

# Как считать слова?

Дмитрий Пионтковский, Максим Прасолов, Григорий Рыбников

## Решения

### 1 Главная задача

1. См. задачу 24.
2. См. задачи 18 и 48.
3. Один из вариантов решения дан в задаче 46, а другой может быть получен с помощью задач 59 и 61.
4. В задаче 1 алфавит  $A$  состоит из  $N = 100$  слов языка племени, роль слов языка  $L$  играют фразы языка племени Винни-Пухов, а запретные слова языка  $L$  — это два магических заклинания. В задаче 2 алфавит  $A$  состоит из  $N = 256$  команд компьютера, а роль слов языка  $L$  играют программы. Единственное запретное слово — программа, ломающая компьютер.

### 2 Как записывать ответ?

5. Слово из  $m$  букв получается выбором на каждом из  $m$  мест любой из  $N$  букв. Перемножая количество возможностей на каждом месте, получаем  $N^m$  слов.
6. Если в допустимом слове первая буква  $x$ , то вторая тоже. Аналогично остальные. Значит, допустимое слово имеет вид  $xx\dots x$ , где  $x$  — одна из  $N$  букв алфавита. Следовательно допустимых слов из фиксированного количества букв  $N$  штук.
7. Пусть в допустимом слове первая буква  $a$ . Так как  $aa$  — запрещённое слово, то вторая буква данного допустимого слова  $b$ . Продолжаем рассуждения и получаем, что на нечётном месте стоит буква  $a$ , а на чётном — буква  $b$ . Если первая буква  $b$ , то на чётном месте стоит буква  $a$ , а на нечётном —  $b$ . Значит, ряд размеров данного языка таков:  $1 + 2x + 2x^2 + 2x^3 + \dots$

### 3 Арифметика языков

8. Набор запретных слов таков: всевозможные слова из двух букв, в которых одна буква из первого алфавита, а другая — из второго, и запретные слова обоих языков. Очевидно, что допустимое слово каждого языка не содержит запретных слов в сумме языков. Если в не содержащем запретное подслово слове суммы языков первая буква принадлежит первому алфавиту, то вторая тоже, и аналогично остальные. Другими словами такое слово состоит из букв одного алфавита. Но тогда оно является допустимым в языке с этим алфавитом, а значит, допустимым в сумме языков.
9. Свободные члены рядов  $L(x)$  и  $L_1(x) + L_2(x) - 1$  равны единице. При  $n > 0$  коэффициент при  $x^n$  ряда  $L_1(x) + L_2(x) - 1$  равен сумме количеств слов длины  $n$  в языках  $L_1$  и  $L_2$ , то есть количеству слов длины  $n$  в языке  $L$ , то есть коэффициенту при  $x^n$  в ряде  $L(x)$ .

10. Имеем

$$(1-x)(1+x+x^2+x^3+\dots) = 1-x + (1-x) \cdot x + (1-x) \cdot x^2 + (1-x) \cdot x^3 + \dots = \\ = 1-x+x-x^2+x^2-x^3+x^3-x^4+\dots = 1,$$

что и требовалось.

**11.** Набор запретных слов таков: все слова из двух букв, в которых первая буква из второго алфавита, а вторая – из первого, и все запретные слова языков произведения. Рассмотрим допустимое слово произведения. В нём буквы второго алфавита следуют после букв первого алфавита, и это слово не содержит запретных слов языков-множителей. Значит, допустимое слово произведения не содержит указанных запретных слов. Теперь возьмём слово, не содержащее указанных запретных слов. В нём буквы второго алфавита следуют после букв первого алфавита, поэтому такое слово имеет вид  $w_1w_2$ , где  $w_1$  – слово из букв первого алфавита, а  $w_2$  – из букв второго. Слово  $w_1w_2$  не содержит запретных слов языков-множителей, поэтому слова  $w_1$  и  $w_2$  допустимы в своих языках, то есть слово  $w_1w_2$  допустимо в произведении языков.

**12.** Коэффициент при  $x^k$  в ряде  $L_1(x) \cdot L_2(x) = L_1(x) \cdot n_0 + L_1(x) \cdot n_1x + \dots = (n_0 \cdot m_0 + n_0m_1x + \dots) + (n_1m_0x + n_1m_1x^2 + \dots) + \dots$  равен  $n_0m_k + n_1m_{k-1} + \dots + n_km_0$ . Количество слов длины  $k$  в множестве слов  $L_1 \cdot L_2$  равно числу способов выбрать пару слов  $m$  из языка  $L_1$  и  $n$  из языка  $L_2$ , суммарное количество букв которых равно  $k$ . Если в слове  $m$   $i$  букв, то в слове  $n$   $k-i$  букв, а количество таких пар равно  $m_i \cdot n_{k-i}$ . Суммируя такие выражения для всех  $i$ , получаем  $n_0m_k + n_1m_{k-1} + \dots + n_km_0$ . Поэтому коэффициенты при  $x^k$  в рядах  $L_1(x) \cdot L_2(x)$  и  $L_1 \cdot L_2(x)$  совпадают, следовательно сами ряды совпадают.

**13.**

а) В решении этой задачи нам потребуются тот факт, что стандартные свойства умножения и сложения многочленов (ассоциативность, коммутативность, дистрибутивность) выполняются и для рядов. Рассмотрим, например, ассоциативность умножения –  $(P \cdot Q) \cdot R = P \cdot (Q \cdot R)$ . Чтобы вычислить коэффициент при  $x^k$  в левой и правой части, достаточно его вычислить для рядов, у которых отброшены члены более высокой степени, то есть для многочленов. Поэтому указанное тождество для рядов следует из того же тождества для многочленов. Аналогично доказываются остальные свойства.

Заметим теперь, что поскольку ряд  $\bar{S}$  начинается с первой степени  $x$ , в ряд  $R \cdot \bar{S}^m$  не входят степени  $x$  от нулевой до  $m-1$ -й. Именно поэтому бесконечные суммы вида  $R - R \cdot \bar{S} + R \cdot \bar{S}^2 - R \cdot \bar{S}^3 + \dots$  имеют смысл – для вычисления  $k$ -го коэффициента можно заменить эту сумму конечной. По той же причине для бесконечных сумм такого вида выполняется свойство дистрибутивности

$$R \cdot (S_1 + S_2 + S_3 + \dots) = R \cdot S_1 + R \cdot S_2 + R \cdot S_3 + \dots$$

Имея это в виду, мы легко получаем тождество

$$(1 + \bar{S})(1 - \bar{S} + \bar{S}^2 - \bar{S}^3 + \dots) = 1.$$

Отсюда

$$\begin{aligned} S \cdot \frac{R}{S} &= S \cdot (R - R \cdot \bar{S} + R \cdot \bar{S}^2 - R \cdot \bar{S}^3 + \dots) = S \cdot R \cdot (1 - \bar{S} + \bar{S}^2 - \bar{S}^3 + \dots) = \\ &= R \cdot (1 + \bar{S})(1 - \bar{S} + \bar{S}^2 - \bar{S}^3 + \dots) = R. \end{aligned}$$

б) Согласно утверждению а),

$$S \cdot (T - \frac{R}{S}) = S \cdot T - S \cdot \frac{R}{S} = R - R = 0$$

Предположим, что ряд  $(T - \frac{R}{S})$  ненулевой. Так как ряд  $S$  начинается с единицы, то первый ненулевой коэффициент ряда  $(T - \frac{R}{S})$  равен первому ненулевому коэффициенту ряда  $S \cdot (T - \frac{R}{S})$ . Значит, ряд  $(T - \frac{R}{S})$  нулевой и  $T = \frac{R}{S}$ .

