How to count words?

Dmitri Piontkovski, Maxim Prasolov, Grigory Rybnikov

Solutions

1 Main problem

1. Cf. Problem 24.

2. Cf. Problems 18 and 48.

3. One version of the solution is given in problem 46; another version can be obtained using Problems 59 and 61.

4. In Problem 1, the alphabet A consists of $N = 100$ words of the tribe language. The phrases of the tribe language play the role of words of L, and the forbidden words of L are the two magic spells. In Problem 2, the alphabet A consists of $N = 256$ commands of the computer, and the programs play the role of words of L.The only forbidden word is the program of 4 commands that breaks the computer.

2 How to write down the answer?

5. To get an arbitrary word of m letters, one choose any of N letters in any of m places. Multiplying the numbers of possibilities in each place, we get N^m words.

6. If the first letter of an admissible word is x, then the second one is x as well. It follows that each admissible word has the form $xx \dots x$, where x is one of the N letters. Therefore, the number of admissible words of any given number of letters is N.

7. Assume that the first letter of an admissible word is a. Since aa is a forbidden word, then the second letter is b. Proceeding by the same way, we get a on the odd places and b on the even places. Similarly, if the first letter is b, then we get a on even places and b one odd ones. Thus, the dimension series of this language is $1 + 2x + 2x^2 + 2x^3 + \dots$

3 The arithmetics of languages

8. The collection of forbidden words is the following: all forbidden words of the both languages and all words of 2 letters such that the first letter is of the first alphabet and the second one is of the the second alphabet. Obviously, the admissible word of each language do not contain a subword which is forbidden in the sum of languages. Let w be a word of sum which does not contain subwords equal to the words described above. If its first letter is, say, in the first alphabet, then the subsequent letters are in the first alphabet as well, that is, each such a word consists of the letters of the same alphabet. It follows that w is admissible in the language of this alphabet, thus, it is admissible in the sum.

9. The initial terms of the series $L(x)$ and $L_1(x) + L_2(x) - 1$ are equal to 1. For $n > 0$, the coefficient of x^n in the series $L_1(x) + L_2(x) - 1$ is equal to the sum of the numbers of words of n letters in the languages L_1 and L_2, that is, the number of words of n letters in the language L, which is equal to the coefficient of x^n in the series $L(x)$.

10. We have

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

as required.

11. The collection of forbidden words is the following: all forbidden words of two given languages and words of two letters such that their first letter is in the second alphabet and the second one is in the first alphabet. Consider an admissible word w of the product. In w, the letters of the second alphabet follow to the letters of the first one, therefore, w has the from $w_1 w_2$, where w_1 is a word of the first language and w_2 is a word of the second one. The word $w_1 w_2$ do not contain subwords which are forbidden in the languages-multipliers, therefore, w_1 and w_2 are admissible in their languages, that is, the word $w_1 w_2$ is admissible in the product of the languages.

12. The coefficient of x^k in the series $L_1(x) \cdot L_2(x) = L_1(x) \cdot n_0 + L_1(x) \cdot n_1 x + \cdots = (n_0 \cdot m_0 + n_0 m_1 x + \dots) + (n_1 m_0 x + n_1 m_1 x^2 + \dots) + \dots$ is $n_0 m_k + n_1 m_{k-1} + \cdots + n_k m_0$. The number of words of length k in the set of words $L_1 \cdot L_2$ is equal to the number of possibilities to get a pair of words, m in L_1 and n in L_2, such that the total number of letters in these words is k. If the word m is of i letters, then the word n is of $k-i$ letters, so that the number of such pairs is equal to $m_i \cdot n_{k-i}$. Taking a sum of all such products for all i, one gets $n_0 m_k + n_1 m_{k-1} + \cdots + n_k m_0$. Thus, the coefficients of x^k in two series $L_1(x) \cdot L_2(x)$ and $L_1 \cdot L_2(x)$ coincide, therefore, the series itself coincide.

13. a) First let us show that the standard properties of addition and multiplication of polynomials (associativity, commutativity, distributivity) hold for series as well. For example, consider associativity relation for multiplication $(P \cdot Q) \cdot R = P \cdot (Q \cdot R)$. To compute the coefficient of x^k in the both sides of this relation, it is enough to make computations for the same series without terms of degree higher than k, i. e., for polynomials. Therefore, the relation for series follows from the same relation for polynomials. The other relations are proved in the same way.

