Задачи и олимпиады

О повторяющихся подсловах

Р. М. Травкин

Известно, что существует бесконечная последовательность букв трехбуквенного алфавита, в которой никакое слово не повторяется подряд (бесконечное бесквадратное слово). Построение такой последовательности — довольно трудная задача: если записывать буквы слова как попало, то можно прийти к конечному слову, которое нельзя продолжить, сохраняя свойство бесквадратности.

Для 57-й московской математической олимпиады (1993 г.) А. В. Спивак предложил следующую задачу.

Задача 1: *Существует ли последовательность из 32 букв русского алфавита, не содержащая двух одинаковых кусков, идущих подряд, такая что при приписывании справа любой буквы указанное свойство нарушается?*

Построим такую последовательность над произвольным алфавитом из k символов $\{A_1, \ldots, A_k\}$ по индукции. Положим

$$z_1 = A_1, \quad z_2 = A_1 A_2 A_1, \quad \ldots, \quad z_k = z_{k-1} A_k z_{k-1}, \quad \ldots$$

Если к слову $z_k = z_{k-1} A_k z_{k-1}$ приписать символ A_k, то получится слово $z_{k-1} A_k z_{k-1} A_k$, содержащее два одинаковых куска (подслова) подряд. Если приписать любой другой символ A_s, два одинаковых куска в z_k найдутся в силу предположения индукции (z_k заканчивается на z_{k-1}). Задача решена.

В связи с задачей 1 естественно возникает вопрос: *Какова минимально возможная длина l_k последовательности, удовлетворяющей условию задачи?* (k — мощность алфавита.) Из приведённого выше решения получается оценка $l_k \leqslant 2^k - 1$. На 2-м фестивале математических боев «Кубок памяти А. Н. Колмогорова» А. Я. Канель-Беловым была предложена задача, в которой предлагалось найти точный ответ на этот вопрос.

В этой статье мы рассматриваем более общую постановку задачи, когда запрещено n-кратное повторение слова. Получен точный ответ: $n^k - 1$.

Данная работа докладывалась на международной конференции школьников ЮНИОР–99, организованной корпорацией INTEL (научный руководитель — А. Я. Канель-Белов).

Обозначения и определения

Напомним определения и обозначения, используемые при работе со словами.

Слово — это конечная последовательность символов (букв) из некоторого алфавита. Слова мы будем обозначать строчными латинскими буквами, а символы алфавита — прописными. Длиной $|u|$ слова u называется количество букв в слове. Удобно также использовать и пустое слово нулевой длины, которое вообще не содержит букв.

Слово uv получается приписыванием после слова u слова v. Подслово — это подпоследовательность символов, идущих подряд. В образной терминологии олимпиадных задач подслово называется куском. Через w^n будем обозначать слово, состоящее из n раз повторенного слова w (w^0 обозначает пустое слово).

Пусть w — непустое слово. Тогда назовем слово вида ww *квадратичным*, вида www — кубическим, w^n — *n-степенным.* Слово, не содержащее подслов вида ww, www, w^n, будем называть соответственно *бесквадратным, бескубным, n-бесстепенным.* Назовем n-бесстепенное слово *n-критическим*, если при добавлении к нему справа любой буквы алфавита удлиненное слово не является n-бесстепенным, то есть имеет вид xw^n.

Минимальная длина критического слова

Основное утверждение данной работы состоит в точной нижней оценке длины n-критического слова.

Теорема 1. *Минимальная длина n-критического слова в алфавите из k букв равна $n^k - 1$.*

Вначале докажем, что эта оценка достигается.

Предложение 1. *Для любых натуральных k и n существует n-критическое слово в алфавите из k букв.*

Доказательство проводится индукцией по количеству букв алфавита. При $k = 1$ утверждение очевидно. Пусть утверждение выполняется для $k = m$. Обозначим через $u(n, m)$ соответствующее n-критическое слово. Добавим к алфавиту еще одну букву A_{m+1}. Тогда слово

$$u(n, m + 1) = (u(n, m)A_{m+1})^{n-1}u(n, m)$$

является n-бесстепенным.

Действительно, пусть $u(n, m + 1)$ содержит слово $v = w^n$. Тогда в v нет буквы A_{m+1}, поскольку в $u(n, m + 1)$ есть всего $n - 1$ буква A_{m+1}. Значит, v содержится в $u(n, m)$, что невозможно по предположению индукции. В то же время при добавлении к слову $u(n, m + 1)$ как буквы A_{m+1}, так и любой другой буквы, получаем слово, заканчивающееся на n-ую степень. □

По индукции легко проверить, что длина слова $u(n, k)$, построенного при доказательстве предложения 1, равна $n^k - 1$.