**14.**

а) Рассуждая аналогично задаче 10, получаем

$$F_A(x) = 1 + Nx + N^2x^2 + \dots = \frac{1}{1 - Nx}.$$

б) Получаем

$$1 + Nx + Nx^2 + Nx^3 + \dots = -N + 1 + N \cdot (1 + x + x^2 + x^3 + \dots) = -N + 1 + \frac{N}{1-x} = \frac{1 + (N-1)x}{1-x}$$

и

$$1 + 2x + 2x^2 + 2x^3 + \dots = 1 + 2x \cdot (1 + x + x^2 + x^3 + \dots) = 1 + \frac{2x}{1-x} = \frac{1+x}{1-x}.$$

**15.** Следует из задач 59 и 61.

## 4 Свободное слово

16. Как следует из задачи 17 ниже, получаем  $L(x) = \frac{1}{1-26x+x^5}$ . Можно и непосредственно получить эту формулу, проведя рассуждения, аналогичные решению задачи 17.

17. Пусть  $a_k$  — число допустимых слов длины  $k$ . Ясно, что  $a_0 = 1$ . Докажем, что при  $k > 0$  выполнено рекуррентное соотношение  $a_k = Na_{k-1} - a_{k-m}$  (мы считаем, что  $a_k = 0$  при  $k < 0$ , так как слов отрицательной длины не существует).

Действительно, приписывая к началу каждого допустимого слова из  $k-1$  буквы каждую букву алфавита, мы получаем  $Na_{k-1}$  слов, среди которых все допустимые слова длины  $k$ . Посмотрим, какие недопустимые слова длины  $k$  получаются таким образом, то есть имеют вид  $cg$ , где  $c$  — буква, а  $g$  — допустимое слово длины  $k-1$ . Ясно, что запретное подслово должно стоять в начале, то есть  $cg = wf$ , где  $w$  — запретное слово, а  $f$  — допустимое. Из того, что  $w$  — свободное слово, следует, что для любого допустимого слова  $f$  слово, получающееся отбрасыванием первой буквы в слове  $wf$ , допустимо (в противном случае  $w$  имело бы зацепление с самим собой). Поэтому множество всех слов вида  $cg$ , где  $c$  — буква, а  $g$  — допустимое слово длины  $k-1$ , является объединением двух непересекающихся множеств: множества допустимых слов длины  $k$  и множества слов вида  $wf$ , где  $f$  — допустимое слово длины  $k-m$ . Отсюда следует нужное нам рекуррентное соотношение.

Рассмотрим сумму соотношения  $a_0 = 1$  и всех соотношений  $a_k x^k = Na_{k-1} x^k - a_{k-m} x^k$  для  $k = 1, 2, 3, \dots$ . Получится

$$L(x) = 1 + NxL(x) - x^m L(x).$$

Решая это уравнение относительно  $L(x)$ , получаем нужную нам формулу.

18. Согласно задаче 17,

$$L(x) = \frac{1}{1-26x+x^4} = 1 + (256x - x^4) + (256x - x^4)^2 + (256x - x^4)^3 + \dots$$

Коэффициент при  $x^7$  равен  $256^7 - 4 \cdot 256^3$ . Поэтому вероятность поломки компьютера равна  $\frac{4 \cdot 256^3}{256^7} = 4 \cdot 256^{-4}$ , что примерно равно  $10^{-10}$ .

## 5 Преобразования слов

19. Первая стрелка определена однозначно; вторая сопоставляет каждому слову из  $N$  то же самое слово, рассматриваемое как элемент  $F_A$ ; третья сопоставляет каждому из оставшихся слов  $F_A$  его же, как элемент  $G$ ; последняя стрелка тривиальна, как и первая.

20. Каждый из школьников съедает поровну конфет, бывших у мальчиков, и конфет, бывших у девочек. Последняя девочка доест все конфеты, бывшие у мальчиков. Поэтому и конфеты, бывшие у девочек, она тоже доест, и учительнице ничего не достанется.

21.

а) Пусть  $M_{\text{odd}} = M_1 \cup M_3 \cup M_5 \cup \dots$  и  $M_{\text{even}} = M_2 \cup M_4 \cup M_6 \cup \dots$ . Каждое преобразование устанавливает взаимно-однозначное соответствие между некоторым подмножеством  $M_{\text{odd}}$  и  $M_{\text{even}}$ , причём, поскольку по краям стоят пустые множества, каждый элемент участвует ровно в одном из этих взаимно-однозначных соответствий. Поэтому во множествах  $M_{\text{odd}}$  и  $M_{\text{even}}$  поровну элементов.

б) Для каждого  $k$  множество  $M_i^{(k)}$  слов из  $M_i$  длины  $k$  конечно; применяя к конечным множествам  $M_i^{(k)}$  с данным  $k$  утверждение пункта а), получаем, что коэффициенты при  $x^k$  в левой и правой частях доказываемой формулы совпадают. Поскольку  $k$  — любое, это означает, что формула верна.

22. Проверим, что множество  $A \cdot G$  является объединением двух непересекающихся множеств: множества  $\overline{G}$  и множества  $B \cdot G$ . Доказательство этого опирается на то, что множество  $B$  — свободное, и почти буквально повторяет соответствующее рассуждение в решении задачи 17.

Теперь точная последовательность строится очевидным образом: первая и последняя стрелки тривиальны, вторая переводит каждый элемент  $B \cdot G$  в себя (мы пользуемся тем, что  $B \cdot G \subseteq A \cdot G$ ), а третья переводит в себя каждый из оставшихся элементов  $A \cdot G$  (мы пользуемся тем, что  $A \cdot G \setminus B \cdot G = \overline{G}$ ). В частности, ядром второго преобразования является пустое множество, а ядром третьего преобразования (и образом второго) — множество  $D \cdot G$ ; образом третьего преобразования является  $\overline{G}$ .

**23.** По задаче 21б), из точной последовательности задачи 22 получаем

$$(B \cdot G)(x) + \overline{G}(x) = (A \cdot G)(x).$$

Заметим, что каждый элемент  $A \cdot G$  записывается в виде  $ag$ , где  $a \in A, g \in G$ , однозначно. Поэтому  $(A \cdot G)(x) = A(x)G(x)$ . Далее, каждый элемент  $B \cdot G$  записывается в виде  $bg$ , где  $b \in B, g \in G$ , однозначно (поскольку запрещённое слово не может быть подсловом другого запрещённого слова). Поэтому  $(B \cdot G)(x) = B(x)G(x)$ . Мы имеем  $A(x) = Nx, G(x) = L(x), \overline{G}(x) = G(x) - 1 = L(x) - 1$ . Отсюда

$$B(x)L(x) + L(x) - 1 = NxL(x).$$

Решая это уравнение относительно  $L(x)$ , получаем требуемую формулу.

**24.** Обозначим встречающиеся в фразах слова буквами: “земля” — А, “стоять” — В, “на” — С, “великий” — D, “крокодил” — Е, “каждый” — F, “вечер” — G, “глотать” — H, “солнце” — I. Тогда заклинаниям отвечают запретные слова “ABCDE” и “FGHИ”. Эти слова свободны (потому что в каждом все буквы различны) и не имеют зацеплений друг с другом (потому что первая и последняя буквы второго слова не встречаются в первом). Следовательно, множество заклинаний свободно, и ряд размеров языка имеет вид

$$L(x) = \frac{1}{1 - 100x + 2x^5}.$$

Из этой формулы легко вывести (обращая рассуждения в решении задачи 17), что числа  $a_k$  (количество фраз из  $k$  слов) можно вычислить из начального условия  $a_0 = 1$  и рекуррентного соотношения  $a_k = 100a_{k-1} - 2a_{k-5}$ . Вычисления приводят к ответу  $a_{20} = 10^{40} - 32 \cdot 10^{30} + 264 \cdot 10^{20} - 448 \cdot 10^{10} + 16$ .

