которое пусто, так как числа без цифр не бывает). Теперь формула включений и исключений даёт ответ:

$$5^{10} - C_5^1 \cdot 4^{10} + C_5^2 \cdot 3^{10} - C_5^3 \cdot 2^{10} + C_5^4 \cdot 1^{10}$$

(последний член с 0^{10} мы опустили). Теперь несложно подсчитать ответ (особенно есть есть калькулятор или какая-нибудь программа для математических вычислений), получится 5 103 000.

(Второе решение) Представим себе, что мы стёрли первую цифру в числе, удовлетворяющем условию (все нечётные цифры входят). Может быть два случая:

- в оставшемся числе по-прежнему есть все пять цифр;
- в оставшемся числе какой-то цифры нет (только одной — той, которую мы стёрли).

Первый случай сводит задачу к паре $(9, 5)$ (количество девятизначных чисел из нечётных цифр, в которых есть все цифры), второй случай разбивается на пять подслучаев в зависимости от отсутствующей цифры. Количество чисел в каждом подслучае соответствует нашей задаче для пары $(9, 4)$.

Эту же альтернативу можно сформулировать иначе: первая буква встречается больше одного раза (дальше есть такие же буквы) или только один раз.

Более точно, пусть $T(n, k)$ при $n \geqslant k$ — количество слов из n букв в k-буквенном алфавите, в которых встречаются все буквы. Тогда

$$T(n, k) = kT(n-1, k) + kT(n-1, k-1).$$

Первое слагаемое соответствует первому случаю: если после удаления первой буквы получается слово, где есть все k букв (что бывает $T(n-1, k)$ способами), то первую букву можно добавить любым из k способов.

Второе слагаемое (для второго случая): если после удаления первой буквы остаётся слово, в котором первой буквы нет, а все остальные есть, то таких слов $T(n-1, k-1)$, и надо ещё умножить на k вариантов для первой буквы.

Эта рекуррентная формула имеет смысл при $n > k$, при $n = k$ мы знаем ответ ($k!$). (Если считать, что $T(n, k) = 0$ при $n < k$, то формула станет верной при всех n и k.) Теперь по ней можно вычислить нужное нам число $T(10, 5)$:

n \ k	1	2	3	4	5
1	1				
2	1	2			
3	1	6	6		
4	1	14	36	24	
5	1	30	150	240	120
6	1	62	540	1560	1800
7	1	126	1806	8400	16800
8	1	•	5796	40824	126000
9	1	•	•	186480	834120
10	1	•	•	•	5103000

$T(n, k)$

(показана только часть таблицы, нужная для вычисления $T(10,5)$). $\triangleleft$

- Два решения этой задачи дают нам две формулы для $T(n,k)$:

$$\begin{aligned}T(n,k) &= C_k^0 k^n - C_k^1(k-1)^n + C_k^2(k-2)^n - \ldots \pm C_k^k(k-k)^n \\ &= k(T(n-1,k) + T(n-1,k-1))\end{aligned}$$

(для единообразия в первой формуле мы записали k^n как $C_k^0 k^n$ и оставили один нулевой член в конце). Вторая формула проще, но использует предыдущие значения S, в отличие от первой.

$\triangleright$ Для вычисления по рекуррентной формуле можно написать простую программу типа такой (это python):

```
N=10
K=5
T = [[0 for i in range(K+1)] for j in range(N+1)]
T[1][1]=1
for n in range(2,N+1):
    for k in range(1,K+1):
        T[n][k]=k*(T[n-1][k-1]+T[n-1][k])
print (T[10][5])
```

В этой программе мы окружаем интересующую нас таблицу нулями, чтобы не разбирать случаи $k = 1$ и $k = n$ отдельно. Благодаря этому единственное ненулевое значение, которое нам надо явно указать, это $T(1,1) = 1$. $\triangleleft$

- Посмотрим ещё раз на десятизначные числа из нечётных цифр, в которых встречаются все нечётные цифры. В таком числе можно переставить цифры по какому-то правилу. Скажем, если применить перестановку (1→5, 3→3, 5→7, 7→1, 9→9) к числу 5533997711, то получится другое число того же типа, а именно 7733991155. (Как сказали бы математики, «на нашем множестве действует группа перестановок пяти цифр».)

Сколько чисел можно получить из одного такими перестановками? («Сколько элементов в орбите одного числа при действии группы перестановок?») Разные перестановки дадут разные числа (поскольку все цифры входят, отличие проявится), поэтому получится $120 = 5!$ чисел в каждой группе. Такая группа соответствует разбиению всех разрядов числа (позиций в его записи) на пять классов (групп) — в каждом классе стоит одна и та же цифра. Можно сказать, что мы сначала разбиваем разряды на пять групп, а потом решаем, какую цифру назначить для каждой группы. Способов такого разбиения (на пять групп) будет $5\,103\,000/120 = 42\,525$ (из нашего рассуждения было сразу ясно, что поделится нацело).

Тот же вывод в общем случае: *число разбиений множества из n элементов на k различных классов равно $T(n,k)/k!$*. Если обозначить его через $S(n,k)$, то получается рекуррентная формула

$$S(n,k) = S(n-1,k-1) + kS(n-1,k).$$

Надо только отчётливо понимать, что́ именно мы подсчитываем, говоря о числе разбиений: разбиение определяется тем, какие позиции входят в один класс, а какие в разные (а, скажем, порядок перечисления классов не учитывается). Математики сказали бы, что мы подсчитываем «отношения эквивалентности на n-элементном множестве с k классами эквивалентности».

▷ Для чисел $S(n,k)$ есть традиционное название: *числа Стирлинга второго рода*. (Это тот же самый английский математик Стирлинг (1692–1770), который нашёл приближённую формулу для факториалов: $n! \approx \sqrt{2\pi n}(n/e)^n$.) ◁