Доказывать нижнюю оценку будем следующим образом. Назовем слово *X/N-нерасширяемым*, если при прибавлении к нему справа буквы X оно приобретает вид xw^N. Здесь N – некоторое натуральное число, которое будем называть *индексом X-нерасширяемости*). Если слово в алфавите из k букв ($k \geqslant 1$) является нерасширяемым по всем буквам алфавита соответственно с индексами N_1, N_2, ..., N_k, то назовем его $(N_1, N_2, \ldots, N_k)$-*нерасширяемым.* Ясно, что n-критическое слово будет нерасширяемым по каждой букве алфавита с индексом n (хотя обратное, вообще говоря, неверно).

Лемма 1. *Пусть слово u является $(N_1, \ldots, N_k)$-нерасширяемым. Тогда*

$$|u| \geqslant N_1 \cdot \ldots \cdot N_k - 1. \tag{1}$$

Из этой леммы вытекает нижняя оценка в теореме 1. В самом деле, любое n-критическое слово является $(n, \ldots, n)$-нерасширяемым, а правая часть формулы (1) при $N_1 = \cdots = N_k = n$ дает $n^k - 1$.

Основная конструкция

Определим на словах преобразование $F_{A,N}$. Здесь A — произвольная буква, $N \geqslant 2$ — натуральное число (для краткости вместо $F_{A,N}$ будем писать F, если понятно, какие значения A, N имеются в виду). Пусть $L \geqslant 1$, $X \neq A$. По определению полагаем $F_{A,N}(A^L) = A^{\min\{L-1,N-2\}}$, $F_{A,N}(A^L X) = A^{\min\{L-1,N-2\}}X$, $F_{A,N}(X)$ — пустое слово. Чтобы определить преобразование $F_{A,N}$ на произвольном слове u, разобьем слово u на A-блоки, которые строятся следующим образом.

Каждое слово однозначно представляется в виде

$$u = A^{r_1} X_1 A^{r_2} X_2 \ldots A^{r_s} X_s A^{r_{s+1}}, \tag{2}$$

где s — число букв в слове u, отличных от A, $r_i \geqslant 0$ при $i = 1, \ldots, s+1$ и $X_i \neq A$ при $i = 1, \ldots, s$ (при $s = 0$ равенство (2) принимает вид $u = A^{r_1}$). Используя представление (2), слово u можно разбить на подслова

$$A^{r_1} X_1, \quad A^{r_2} X_2, \quad \ldots, \quad A^{r_s} X_s, \quad A^{r_{s+1}},$$

которые называются A-*блоками* (если u не кончается на A, то подслово $A^{r_{s+1}}$ пусто и не считается A-блоком). Например, слово $A^2B^2ACA^2$ разбивается на A-блоки A^2B, B, AC, A^2, а слово A^2B^2AC — на A-блоки A^2B, B, AC.

Если слово u разложено на A-блоки $b_1, b_2, \ldots, b_s$, то по определению

$$F_{A,N}(u) = F_{A,N}(b_1) F_{A,N}(b_2) \ldots F_{A,N}(b_s).$$

Преобразование $F_{A,N}$ обладает следующими свойствами:

(F1) если u кончается на A и буква $X \neq A$, то $F(uX) = F(u)X$;

(F2) если u не кончается на A, то $F(uv) = F(u)F(v)$;

(F2′) (следствие свойства F2) пусть $u_1, \ldots, u_r$ — произвольные слова, причем $u_1, \ldots, u_{r-1}$ не кончаются на A. Тогда $F(u_1 \ldots u_r) = F(u_1) \ldots F(u_r)$;

(F3) $F(u)$ является началом слова $F(uv)$;

(F4) $F(v)$ является концом слова $F(uv)$;

(F5) каждый A-блок уменьшается на одну или более букв;

(F6) результат преобразования каждого A-блока имеет длину $\leqslant N-1$.

Несложная проверка этих свойств предоставляется читателю.

Доказательство леммы 1 использует следующую лемму

Лемма 2. *Пусть u — X/M-нерасширяемое слово, кончающееся на A и $N \geqslant 2$ — натуральное. Тогда слово $F_{A,N}(u)$ является: a) X/M-нерасширяемым, если $A \neq X$; b) $X/(M-1)$-нерасширяемым, если $A = X$ и $M = N$.*

Вывод леммы 1 из леммы 2

Теорема 1 доказывается индукцией по сумме индексов $N_1 + \cdots + N_k$. Без ограничения общности можно считать, что u оканчивается на букву A, имеющую номер 1 в алфавите.