**25.** В каждом запретном слове буква  $v$  встречается только на первом месте и все слова имеют длину 4, поэтому множество запретных слов свободно. По задаче 23, мы имеем

$$L(x) = \frac{1}{1 - 26x + 3x^4}.$$

**26.** Если множество запретных слов свободно, то, в частности, нет простых сцепок. Докажем обратное: если нет простых сцепок, то множество запретных слов свободно. Предположим противное: пусть множество запретных слов свободным не является. Тогда найдётся зацепление двух запретных слов, то есть найдутся такие три непустых слова  $s, t, r$ , что слова  $st$  и  $tr$  — запретные. Выберем такую тройку  $(s, t, r)$  так, чтобы она имела минимально возможную суммарную длину. Если это не простая сцепка, то  $str$  содержит запретное подслово  $w$ , отличное от  $st$  и  $tr$ . Заметим, что конец  $w$  не совпадает с концом  $tr$ , поскольку иначе либо  $w$  является подсловом  $tr$ , либо  $tr$  является подсловом  $w$ , что невозможно, так как никакое запретное слово не содержит другое запретное в качестве подслова. Аналогично, начало  $w$  не совпадает с началом  $st$ . Из тех же соображений, подслово  $w$  имеет общую часть как с подсловом  $s$ , так и с подсловом  $r$ . Обозначим через  $t'$  общую часть подслов  $st$  и  $w$ , через  $s'$  — остаток подслова  $st$ , а через  $r'$  — остаток подслова  $w$ . Эти слова непусты, их суммарная длина меньше суммарной длины слов  $s, t, r$ , а слова  $s't' = st$  и  $t'r' = w$  — запретные. Получено противоречие. Таким образом, из отсутствия простых сцепок следует, что множество запретных слов — свободно.

**27.** Будем строить преобразования последовательно, начиная с конца (с самой правой стрелки). Так как последнее множество пусто, то и область определения последнего преобразования — пустое множество. Поэтому предпоследнее преобразование имеет образом всё множество  $\overline{G}$ . Так как  $\overline{G} \subseteq A \cdot G$ , мы можем взять  $\overline{G}$  в качестве области определения предпоследнего преобразования и считать, что каждый элемент  $g \in \overline{G}$  переводится этим преобразованием в себя. Ядро этого преобразования состоит из не являющихся допустимыми слов вида  $ag$ , где  $a$  — буква и слово  $g$  является допустимым. Ясно, что для любого такого слова найдутся запретное слово  $w$  и допустимое слово  $f$  такие, что  $ag = wf$  (мы уже использовали это соображение при решении задач 17 и 22). Это позволяет построить третью с конца стрелку (это преобразование также переводит каждый элемент своей области определения в себя). Рассмотрим ядро этого преобразования. Оно состоит из тех слов вида  $wf$ , где  $w$  — запретное, а  $f$  — допустимое, которые имеют вид  $av$ , где  $a$  — буква, а слово  $v$  допустимым не является. Выберем в  $v$  самое первое (начинающееся левее всех) запретное подслово  $u$ . Легко видеть, что подслово  $u$  слова  $av = wf$  имеет общую часть с подсловом  $w$  и образует вместе с ним простую сцепку. Таким образом, ядро третьей с конца стрелки целиком содержится в  $S \cdot G$ , что позволяет построить преобразование  $S \cdot G \implies B \cdot G$  (также переводящее каждый элемент своей области определения в себя).

**28.** Из решения предыдущей задачи получаем, что язык является незапутанным тогда и только тогда, когда допустимы все слова вида  $rg$ , где  $r$  — хвост простой сцепки, а  $g$  — допустимое слово. Эквивалентное условие на множество запретных слов звучит так: не существует таких слов  $p, q, r, s, t$ , где слова  $p, q, s, t$  непусты, слова  $pq, qrs, st$  — запретные, причём  $pqrs$  — простая сцепка.

Заметим, что любой элемент множества  $S \cdot G$  однозначно представляется в виде произведения простой сцепки на допустимое слово (это легко следует из определения простой сцепки и того, что никакое запретное слово не является подсловом другого запретного слова). Аналогично решению задачи 23, для незапутанного языка  $L$  из точной последовательности мы получаем уравнение

$$S(x)L(x) + NxL(x) = B(x)L(x) + L(x) - 1,$$

откуда и получается искомая формула

$$L(x) = \frac{1}{1 - Nx + B(x) - S(x)}.$$

**29.** Простые сцепки имеют вид  $abbc, abbac$ , их хвосты —  $c, ac$ . Видно, что ни один из этих хвостов не оканчивается на непустое начало запретного слова. Поэтому язык является незапутанным. Соответственно, ряд размеров имеет вид

$$L(x) = \frac{1}{1 - 3x + 3x^3 - x^4 - x^5}.$$

**30.** Простые сцепки имеют вид  $x_i y_j z_k$ , где  $1 \leq i, j, k \leq n$ , их хвосты  $z_k$ . Так как ни одно запретное слово не начинается на  $z_k$ , язык является незапутанным. Соответственно, ряд размеров имеет вид

$$L(x) = \frac{1}{1 - 3nx + 2n^2x^2 - n^3x^3}.$$

**31.** Пусть  $w$  — единственное запретное слово, и пусть  $L$  — его длина. Предположим, что оно не является свободным, и  $pqr, pq = qr = w$ , — простая сцепка. Мы имеем  $wr = pqr = pw$ . Следовательно, конечный участок длины  $L$  у каждого из слов  $wr = pw, wrr = pwr = prw, wrrr = prwr = prrw, \dots$  равен  $w$ . Возьмём из этих слов первое, имеющее длину не меньше  $2L$ . Мы получим, что слово вида  $rrr \dots r$  имеет конечный участок, равный  $w$ . Но тогда у слова  $r$  есть непустое окончание, являющееся началом слова  $w$ . Следовательно наш язык не является незапутанным (см. решение задачи 28).

## 6 Ещё о свободных множествах

**32.** Например, если алфавит состоит из букв  $a$  и  $b$ , то свободным будет множество слов вида  $a^n b^n ab$ , где  $n \geq 2$ . Докажем это. Очевидно, что никакие два разных слова такого вида не являются подсловами друг друга. Осталось доказать, что между ними нет нетривиальных зацеплений (т.е. что любое зацепление есть просто зацепление слова с самим собой по всей длине). Действительно, если  $w$  — зацепление слов  $a^n b^n ab$  и  $a^m b^m ab$ , то легко видеть, что  $w$  содержит не менее трёх букв. Как конец слова  $a^n b^n ab$ , оно имеет вид либо  $b^k ab$ , либо  $a^k b^n ab$ , где  $1 \leq k \leq n$ . Поскольку оно также должно быть началом слова  $a^m b^m ab$ , получаем, что  $k = m = n$  и  $w = a^n b^n ab = a^m b^m ab$  — тривиальное зацепление.

**33. Лемма.** Если  $p(x) = 1 + p_1x + p_nx^n$  — многочлен с единичным свободным членом степени  $n \geq 1$ , то ряд  $f(x) = 1/p(x)$  не может быть многочленом (т.е. этот ряд содержит бесконечное множество ненулевых членов).

*Доказательство леммы.* Предположим, от противного, что ряд  $f(x)$  — многочлен, т.е.  $f(x) = f_0 + f_1x + \dots + f_mx^m$ , где старший коэффициент  $f_m$  ненулевой. Согласно задаче 13 а), получаем  $1 = f(x)p(x) = 1 + (f_0p_1 + f_1p_0)x + \dots + f_m p_n x^{m+n}$  — противоречие.

Перейдём к решению задачи. Согласно задаче 23, имеем

$$L(x) = \frac{1}{1 - Nx + B(x)}.$$

Если бы язык  $L$  был конечным, то ряд  $L(x)$  был бы многочленом, что невозможно согласно доказанной лемме. Следовательно, множество допустимых слов языка бесконечно.