Now note that, since the series $\overline{S}$ starts with x^1, the series $R \cdot \overline{S}^m$ has no terms of degree less than m. That is why infinite sums of the form $R - R \cdot \overline{S} + R \cdot \overline{S}^2 - R \cdot \overline{S}^3 + \dots$ make sense: to find the kth coefficient, the sum can be replaced by a finite one. For the same reason, the sums of this type satisfy the distributivity relation

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

Having this in mind, we easily get

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

Hence

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

b) By the above, we have

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

Assume that the series $(T - \frac{R}{S})$ is nonzero. Since S starts with 1, the first nonzero coefficient of $(T - \frac{R}{S})$ is equal to the first nonzero coefficient of $S \cdot (T - \frac{R}{S})$. Therefore $T - \frac{R}{S} = 0$ and $T = \frac{R}{S}$.

14. a) Similarly to Problem 10, we obtain

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

b) We have

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

and

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

15. The solution follows from Problems 59 and 61.

4 Free word

16. It follows from Problem 17 below that

$$L(x) = \frac{1}{1 - 26x + x^5}.$$

Also, one can directly obtain the above formula in a similar way as in the solution of Problem 17.

17. Let a_k be the number of admissible words of length k. Clearly, $a_0 = 1$. Let us prove the recurrent relation $a_k = Na_{k-1} - a_{k-m}$ for $k > 0$ (we have $a_i = 0$ for $i < 0$, since there are no words of negative length).

Indeed, by adding each letter of the alphabet to the beginning of each admissible word of length $k - 1$, we obtain Na_{k-1} words, among which are all admissible words of length k. Let us find which non-admissible words of length k can be obtained in this way, i. e., can be written as cg, where c is a letter and g is an admissible word of length $k - 1$. Clearly, the forbidden subword must stand at the beginning, so $cg = wf$, where w is the forbidden word and f is admissible. Since w is free, we conclude that, for any admissible word f, the word obtained from wf by cutting its first letter is admissible (otherwise w would have an overlap with itself). Therefore, the set of all words of the form cg, where c is a letter and g is an admissible word of length $k - 1$, is the union of two non-intersecting sets: the set of all admissible words of length k and the set of all words of the form wf, where f is an admissible word of length $k - m$. Hence we get the recurrent relation.

Consider the sum of relation $a_0 = 1$ and all relations $a_k x^k = Na_{k-1}x^k - a_{k-m}x^k$ for $k = 1, 2, 3, \dots$. We obtain

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

Solving this equation with respect to $L(x)$, we get the required formula.

18. According to Problem 17,

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

Here the coefficient of x^7 is $256^7 - 4 \cdot 256^3$. Thus, the probability of computer break is $\frac{4 \cdot 256^3}{256^7} = 4 \cdot 256^{-4}$, or, approximately, 10^{-10}.

5 Transformations of words

19. The first arrow is determined uniquely; the second one maps each word in N to the same word regarded as an element of F_A; the third one maps each of the remaining words in F_A to the same word as an element of G; the last arrow is as trivial as the first one.

20. Each of the children eats as many bonbons that belonged to boys as bonbons that belonged to girls. The last girl eats the last bonbons that belonged to boys. Thus she also eats the last bonbons that belonged to girls, and the teacher gets nothing at all.

21. a) Let $M_{\text{odd}} = M_1 \cup M_3 \cup M_5 \cup \dots$ and $M_{\text{even}} = M_2 \cup M_4 \cup M_4 \cup \dots$. Each transformation establishes a one-to-one correspondence between a subset of M_{odd} and a subset of M_{even}, besides, since both the rightmost and the leftmost sets are empty, each element participates in exactly one of these correspondences. Hence the sets M_{odd} and M_{even} have the same number of elements.
b) For each k, the set $M_i^{(k)}$ of words in M_i of length k is finite; by applying assertion a) to the finite sets $M_i^{(k)}$ with the same k, we conclude that the coefficient of x^k in the left-hand side of the formula is the same as in its right-hand side. Since k is arbitrary, it means that the formula is correct.

22. Let us verify that the set $A \cdot G$ is the union of two non-intersecting sets: the set $\overline{G}$ and the set $B \cdot G$. The proof, which is based on the fact that the set B is free, is almost literally the same as the corresponding reasoning in the solution of Problem 17.