Случай 1: $N_1 = 1$.

Пусть $\tilde{u}$ — слово, получающееся из u отбрасыванием всех букв A. Тогда $\tilde{u}$ является словом над алфавитом, состоящим из всех букв исходного алфавита, кроме A. Очевидно, что $\tilde{u}$ является $(N_2, \ldots, N_k)$-нерасширяемым и имеет сумму индексов ме́ньшую, чем u. Следовательно, для $\tilde{u}$ справедливо предположение индукции. Поэтому $|u| \geqslant |\tilde{u}| \geqslant N_2 \cdot \ldots \cdot N_k - 1 = N_1 \cdot N_2 \cdot \ldots \cdot N_k - 1$, что и требовалось доказать.

Случай 2: $N_1 > 1$.

В силу леммы 2 слово $F(u)$ является $(N_1 - 1, N_2, \ldots, N_k)$-нерасширяемым, поэтому к нему можно применить предположение индукции. Получим неравенство:

$$|F(u)| \geqslant (N_1 - 1) \cdot N_2 \cdot \ldots \cdot N_k - 1. \tag{3}$$

Пусть X — любая буква алфавита, отличная от A. Поскольку u кончается на A, то

$$F(u)X = F(uX). \tag{4}$$

Ввиду (3) и (4)

$$|F(uX)| = |F(u)X| \geqslant (N_1 - 1) \cdot N_2 \cdot \ldots \cdot N_k. \tag{5}$$

Ввиду (F6) и (5) количество A-блоков в слове $F(uX)$ не менее $N_2 \cdot \ldots \cdot N_k$, поэтому с учетом (F5) получаем оценку:

$$|uX| \geqslant (N_1 - 1) \cdot N_2 \cdot \ldots \cdot N_k + N_2 \cdot \ldots \cdot N_k = N_1 \cdot N_2 \cdot \ldots \cdot N_k,$$

откуда $|u| \geqslant N_1 \cdot N_2 \cdot \ldots \cdot N_k - 1$, что и требовалось доказать.

Доказательство леммы 2

Доказательство a). Поскольку u является X/M-нерасширяемым словом, то найдется такое слово x и такое непустое слово w, что $uX = xw^M$. Согласно свойству (F4), слово $F(uX)$ кончается на $F(w^M)$. Далее, согласно свойству (F2$'$), поскольку w кончается на X, $F(w^M) = (F(w))^M$. Поскольку w кончается на AX (за исключением тривиального случая $M = 1$), то $F(w)$ кончается на X, и, следовательно, не пусто. Применяя свойство (F1), получаем, что $F(u)X = F(uX)$.

Таким образом, получаем, что $F(u)X$ кончается на $(F(w))^M$. Следовательно, поскольку $F(w)$ не пусто, $F(u)$ является X/M-нерасширяемым, что и требовалось доказать.

Доказательство b). Поскольку u является A/N-нерасширяемым, то $uA = xw^N$ для некоторого слова x и непустого слова w. Возможны два случая. Если u кончается на A^{N-1}, то $F(u)$ кончается на A^{N-2}, и, следовательно, является $A/(N-1)$-нерасширяемым.

В противном случае выполняется равенство $F(u)A = F(uA)$. Кроме того, этот случай возможен лишь при $N > 2$. Если бы w состояло из одних букв A, то слово xw^N кончалось бы на A^N. Но это невозможно, поскольку мы предположили, что u не кончается на A^{N-1}. Следовательно, в w есть буква, отличная от A. Поэтому w можно представить в виде zA^k, где $k \geqslant 1$, а z не кончается на A. Тогда слово $uA = x(zA^k)^N$ кончается на $A^k(zA^k)^{N-1} = (A^kz)^{N-1}A^k$. Согласно свойству (F3), найдется непустое слово t, такое, что $F(A^kz) = F(A^k)t$.

Согласно (F2′),

$$F((A^kz)^{N-1}A^k) = (F(A^kz))^{N-1}F(A^k) = \\ = (F(A^k)t)^{N-1}F(A^k) = F(A^k)(tF(A^k))^{N-1}.$$

Поскольку uA кончается на $(A^k)^{N-1}A^k$, то $F(u)A = F(uA)$ кончается на $F((A^kz)^{N-1}A^k)$, которое, в свою очередь, кончается на $(tF(A^k))^{N-1}$. Таким образом, $F(u)$ является $A/(N-1)$-нерасширяемым, что и требовалось доказать.

Р. М. Травкин, студент ВКМ НМУ