**34.** Очевидно, для любых рядов неравенство  $A \geq B$  равносильно неравенству  $A - B \geq 0$ , т.е. условию, что коэффициенты ряда  $A - B$  неотрицательны. Обозначим ряд  $P - Q$  через  $A = a_0 + a_1x + a_2x^2 \dots$ , а ряд  $R$  как  $R = r_0 + r_1x + r_2x^2 \dots$ . Тогда произвольный  $n$ -й коэффициент ряда  $AR$ , вычисляемый по формуле  $a_0r_n + a_1r_{n-1} + \dots + a_nr_0$ , представим в виде суммы неотрицательных чисел, и потому сам неотрицательный. Это означает, что выполняется неравенство  $AR \geq 0$ , равносильное неравенству  $PR - QR \geq 0$ , или  $PR \geq QR$ .

35. По формуле из задачи 23,

$$L(x) = \frac{1}{1 - A(x) + B(x)}.$$

Согласно задаче 27, существует точная последовательность

$$\emptyset \implies K \implies B' \cdot G' \implies A \cdot G' \implies \overline{G'} \implies \emptyset,$$

где  $K$  — ядро отображения  $B \cdot G \implies A \cdot G'$ , а  $G'$  — множество допустимых слов языка  $L'$ . Из этой последовательности получаем (зад. 21) равенство

$$B'(x)G'(x) - A(x)G'(x) + G'(x) - 1 = K(x),$$

откуда (т.к.  $B'(x) = B(x)$ ,  $L'(x) = G'(x)$  и  $K(x) \geq 0$ )

$$L'(x)(B(x) - A(x) + 1) \geq 1.$$

Умножая это неравенство на ряд  $L(x) \geq 0$ , получаем (ввиду зад. 33)

$$L'(x)(B(x) - A(x) + 1) \cdot \frac{1}{1 - A(x) + B(x)} \geq L(x),$$

т. е.

$$L'(x) \geq L(x).$$

36. Пусть алфавит  $A = \{a, b\}$ . Обозначим второе слово через  $v$ . Очевидно, чтобы слова  $v$  и  $w$  не были свободными, необходимо, чтобы их начальные и конечные буквы были различные. Если при этом  $w$  начинается на одну букву, а  $v$  — на другую, то последняя буква слова  $v$  совпадает с первой буквой слова  $w$ , и эти слова зацепляются, так что множество  $B$  не будет свободным. Итак, осталось рассмотреть случай, когда оба слова  $v$  и  $w$  начинаются на одну букву (скажем,  $a$ ), а заканчиваются на другую ( $b$ ).

а) Имеем  $w = ab$  и  $v = a...b$ . Очевидно, если первая буква  $b$  в слове  $v$  находится на  $k$ -м месте, то его подслово из букв на  $(k-1)$ -м и  $k$ -м местах равно  $w$ , так что множество  $B$  не свободно.

б) Ответ: нет. Пусть  $w = aab$  (случай  $w = abb$  аналогичен, с точностью до левой-правой симметрии и замены букв). Поскольку слово  $w$  не является подсловом в  $v$ , то в слове  $v$  за парой букв  $aa$  всегда следует ещё одна  $a$ . Так как слова  $v$  и  $w$  не зацепляются, то  $v$  не может начинаться на  $ab$ , т. е. оно начинается на  $aa$ . Следовательно, третья буква в  $v$  снова  $a$ , затем четвёртая и т. д., откуда  $v = aa...a$  — противоречие.

37. Достаточно показать, что существует свободное множество  $B$  из  $g = \lfloor n^2/4 \rfloor$  двухбуквенных слов. Пусть  $k = \lfloor n/2 \rfloor$ , т. е.  $n = 2k$  или  $n = 2k + 1$ . Положим  $B = \{x_i x_j | 1 \leq i \leq k, k+1 \leq j \leq n\}$ . Очевидно, оно свободно. Тогда в случае четного  $n = 2k$  множество  $B$  насчитывает  $k^2 = n^2/4$  элементов, а в случае нечётного  $n = 2k+1$  оно содержит  $k(k+1) = (n-1)(n+1)/4 = n^2/4 - 1/4 = \lfloor n^2/4 \rfloor$  элементов, что и требовалось.

38. Достаточно показать, что существует свободное множество  $B$  из  $m \leq k^d(d-1)^{d-1}$  слов длины  $d$ . Разобьём основной алфавит  $A = \{x_1, \dots, x_n\}$  на два подмножества  $P = \{x_1, \dots, x_k\}$  и  $Q = \{x_{k+1}, \dots, x_n\}$ . Тогда легко видеть, что множество  $B = \{pq | p \in P, q \in Q\}$  — слово длины  $d-1$  — искомого.

39.

а) Согласно задаче 23,  $\frac{1}{1-nx+B(x)} = L(x) \geq 1$ .

б) Ответ: неверно. Например, в случае двухбуквенного алфавита  $A$  такого множества  $B$  не существует при  $p(x) = x^3 + x^{10}$  (это доказано в задаче 36 б). Осталось убедиться, что ряд

$$f(x) = \frac{1}{1 - 2x + x^3 + x^{10}}$$

имеет неотрицательные коэффициенты (поскольку свободный член этого ряда единичный, это условие равносильно требуемому неравенству  $f(x) \geq 1$ ). Оказывается, для этих коэффициентов справедливо даже более сильное неравенство  $a_n \geq (3/2)^n$ . Доказательство последнего неравенства можно провести по индукции, воспользовавшись рекуррентным соотношением  $a_{n+10} = 2a_{n+9} - a_{n+7} - a_n$ , которое следует из равенства  $(1 - 2x + x^3 + x^{10})f(x) = 1$ .

Другой подобный пример — случай  $p(x) = 4x^6$ , снова над алфавитом из двух букв.

40. Решение содержится в теореме 5.1 (эквивалентность  $A \iff B$ ) и предложении 5.6 в статье: David Anick, *Generic algebras and CW-complexes*, Proceedings of 1983 Conference on algebra, topology and K-theory in honor of John Moore. Princeton University, 1988, p. 247–331.

41. Вопрос остаётся открытым.

## 7 Слова и цепи

42. Слово “of” не участвует в построении цепей длины больше 1, потому что данные слова не начинаются на букву f и не заканчиваются на букву o. Буквы s нет в слове tournament, поэтому цепь может содержать слово towns только в конце. Буква t в слове tournament стоит лишь на первом и последних местах, поэтому это слово может зацепляться с собой только по одной букве. Исходя из сказанного, мы находим все цепи: tournament, of, towns; tournamenttournament, tournamentowns;... Цепей длины больше 1 будет ровно две.

43. Рассмотрим цепь  $c$  длины  $n$ . От его правого конца можно строить антицепь по направлению к левому концу единственным образом. Будем присоединять звенья слева по очереди. Докажем по индукции, что начало  $i$ -го звена антицепи лежит между началом  $i$ -го справа звена цепи и концом  $i+1$ -го справа звена цепи. Для  $i = 1$  утверждение верно, потому что в этом случае соответствующие звенья цепи и антицепи совпадают. Проверим базу также для  $i = 2$ . Второе звено антицепи лежит правее второго справа звена цепи, потому что между первым и вторым звеном нет запретных слов. Начало второго звена антицепи лежит левее конца третьего справа звена цепи, потому что правее третьего справа звена цепи нет запретных слов по определению цепи. Таким образом для  $i = 2$  утверждение верно. Докажем шаг. По предположению индукции  $i$ -е справа звено цепи пересекает  $i-1$ -е звено антицепи, но  $(i+1)$ -е звено антицепи не пересекает  $(i-1)$ -го звена антицепи, поэтому  $(i+1)$ -е звено антицепи левее  $i$ -го справа звена цепи. Между концом  $(i+2)$ -го и началом  $i$ -го справа звена цепи не начинается никакое запретное слово, поэтому начало  $(i+1)$ -го звена антицепи левее конца  $(i+2)$ -го звена цепи. Так как  $(i-1)$ -е звено антицепи не левее  $i-1$ -го звена цепи, то  $(i+1)$ -е звено цепи не пересекает  $(i-1)$ -е звено антицепи, но пересекает  $i$ -е звено, поэтому  $(i+1)$ -е звено антицепи не левее  $(i+1)$ -го звена цепи.