Now it is easy to construct the required exact sequence: the first and the last arrows are trivial, the second one maps each element of $B \cdot G$ to itself (here we use that $B \cdot G \subseteq A \cdot G$), and the third one maps each of the remaining elements of $A \cdot G$ to itself (here we use that $A \cdot G \setminus B \cdot G = \overline{G}$). In particular, the kernel of the second transformation is empty, and the kernel of the third transformation, which is the same as the image of the second one, is $D \cdot G$; the image of the third transformation is $\overline{G}$.

23. By Problem 21b), the exact sequence in Problem 22 implies

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

Note that each element of $A \cdot G$ can be *uniquely* written as ag, where $a \in A, g \in G$. Hence $(A \cdot G)(x) = A(x)G(x)$. Further, each element of $B \cdot G$ can be *uniquely* written as bg, where $b \in B, g \in G$, (since no forbidden word is a subword of another forbidden word). Hence $(B \cdot G)(x) = B(x)G(x)$. We have $A(x) = Nx$, $G(x) = L(x)$, $\overline{G}(x) = G(x) - 1 = L(x) - 1$. Therefore,

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

Solving this equation with respect to $L(x)$, we get the required formula.

24. Denote the words that occur in the spells by letters: "earth" — A, "stand" — B, "on" — C, "great" — D, "crocodile" — E, "every" — F, "evening" — G, "swallow" — H, "sun" — I. Then the spells correspond to forbidden words "ABCDE" and "FGEHI". These words are free (since all letters in each of them are distinct) and have no overlaps with each other (since both the first and the last letters of the second word do not occur in the first one). Therefore, the set of spells is free, and and the dimension series for the language is

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

Using this formula, it is not hard to show (see the solution of Problem 17), that a_k (the number of sentences of k words) can be computed from the initial condition $a_0 = 1$ and the recurrent relation $a_k = 100a_{k-1} - 2a_{k-5}$. The computations provide us with the answer $a_{20} = 10^{40} - 32 \cdot 10^{30} + 264 \cdot 10^{20} - 448 \cdot 10^{10} + 16$.

25. Letter v occurs only as the first letter of each forbidden word and all forbidden words are of length 4. Hence the set of forbidden words is free. By Problem 23, we have

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

26. If the set of forbidden words is free, then, in particular, there are no simple linkages. Let us prove that if there are no simple linkages, then the set of forbidden words is free. Assume the contrary, i. e., that the set of forbidden words is not free. Then there is an overlap of two forbidden words, that is, there exist three nonempty words s, t, r such that the words st and tr are forbidden. Choose such a triple (s, t, r) so that the length of str be minimal. If it is not a simple linkage, then str has a forbidden subword w other than st and tr. Note that the end of w is not he end of tr since otherwise either w would be a subword of tr or tr would be a subword of w, which is impossible as no forbidden word is a subword of another forbidden word. Similarly, the beginning of w is not the beginning of st. For the same reason, the subword w overlaps with both s and r. Denote the common part of st and w by t', the remaining part of st by s', and the remaining part of w by r'. These words are nonempty, the length of $s't'r'$ is less than the length of str, the words $s't' = st$ and $t'r' = w$ are forbidden. We obtain a contradiction. Thus, if there are no simple linkages, the set of forbidden words is free.

27. We construct the transformations starting from the end (from the rightmost arrow). Since the last set is empty, the domain of definition of the last transformation is also empty. Hence the image of the last but one transformation is the whole set $\overline{G}$. Since $\overline{G} \subseteq A \cdot G$, we can take $\overline{G}$ to be the domain of definition of the last but one transformation, and define the corresponding function to map each element $g \in \overline{G}$ to itself. The kernel of this transformation consists of all non-admissible words of the form ag, where a is a letter and g is an admissible word. It is readily seen that, for any word of this type, there exist a forbidden word w and an admissible word f such that $ag = wf$ (we already used similar reasoning in the solutions of Problems 17 and 22). So we can construct the third arrow from the right (this transformation also maps each element of its domain to itself). Consider the kernel of this transformation. It consists of those words of the form wf, where w is forbidden and f is admissible, which also have the form av, where a is a letter and v is a non-admissible word. Choose the leftmost forbidden subword u in v. Clearly, the subword u of the word $av = wf$ overlaps with the subword w and forms a simple linkage with it. Thus the kernel of the third arrow from the right is contained in $S \cdot G$. Hence it is possible to construct transformation $S \cdot G \Longrightarrow B \cdot G$ (which also maps each element of its domain to itself).