44. Пусть цепь  $c'$  – подслово цепи  $c$ . Докажем по индукции, что  $i$ -е звено цепи  $c$  не правее  $i$ -го звена цепи  $c'$ . База  $i = 1$  очевидна. Если первые звенья цепей совпадают, и одна является подсловом другой, то все остальные звенья также совпадают, и, в частности, если такие цепи имеют одинаковую длину, то они совпадают. Поэтому цепь  $c'$  начинается не с начала, а поэтому первое звено цепи  $c'$  правее или совпадает со вторым звеном цепи  $c$ . Если бы второе звено цепи  $c$  было правее первого звена цепи  $c'$ , то тогда бы первое звено цепи  $c'$  можно было бы выбрать в качестве второго звена цепи  $c$ . Значит, второе звено цепи  $c$  левее второго звена цепи  $c'$ , и база индукции верна для  $i = 2$ . Перейдём к шагу индукции. Если  $(i+1)$ -е звено цепи  $c'$  не пересекается с  $i$ -ым звеном цепи  $c$ , то  $(i+1)$ -е звено цепи  $c'$  правее  $(i+1)$ -го звена цепи  $c$ . Также добавленное звено цепи  $c'$  не пересекается с  $(i-1)$ -м звеном цепи  $c'$ , а значит, по предположению индукции с  $(i-1)$ -м звеном цепи  $c$ . Поэтому, если  $i+1$ -е звено цепи  $c'$  пересекает  $i$ -е звено цепи  $c$ , то  $i+1$ -е звено цепи  $c$  всё равно не правее  $(i+1)$ -го звена цепи  $c'$ , потому что иначе то звено можно было бы выбрать вместо этого. Шаг доказан. Значит, последние звенья цепей совпадают, но по прошлой задаче данная пара цепей является также парой антицепей одинаковой длины с совпадающими первыми звеньями, откуда они совпадают.

45. Пусть слово представлено в виде  $gc$ , где  $g$  – допустимое слово, а  $c$  – цепь длины не менее двух. По задаче 43 цепь  $c$  также является антицепью. Если при убиении первого звена цепи  $c$  появляется запретное слово, не зацепленное с укороченной цепью  $c$ , то антицепь  $c$  можно продолжить влево и получить новое представление  $g'c'$ , где длина цепи увеличилась. Если такого запретного слова не появилось, то антицепь  $c$  можно укоротить на самое левое звено и получить новое представление  $g'c'$ , в котором длина цепи уменьшилась. Если есть ещё одно представление  $g''c''$ , то заметим, что дуги антицепей  $c, c', c''$  совпадают, а значит, длина каких-то двух цепей отличается минимум на два, но тогда допустимое слово перед короткой антицепью содержит самое левое звено более длинной антицепи, противоречие.

46. Для решения этой задачи мы применим задачу 21 для точной последовательности

$$\dots \implies C_{n+1} \cdot G \implies C_n \cdot G \implies C_{n-1} \cdot G \implies \dots C_1 \cdot G \implies A \cdot G \implies \bar{G}$$

Построим эту последовательность. Возьмём слово  $cg$  из  $C_n \cdot G$ . Хвост цепи  $c$  припишем в начало слова  $g$  и получим  $c'g'$ . Если  $g'$  допустимое слово, то слово  $c'g'$  принадлежит  $C_{n-1} \cdot G$ . Если слово  $g'$  содержит запретное слово, то цепь  $c$  можно продолжить вправо до цепи  $c''$ , остаток обозначим  $g''$ . Тогда слово  $c''g''$  принадлежит  $C_{n+1} \cdot G$ . Цепь можно единственным образом продолжать внутри слова, поэтому построенные отображения взаимно-однозначны на частях языков  $C_n \cdot G$ .

Стрелки  $C_1 \cdot G \implies A \cdot G \implies \bar{G} \implies \emptyset$  мы берём из задачи 27. По задаче 21

$$1 - L(x)(1 - Nx + C_1(x) - C_2(x) + C_3(x) - \dots) = 0.$$

Откуда получаем требуемое.

47. По задаче 42  $C_1(x) = x^2 + x^5 + x^{10}$ ,  $C_n(x) = x^{9(n-1)}(x^5 + x^{10})$ . По формуле из задачи 46

$$L(x) = \frac{1}{1 - 26x + x^2 + (x^5 + x^{10})(1 - x^9 + x^{18} - \dots)} = \frac{1}{1 - 26x + x^2 + \frac{x^5 + x^{10}}{1 + x^9}} = \frac{1 + x^9}{1 - 26x + x^2 + x^5 + x^9 - 25x^{10} + x^{11}}$$

48. Рассмотрим четыре варианта запретного слова из четырёх букв с точностью до замены букв.

1) Запретное слово имеет вид аaaa. В этом случае  $C_{2n} = x^{4n+1}$ ,  $C_{2n-1} = x^{4n}$ . По формуле задачи 46

$$L(x) = \frac{1}{1 - 256x + x^4 - x^5 + \dots} = \frac{1}{1 - 256x + \frac{x^4 - x^5}{1 - x^4}} = \frac{1 - x^4}{1 - 256x + 255x^5} = (1 - x^4)(1 + (256x - 255x^5) + (256x - 255x^5)^2 + \dots).$$

Коэффициент при  $x^7$  равен  $256^7 - 3 \cdot 256^2 \cdot 255 - 256^3 = 256^7 - 4 \cdot 256^3 + 3 \cdot 256^2$ .

2) Запретное слово имеет вид abca и хотя бы две различные буквы. В этом случае  $C_n = x^{3n+1}$ .

$$L(x) = \frac{1}{1 - 256x + x^4 - x^7 + \dots} = 1 + (256x - x^4 + x^7 - \dots) + (256x - x^4 + x^7 - \dots)^2 + \dots$$

Коэффициент при  $x^7$  равен  $256^7 - 4 \cdot 256^3 + 1$

3) Запретное слово имеет вид abab. В этом случае  $C_n = x^{2(n+1)}$ .

$$L(x) = \frac{1}{1 - 256x + x^4 - x^6 + x^8(\dots)}$$

Коэффициент при  $x^7$  равен  $256^7 - 4 \cdot 256^3 + 2 \cdot 256$

4) Запретное слово свободно. А случай свободного запретного слова разобран в задаче 18.

49. Из задачи 46 следует, что

$$\frac{1}{1 - Nx} - L(x) = L(x) \cdot C_1(x) \cdot \frac{1}{1 - Nx} - L(x) \cdot C_2(x) \cdot \frac{1}{1 - Nx} + L(x) \cdot C_3(x) \cdot \frac{1}{1 - Nx} - \dots$$

(Бесконечная сумма имеет смысл, поскольку степени начальных членов слагаемых растут.) В левой части равенства стоит ряд размеров множества недопустимых слов. В правой части равенства стоит знакопеременная сумма рядов размеров множеств  $L \cdot C_n \cdot F_A$ . Определено отображение для каждого из этих языков в множество недопустимых слов, и наоборот, каждое недопустимое слово можно представить в виде  $gc_1u$ , где  $g$  – допустимое,  $c_1$  – запретное слово. Но недопустимое слово может быть представлено в виде  $gc_nu$ , где  $g$  – допустимое, а  $c_n$  – цепь длины  $n$ , несколькими способами. Пусть для слова  $w$  число таких способов равно  $w_n$ . Тогда сумма чисел  $(w_1 - w_2 + w_3 - w_4 + \dots)$  по всем словам длины  $k$  равна коэффициенту при  $x^k$  в правой части равенства выше, а значит, и в левой, что равно количеству недопустимых слов длины  $k$ .

Заметим, что из представления слова  $w$  в виде  $gc_nu$  можно получить другое представление с длиной цепи на единицу меньше, отбросив хвост цепи  $c_n$  в остаток  $u$ . Поэтому два таких варианта сократятся в сумме  $(w_1 - w_2 + w_3 - w_4 + \dots)$ . Другими словами рассматриваемая сумма равна числу представлений слова  $w$  в виде  $gsu$ , где подслово  $s$  – это максимальная подцепь слова  $w$ , причём цепь  $s$  имеет нечётную длину.

Рассмотрим представление слова  $w$  в виде  $gsu$ , где  $s$  – максимальная цепь с самым правым последним звеном. По задаче 45 либо в слове  $gs$  есть максимальная цепь длины 1, либо оно представимо в виде  $g'c'$ , где  $g'$  – допустимо, а  $c'$  – цепь, которая либо длиннее, либо короче цепи  $s$  на одно звено. Заметим, что  $c'$  – также максимальная подцепь слова  $w$ . Одна из цепей  $s$  и  $c'$  имеет нечётную длину. Тем самым мы показали, что сумма  $(w_1 - w_2 + w_3 - w_4 + \dots)$  не меньше единицы. Но сумма таких величин по всем недопустимым словам длины  $k$  равна их количеству. Значит, каждая величина равна единице, и в каждом недопустимом слове максимальных цепей нечётной длины ровно одна.

50. Применим формулу задачи 46.

$$L'(x) = \frac{1}{1 - (N+1)x + C_1(x) - C_2(x) + \dots} = \frac{1}{\frac{1}{L(x)} - x}$$

51. Применим формулу задачи 46.

$$W(x) = \frac{1}{1 - (N+N')x + (C_1(x) + C_1'(x)) - (C_2(x) + C_2'(x)) + \dots} = \frac{1}{\frac{1}{L(x)} + \frac{1}{L'(x)} - 1}$$

52. Допустимые слова языка  $M$  – это цепи языка  $L$ . Цепи длины  $n$  этих языков состоят из  $n+1$ -й буквы, поэтому  $C_n(-x) = (-1)^{n+1}C_n(x)$ . Имеем  $M(-x) = 1 - Nx + C_1(x) - C_2(x) + \dots$ , то есть  $L(x)M(-x) = 1$ .

## 8 Дополнительные задачи

**53.** Существование свободного множества при условии  $m \leq k^d(d-1)^{d-1}$  доказано в задачах 37 и 38. Осталось доказать, что при  $m > k^d(d-1)^{d-1}$  искомого свободного множества не существует.

а) Если задано  $m > n^2/4$  слов из двух букв, то они могут составлять свободное множество  $S$  лишь при условии, что первая буква никакого слова не совпадает со второй буквой никакого другого слова, т.е. есть два непересекающихся множества букв,  $P$  и  $Q$ , элементы которых могут служить, соответственно, только начальными или только вторыми буквами слова из  $S$ . Обозначим  $r = |P| + |Q| \leq n$  и  $s = |P| \cdot |Q| \geq |S| = m > n^2/4$ . По теореме Виета, натуральные числа  $|P|$  и  $|Q|$  являются корнями квадратного уравнения  $x^2 - rx + s = 0$ , дискриминант которого  $D = r^2 - 4s$  отрицателен при выполнении указанных ограничений на  $r$  и  $s$  — противоречие.

б) Пусть  $B$  — свободное множество  $m$  слов длины 3. Поскольку в свободном множестве никакая первая буква слова не может быть также и последней буквой какого-либо слова, в алфавите  $A$  есть два непересекающихся подмножества  $X$  и  $Y$ , элементы которых встречаются, соответственно, только в начале и только в конце слов из  $B$ . Если есть буквы, которые не встречаются ни в конце, ни в начале слова, присоединим их произвольным образом к одному из множеств  $X$  и  $Y$ . Пусть, для определённости, число  $s$  элементов множества  $X = \{x_1, \dots, x_s\}$  не превосходит количество элементов  $t$  множества  $Y = \{y_1, \dots, y_t\}$ . Каждый элемент множества  $B$  имеет вид  $x_{i_1}x_{i_2}y_{j_1}$  или  $x_{i_1}y_{j_1}y_{j_2}$ , причём никакое финальное подслово вида  $xy$  элемента первого типа не может быть началом элемента второго типа. Проведём преобразование множества  $B$ , заменив для каждого финального подслова  $xy$  все слова первого типа, оканчивающиеся на него, на слова второго типа, по правилу  $x_i xy \mapsto xy y_i$ , где  $1 \leq i \leq s \leq t$ . Легко видеть, что при таком преобразовании никакие элементы не переходят в другие элементы множества  $B$  и никакие различные элементы не переходят в один и тот же, причём получившееся множество  $B'$  останется свободным. При этом в  $B'$  все элементы имеют вид  $x_{i_1}y_{j_1}y_{j_2}$ . Следовательно, количество элементов множества  $B'$  (равное по-прежнему числу  $m$ ) не превосходит  $st^2$ , откуда

$$m \leq st^2 \leq (n-t)t^2 \leq \left(n - \frac{2n}{3}\right) \left(\frac{2n}{3}\right)^2 = \frac{4n^3}{27} = 4k^3,$$

что и требовалось.

в) Для доказательства потребуется следующая аналитическая

**Лемма.** Пусть  $R(x) = 1 + a_1x + a_2x^2 + \dots$  — ряд с натуральными коэффициентами такой, что  $R(x) = 1/p(x)$  для некоторого многочлена  $p(x)$  с единичным свободным членом. Обозначим  $R_n(x) = 1 + a_1x + a_2x^2 + \dots + a_nx^n$ . Если для всех  $x \in [0, x_0]$ , где  $x_0 > 0$ , выполняется условие  $p(x) \geq m > 0$ , то для всех  $n > 0$  справедливы неравенства  $R_n(x_0) \leq 1/m$ .

Не доказывая лемму, перейдём к решению задачи. Обозначим  $s = mk^{-d} - (d-1)^{(d-1)}$ ; требуется доказать, что при  $s > 0$  свободного множества не существует. Предположим, что это не так. Тогда, согласно задаче 39 а), ряд  $1/p(x)$ , где  $p(x) = 1 - d k x + m x^d$ , имеет натуральные коэффициенты (нулевых коэффициентов быть не может, поскольку этот ряд бесконечен согласно задаче 33). Отметим, что при  $s > 0$  многочлен  $p(x)$  положителен на отрезке  $[0, 1]$  (доказательство: минимум этого многочлена на отрезке  $[0, 1]$  достигается либо в концах отрезка, где  $p(x)$  положителен, либо в такой точке  $x_0$ , в которой  $p'(x_0) = 0$ , т.е. при  $x_0 = \frac{1}{k(d-1)}$ ; при этом  $p(x_0) = s x_0^d > 0$ ). Это означает, что существует такое число  $m > 0$ , что  $p(x) \geq m$  при  $x \in [0, 1]$ . Согласно лемме, это означает, что для всех  $n$  количество слов длины не выше  $n$ , равное  $L_n(1)$ , ограничено константой  $1/m$ .

**54.** а) По определению, запретными словами языка  $L^!$  являются все двухбуквенные слова, не запретные в  $L$ . Следовательно, запретными словами языка  $(L^!)^!$  будут все двухбуквенные слова, не запретные в  $L^!$ , т.е. в точности запретные слова языка  $L$ . Таким образом, и алфавиты, и наборы запретных слов языков  $L$  и  $(L^!)^!$  совпадают, а потому и сами языки равны.

б) Поскольку множество запретных слов языка  $M = (L_1 + L_2)^!$  есть объединение множеств запретных слов языков  $L_1^!$  и  $L_2^!$ , а алфавит языка  $M$  представляет собой объединение их (непересекающихся) алфавитов, то язык  $M$  есть свободное произведение (см. определение в зад. 51) языков  $L_1^!$  и  $L_2^!$ .