28. By the solution of the previous problem, we see that a language is non-tangled if and only if all words of the form rg, where r is the tail of a simple linkage and g is an admissible word, are admissible. An equivalent condition for the set of forbidden words writes as follows: there exist no such words p, q, r, s, t, where the p, q, s, t are nonempty, the words pq, qrs, st are forbidden, and $pqrs$ is a simple linkage.

Note that any element of the set $S \cdot G$ can be uniquely represented as the product of a simple linkage by an admissible word (it follows easily from the definition of simple linkage and the fact that no forbidden word is a subword of another forbidden word). In the same way as in the solution of Problem 23, for a non-tangled language L, we use the exact sequence to obtain the following equation:

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

whence follows the required formula

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

29. Simple linkages are *abbc* and *abbac*, and their tails are c and ac. It is clear that none of these tails ends with the beginning of a forbidden word. Thus the language is non-tangled. Therefore, the dimension series is

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

30. Simple linkages are $x_i y_j z_k$, where $1 \le i, j, k \le n$, their tails are z_k. Since no forbidden word starts with z_k, the language is non-tangled. Therefore, the dimension series is

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

31. Let w be the unique forbidden word and let L be its length. Suppose that w is not free; let pqr, where $pq = qr = w$, is a simple linkage. We have $wr = pqr = pw$. Therefore, the last subword of length L in each of the words $wr = pw, wrr = pwr = ppw, wrrr = ppwr = pppw, \dots$ is equal to w. Take the first word in this sequence which has length at least $2L$. Then a word of the from $rrr \dots r$ has the final subword equal to w. But this means that the word r has a nonempty ending which is an initial subword of w. Therefore, the language under consideration is tangled (Cf. the solution of Problem 28).

6 Free sets revisited

32. For example, if the alphabet consists of the letters a and b, then the set of words $a^n b^n ab$, where $n \ge 2$, is free. Let us prove this. Obviously, no two words are subwords of each other. It remains to prove that there is no nontrivial overlap (i. e., each overlap is the letter-by-letter application of a word on itself). Let w is an overlap of the words $a^n b^n ab$ and $a^m b^m ab$. Then it is easy to see that w has at least three letters. Since w is an end of the word $a^n b^n ab$, it has the form either $b^k ab$ or $a^k b^n ab$, where $1 \le k \le n$. Because w is also a begin of the word $a^m b^m ab$, we get $k = m = n$ and $w = a^n b^n ab = a^m b^m ab$ is a trivial overlap.

33. Lemma. Let $p(x) = 1 + p_1 x + p_n x^n$ be a polynomial of degree $n \ge 1$. Then the series $f(x) = 1/p(x)$ cannot be a polynomial (i. e., this series has an infinite set of nonzero terms).

Proof of Lemma. Suppose (ad absurdum) that the series $f(x)$ is a polynomial, that is, $f(x) = f_0 + f_1 x + \dots f_m x^m$, where the leading coefficient f_m is nonzero. According to Problem 13 a), we have $1 = f(x)p(x) = 1 + (f_0 p_1 + f_1 p_0)x + \dots + f_m p_n x^{m+n}$, a contradiction.

Return to Problem 33. According to Problem 23, we have

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

If the language L was finite, the series $L(x)$ would be a polynomial, in contradiction with the above Lemma. It follows that the set of admissible words is infinite.

34. Obviously, for series A and B the inequality $A \ge B$ is equivalent to an inequality $A - B \ge 0$, which is equivalent to the condition that the coefficients of the series $A - B$ are nonnegative. Denote the series $P - Q$ by $A = a_0 + a_1 x + a_2 x^2 \dots$, and the series R by $R = r_0 + r_1 x + r_2 x^2 \dots$ Then an n-th coefficient of the series AR is given by the formula $a_0 r_n + a_1 r_{n-1} + \dots + a_n r_0$. So, a_n is a sum of nonnegative numbers, so that $a_n \ge 0$. This means that it holds an inequality $AR \ge 0$. Equivalently, we have $PR - QR \ge 0$, or $PR \ge QR$.