в) Запретными словами языка  $(L_1 \cdot L_2)^!$  будут разрешённые двухбуквенные слова языков  $L_1$  и  $L_2$ , а также слова вида  $aB$ , где  $a$  — буква из алфавита языка  $L_1$ , а  $B$  — буква из алфавита языка  $L_2$ . Это означает, что

$$(L_1 \cdot L_2)^! = L_2^! \cdot L_1^!.$$

**55.** Пусть  $w$  — слово длины  $nk$  (где  $k \geq 1$ ) над алфавитом языка  $L$ ,  $w^{(n)}$  — соответствующее ему слово языка  $L^{(n)}$ . Разобьём  $w$  на подслова  $w = w_1 \dots w_k$ , каждое из которых соответствует букве языка  $L^{(n)}$ . Легко видеть, что слово  $w$  содержит какое-то запретное подслово  $u$  (состоящее, по определению, из не более чем  $d$  букв) в том и только том случае, когда в каком-то подслове  $w' = w_p \dots w_{p+m-1}$  каждое из подслов  $w_i$

либо содержится в подслове  $u$ , либо пересекается с ним, так что количество  $n$ -буквенных фрагментов  $m$  в подслове  $w'$  удовлетворяет неравенству  $m \leq s$ , где

$$s = 2 + \left\lfloor \frac{d-2}{n} \right\rfloor.$$

Таким образом, любое недопустимое слово  $w^{(n)}$  языка  $L^{(n)}$  содержит недопустимое подслово из не более чем  $s$  букв, т. е. язык  $L^{(n)}$  задаётся конечным множеством запретных слов, причём длины запретных слов этого языка не превосходят  $s$ . Это доказывает утверждение а).

б) Ответ: не всегда.

Докажем, что при  $d \geq 3$  и  $n \geq 2$  длины запретных слов языка  $L^{(n)}$  всегда меньше  $d$ ; в частности, этот язык не может быть  $d$ -определённым, т. е. ответ на вопрос из п. б) отрицательный. Достаточно доказать неравенство  $s < d$ , или

$$2 + \frac{d-2}{n} < d.$$

Последнее неравенство равносильно неравенству  $(d-2)(1-1/n) > 0$ , которое, очевидно, верно при заданных ограничениях на  $d$  и  $n$ .

в) Ответ:  $n = d - 1$ . По доказанному выше, язык  $L^{(n)}$  является квадратичным или свободным (т. е. длины запретных слов не превосходят 2) при условии  $s \leq 2$ , которое равносильно неравенству  $2 + \frac{d-2}{n} < 3$ , или  $n > d - 2$ , т. е.  $n \geq d - 1$ . Если же  $n \leq d - 2$ , то существуют такие  $d$ -определённые языки  $L$ , для которых язык  $L^{(n)}$  имеет запретные слова из более чем трёх букв: примером служит язык  $L$  с трёхбуквенным алфавитом  $\{a, b, c\}$  и единственным запретным словом  $abc^{d-2}$ .

56. См. решение задачи 58.

57. Ответ: да. Например, пусть  $A$  — алфавит из  $n \geq 2$  букв. Рассмотрим язык  $L = F_A \cdot F_A^!$ . Поскольку язык  $F_A$  имеет экспоненциальный рост (для него в задаче 55с) можно выбрать  $c_1 = n + 1$  и  $c_2 = n$ ), причём  $2F_A(x) \geq L(x) \geq F_A(x)$ , то язык  $L$  также имеет экспоненциальный рост. Согласно задаче 53в), имеем  $L^! = (F_A^!)^! \cdot F_A^! = L$ , так что оба языка  $L$  и  $L^!$  имеют экспоненциальный рост.

58. Докажем сначала следующее утверждение (при решении задачи 56 без него можно обойтись).

**Лемма.** Пусть  $a = \{a_0, a_1, a_2, \dots\}$  — последовательность натуральных чисел, в которой  $a_0 = 1$  и для некоторого натурального  $N$  имеем  $a_1 \geq 2, \dots, a_N \geq 2$ . Тогда последовательность  $a$  имеет полиномиальный (соответственно, экспоненциальный) рост в том и только том случае, когда соответствующие неравенства из утверждений б) и с) выполняются для всех  $a_k$  при  $k \geq N$ .

*Доказательство леммы.* Пусть  $M = \max_{i \leq N} \{a_i\}$ . Очевидно, если для некоторых многочленов  $p, q$  степени  $d$  выполняются неравенства  $p(k) \geq a_k \geq q(k)$  при  $k \geq N$ , то выполняются также и неравенства  $p(k) + M \geq a_k \geq q(k) - M$  при всех  $k$ , что доказывает лемму в случае полиномиального роста. Аналогично, если  $c_1^k \geq a_k \geq c_2^k$  при  $k \geq N$ , то  $(M + c_1)^k \geq a_k \geq g^k$  при всех  $k$ , что полностью доказывает лемму.

Перейдём к решению задачи. Очевидно, все допустимые слова длины  $\geq d - 1$  получаются, если, начиная со слова в вершине графа, приписывать к нему справа буквы, которые прочитываются при прохождении какого-либо маршрута, начинающегося в этой вершине, причём разные слова соответствуют разным маршрутам. Очевидно, язык конечен тогда и только тогда, когда никакой маршрут не возвращается в начальную вершину, т. е. в графе нет циклов (что доказывает утверждение а)). Осталось рассмотреть случай, когда язык бесконечен и в графе есть цикл. В этом случае количество  $a_j$  слов длины  $j \geq d$  равно числу маршрутов длины  $j - d + 1$ .

Предположим, что есть два пересекающихся цикла; пусть их длины суть  $d_1$  и  $d_2$ , и  $v$  — общая вершина, исходящие из которой ребра различны для обоих циклов (скажем, отвечающие буквам  $x$  и  $y$ ). Слова, которые прочитываются на ребрах  $k$ -звенных маршрутов, начинающихся в  $v$  и проходящим по каждому из циклов, различны, поэтому  $a_k \geq 2$  при всех  $k \geq 0$ . Кроме того, для любого  $j = (d - 1) + q(d_1 + d_2) + r$ , где  $r < d_1 + d_2$  — остаток от деления числа  $j - d + 1$  на  $d_1 + d_2$ , существует по крайней мере  $2^q$  различных маршрутов длины  $j - d + 1$  (на каждом из  $q$  шагов проходим оба цикла в произвольном порядке, а затем делаем  $r$  шагов в произвольном цикле), так что при  $j \geq 2d$  имеем  $a_j \geq 2^q = 2^{\lfloor \frac{j-d+1}{d_1+d_2} \rfloor} \geq c^j$ , где  $c = 2^{1/2(d_1+d_2)}$ . Поскольку всегда  $a_j \leq n^j$ , из доказанной леммы (при  $N = 2g$ ) следует, что рост экспоненциальный.

Осталось разобрать случай, когда в графе  $\Gamma_L$  циклы есть, но они не пересекаются между собой. Достаточно проверить условия полиномиальности для количеств маршрутов  $b_k = a_{k+d-1}$  длины  $k$  в графе  $\Gamma_L$  (т. к. если соответствующие неравенства выполняются для чисел  $b_k$ , то они выполняются и для  $a_k$  при  $k \geq d - 1$  при замене многочленов  $p(x)$  и  $q(x)$  на многочлены той же степени  $p_1(x) = p(x + d - 1)$  и  $q_1(x) = q(x + d - 1)$ ). Мы докажем, что каждый член последовательности  $b_k$  равен значению некоторого многочлена  $b(k)$  с положительным старшим коэффициентом (будем называть такие последовательности полиномиальными).