35. According to Problem 23,

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

By Problem 27, there is an exact sequence

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

where K is a kernel of the transformation $B \cdot G \Longrightarrow A \cdot G'$ and G' is the set of admissible words of the language L'. It follows from this exact sequence (Problem 21) that

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

therefore (since $B'(x) = B(x)$, $L'(x) = G'(x)$ and $K(x) \geq 0$),

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

Multiplying this by the series $L(x) \geq 0$, we get (using Problem 33)

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

i. e.,

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

36. Let $A = \{a, b\}$ be the alphabet. Denote the second word by v. Obviously, if the words v and w are free, then their initial and last letters differ. If, in addition, w the initial letters of the two words differ, then the last letter of v coincides with the first one of w, and , so that the set B is not free. So, it remains to consider the case when v and w begin with the same letter (say, a) and end with another one (b).

a) We have $w = ab$ and $v = a...b$. Obviously, if the first appearing of b in the word v is in k-th place, then the subword of v which consists of the $(k-1)$-th and k-th letters is w. It follows that B is not free.

b) Answer: no. Let $w = aab$ (the case $w = abb$ is analogous, up to the right–left symmetry and the interchanging the letters). Since the word w is not a subword of v, in v the letters that follows the pair of letters aa is again a. Since the words v and w have overlaps, v cannot begin with ab, i. e., v begins with aa. Therefore, the 3rd letter of v is a, as well as the 4th etc. It follows that $v = aa \dots a$, a contradiction.

37. It is sufficient to show that there exist a free set B of $g = \lceil n^2/4 \rceil$ two-letter words. Let $k = [n/2]$, i. e., $n = 2k$ or $n = 2k + 1$. Put $B = \{x_i x_j | 1 \leq i \leq k, k+1 \leq j \leq n\}$. Obviously, the set B is free. Then for an even $n = 2k$, the set B consists of $k^2 = n^2/4$ elements, and in the case of odd $n = 2k+1$ the set B is of $k(k+1) = (n-1)(n+1)/4 = n^2/4 - 1/4 = \lceil n^2/4 \rceil$ elements, as required.

38. It is sufficient to show that there exist a free set B of $m \leq k^d (d-1)^{d-1}$ words of length d. Let us divide the alphabet $A = \{x_1, \dots, x_n\}$ by two subsets $P = \{x_1, \dots, x_k\}$ and $Q = \{x_{k+1}, \dots, x_n\}$. Then is es easy to see that the set $B = \{pq | p \in P, q \in F_Q —\quad d-1\}$ is as needed.

39.
a) According to Problem 23, $\frac{1}{1-nx+B(x)} = L(x) \geq 1$.
b) Answer: no. For example, over a 2-letters alphabet A there is no such a set B with $p(x) = x^3 + x^{10}$ (it is shown in Problem 36 b). It remains to see that the series

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

has nonnegative coefficients (since the initial term of the above series is 1, the above condition is equivalent to the required inequality $f(x) \geq 1$). We will show that the coefficients of the series satisfy the stronger inequality $a_n \geq (3/2)^n$. One can provide the proof of the last inequality by induction, using the reccurent relation $a_{n+10} = 2a_{n+9} - a_{n+7} - a_n$, which follows from the condition $(1 - 2x + x^3 + x^{10})f(x) = 1$.

Another similar example is the case $p(x) = 4x^6$, again over the two-letter alphabet.

40. The proof is given in Theorem 5.1 (equivalence A$\Longleftrightarrow$B) and Proposition 5.6 in the paper: 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. The question is still open.

7 Words and chains

42. The word “of” has not any overlaps with the other words because no words begin with the letter f and do not end with the letter o. There is no letter s in the word “tournament” so a chain can consist the word “towns” only as the last arc. A letter t is only on the first and the last position in the word “tournament”. So overlaps of the word “tournament” with itself are only “tournamentournament”.

Now we find all chains: tournament, of, towns; tournamentournament, tournamentowns;... There are two chains of the length n for $n > 1$.