Рассмотрим другой граф  $\Gamma'_L$ , вершинами которого служат циклы графа  $\Gamma_L$  и вершины графа  $\Gamma_L$ , не входящие в циклы (назовём их обособленными), а ребра соответствуют ребрам, соединяющим соответствующие

компоненты (вершины или циклы) графа  $\Gamma_L$ . Очевидно, в графе  $\Gamma'_L$  циклов нет, т.е. множество маршрутов в нём конечно. Пусть  $Q^v$  — множество маршрутов в графе  $\Gamma'_L$ , исходящих из данной вершины  $v$ , и пусть  $q_k^v$  — количество маршрутов длины  $k$ , которое им соответствует в графе  $\Gamma_L$ . Поскольку  $b_k = \sum_v q_k^v$ , достаточно доказать, что последовательность  $\{q_k^v\}$  для каждой вершины  $v$  полиномиальна. Воспользуемся индукцией по длине  $D = D(v)$  максимального маршрута, исходящего из  $v$ . Если  $D = 0$ , то либо  $q_k^v = 0$  при  $k > 0$  (если  $v$  — обособленная вершина), либо  $q_k^v = 1$  для всех  $k$  (если  $v$  — цикл), т.е. соответствующая последовательность всегда полиномиальна. Пусть теперь  $v$  — какая-то начальная вершина графа  $\Gamma'_L$ , из которой исходят  $r$  стрелок  $a_1, \dots, a_r$  к вершинам  $v_1, \dots, v_r$  (возможно, повторяющимся). По индукции, считаем  $q_k^{v_i} = b_i(k)$  — многочлен с положительным старшим коэффициентом. Если  $v$  — обособленная вершина, то  $q_k^v = \sum_{i=1}^r q_k^{v_i}$ , т.е. эта последовательность полиномиальна как сумма полиномиальных последовательностей. Если же  $v$  — цикл, то перед переходом к вершинам по ребрам  $a_1, \dots, a_r$  возможно слово любой длины в цикле, поэтому  $q_k^v = \sum_{i=1}^r \left( \sum_{j=1}^k q_{k-j}^{v_i} \right) = \sum_{i=1}^r \left( \sum_{j=1}^k b_i(k-j) \right)$  — сумма многочленов с положительными старшими коэффициентами. Утверждение доказано.

*Примечание.* Аналогичным образом можно определить тип роста любого регулярного множества. Для этого используется отвечающий этому множеству конечный автомат.

**59.** Обозначим множество допустимых слов через  $M$ . Пусть язык является  $d$ -определённым. Докажем, что каждое слово  $M$ -эквивалентно слову не более чем из  $d$  букв.

Действительно, если слово  $v$  не является допустимым, то какое слово к нему не припиши, результат не будет допустимым. Поэтому все недопустимые слова эквивалентны. В частности, любое из них эквивалентно какому-нибудь запретному слову, то есть слову длины не большей  $d$ .

Пусть слово  $u$  допустимо и имеет длину, большую  $d$ . Обозначим через  $v$  подслово слова  $u$ , состоящее из его последних  $d$  букв. Пусть  $w$  — произвольное слово. Если слово  $uw$  содержит запретное подслово, то это подслово содержится в  $vw$ , так как длина запретного подслова не больше  $d$ . Поэтому слова  $u$  и  $v$  эквивалентны.

Пусть в алфавите  $k$  букв. Тогда число слов длины не большей  $d$  не превосходит  $(k+1)^d$ . Положим  $n = (k+1)^d + 1$ . В любом наборе из  $n$  слов найдутся два,  $M$ -эквивалентные одному и тому же слову длины не большей  $d$  и, тем самым, эквивалентные друг другу. Поэтому множество допустимых слов регулярно.

**60.** а) Пусть  $S$  — максимальное множество слов, из которых никакие два не  $M$ -эквивалентны. Тогда любое другое слово эквивалентно какому-то слову из  $S$ . Построим конечный автомат. Возьмём  $S$  в качестве множества вершин графа. Для всех  $s \in S$ ,  $a \in A$ , проведём из вершины  $s$  стрелку, помеченную  $a$ , в вершину,  $M$ -эквивалентную  $sa$ . В полученном графе назовём начальной вершиной ту, которая  $M$ -эквивалентна пустому слову, а принимающими вершинами — все слова из  $S$ , которые принадлежат  $M$ . Легко видеть, что построенный конечный автомат принимает те и только те слова, которые принадлежат  $M$ .

б) Любому слову отвечает путь по стрелкам конечного автомата. Ясно, что если для двух слов такие пути заканчиваются в одной вершине, то слова  $M$ -эквивалентны, где  $M$  — множество принимаемых автоматом слов. Поэтому в качестве  $n$  из определения регулярного множества можно взять число, на единицу большее числа вершин в автомате.

**61.** Рассмотрим конечный автомат  $(\Gamma, v_0, W)$ , принимающий множество  $M$ . Для каждой вершины  $v$  этого конечного автомата обозначим через  $T_v$  множество слов, для которых соответствующие пути по графу  $\Gamma$  заканчиваются в  $v$ .

Далее, для каждой вершины  $v$  и каждой буквы  $a$  обозначим через  $U(v, a)$  множество таких вершин  $u$  графа  $\Gamma$ , что из  $u$  в  $v$  идёт стрелка, помеченная  $a$ . Тогда выполнены следующие равенства:

$$T_{v_0}(x) = 1 + \sum_{a \in A} \sum_{u \in U(v_0, a)} x T_u(x) \quad (1)$$

и

$$T_v(x) = \sum_{a \in A} \sum_{u \in U(v, a)} x T_u(x) \quad (2)$$

для  $v \neq v_0$ .

Перенумеруем вершины графа  $\Gamma$ , начиная с  $v_0$ :  $V = \{v_0, v_1, v_2, \dots, v_k\}$ . Заметим, что каждое из равенств (1), (2) можно воспринимать как уравнение вида

$$(1 + x P_i(x)) T_{v_i}(x) = \sum_{j \neq i} x Q_{ij}(x) T_{v_j}(x) + R_i(x), \quad (3)$$

где  $P_i(x), Q_{ij}(x), R_j(x)$  — некоторые известные многочлены, относительно неизвестных рядов  $T_{v_0}(x), \dots, T_{v_k}(x)$ .

Попробуем решить уравнения (3). Выразим из последнего уравнения  $T_{v_k}(x)$  через остальные неизвестные ряды,

$$T_{v_k}(x) = \sum_{j \neq k} x \frac{Q_{kj}(x)}{(1 + x P_k(x))} T_{v_j}(x) + \frac{R_k(x)}{(1 + x P_k(x))},$$

подставим это выражение вместо  $T_{v_k}(x)$  в остальные уравнения, и домножим их все на  $(1 + xP_k(x))$ . Мы получим уравнения того же вида, но число их (как и число неизвестных) станет на единицу меньше. Прделаем то же самое для  $T_{v_{k-1}}(x)$ ,  $T_{v_{k-2}}(x)$ , и т. д., мы получим в конце концов выражение для  $T_{v_0}(x)$  в виде отношения двух многочленов. Подставив его в выражение для  $T_{v_1}(x)$ , получаем, что и этот ряд есть отношение двух многочленов. Продолжая этот процесс, получаем выражения того же вида для всех  $T_{v_i}(x)$ . Осталось заметить, что  $M(x) = \sum_{v \in W} T_v(x)$ .

**62.** Для каждого слова  $v$  обозначим через  $v^{opp}$  слово, состоящее из тех же букв в противоположном порядке. Для каждого множества слов  $M$  положим  $M^{opp} = \{v^{opp} \mid v \in M\}$ . Ясно, что  $M^{opp}(x) = M(x)$  для любого  $M$ . Если  $L$  — язык с множеством запретных слов  $B$ , то через  $L^{opp}$  мы обозначаем язык с множеством запретных слов  $B^{opp}$ .

Вернёмся к задаче. Очевидно, что  $M_w^{opp}$  — это множество допустимых слов языка  $L^{opp}$ , начинающихся с подслова, равного  $w^{opp}$ . Это множество регулярно (доказательство аналогично решению задачи 59). Поэтому ряд  $M_w(x) = M_w^{opp}(x)$  равен отношению двух многочленов.

На самом деле, множество  $M_w$  тоже регулярно, но доказательство этого заняло бы больше места.