43. From the arc on the right of a chain construct an antichain leftward arc by arc. We obtain a unique antichain of the length i if it exists. We prove by induction that the beginning of the ith arc of the antichain lies between the beginning of the ith from the right (numeration of arcs is from the right) and the end of the (i+1)th chain arcs. The first arcs of chain and antichain coincide, therefore the base of induction holds for $i = 1$. Now check the induction statement for $i = 2$. The second antichain arc is on the right from the second chain arc because there are not forbidden words between the first and the second antichain arcs. The beginning of the second antichain arc is on the left from the end of the third (from the right) chain arc because there is no forbidden words after the end of the third chain arc by the definition of a chain. So the base holds for $i = 2$. Then we prove the induction step. Suppose the induction statement is true for 1,2,...,i. By the induction assumption the ith chain arc intersects the (i−1)th antichain arc, but the (i+1)th antichain arc does not intersect the (i−1)th antichain arc, therefore the (i+1)th antichain arc is on the left from the ith chain arc. Between the end of the (i+2)th and the beginning of the ith chain arcs no forbidden words begin, therefore the beginning of the (i+1)th antichain arc is on the left from the end of the (i+2)th chain arc. The (i−1)th antichain is not on the left from the (i−1)th chain arc, so the (i+1) chain arc does not intersect the (i−1)th antichain arc, but intersects the ith arc. Therefore the (i+1)th antichain arc is not on the left from the (i+1)th chain arc. This finishes the proof.

44. Assume that a chain c' is a subword of a chain c. Let us prove by induction that the ith c-arc is not on the right from the ith c'-arc. The statement is obvious for $i = 1$. If the first arcs of considered chains coincide and the chains have the same length then their arcs coincide and so the whole chains are equal. Hence the chain c' does not start from the beginning of the word c. Between the first and the second c-arcs there is no forbidden words. Therefore the first c'-arc is not on the left from the second c-arc and then the second c'-arc is on the right from the second c-arc. So we proved the base for $i = 2$. Now we prove an induction step. We consider two cases.

The first case. The (i+1)th c'-arc does not intersect the ith c-arc. Then the (i+1)th c'-arc is on the right from the (i+1)th c-arc.

The second case. The (i+1)th c'-arc intersects the ith c-arc. The (i+1)th c'-arc does not intersect the (i−1)th c'-arc, therefore by induction hypothesis the (i+1)th does not intersect the (i−1)th c-arc. Summarizing observations we obtain that by definition of chain the (i+1)th c-arc is not on the right from the (i+1)th c'-arc.

The induction statement is proved and we must only consider the case when the last c- and c'- arcs coincide. In this case we note that these chains as antichains (by the previous problem) have the same arcs on the right and the same length. Therefore they coincide.

45. Assume that a word is decomposed as gc where g is admissible word and c is a chain of the length at least two. By the problem 43 the chain c is also an antichain. If you reduce the chain c by the beginning of the first arc and it will appear a forbidden word that is on the left from the reduced chain then the antichain c can be continued to the left and we obtain a new decomposition $g'c'$ where the chain length is increased by one. If a forbidden word does not appear then the antichain c can be reduced by the beginning of the arc on the left and we obtain a new decomposition $g''c''$ where the chain length is decreased. If the third decomposition $g''c''$ exists then note that the arcs of the antichains c, c', c'' are the same so there are two chains among them which lengths differ at least by two. But in this case the arc on the left of the longest antichain does not intersect the shortest antichain and we have a contradiction.

46. We will apply the problem 21 for an exact sequence

$$\ldots \Longrightarrow C_{n+1} \cdot G \Longrightarrow C_n \cdot G \Longrightarrow C_{n-1} \cdot G \Longrightarrow \ldots C_1 \cdot G \Longrightarrow A \cdot G \Longrightarrow \bar{G}$$

Let us construct this sequence. Choose a word cg from $C_n \cdot G$. Add a tail of the chain c to the beginning of the word g and obtain a decomposition $c'g'$. If g' is an admissible word then the word $c'g'$ belongs to $C_{n-1} \cdot G$. If g' contains a forbidden word then the chain c can be continued to the right to the chain c'', the rest of the word denote by g''. Then the word $c''g''$ belongs to $C_{n+1} \cdot G$. A chain can be continued in unique way in the word, so the constructed maps are biunique on the parts of $C_n \cdot G$.

The arrows $C_1 \cdot G \Longrightarrow A \cdot G \Longrightarrow \bar{G} \Longrightarrow \emptyset$ are the same as in the problem 27. By the problem 21,

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

So we obtain the formula.

47. By the problem 42 $C_1(x) = x^2 + x^5 + x^{10}, C_n(x) = x^{9(n-1)}(x^5 + x^{10})$. By the formula from the problem 46

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

$$= \frac{1 + x^9}{1 - 26x + x^2 + x^5 + x^9 - 25x^{10} + x^{11}}$$

48. Consider four cases of a forbidden word of four letters.

1) The forbidden word is of the form aaaa. In that case $C_{2n} = x^{4n+1}, C_{2n-1} = x^{4n}$. By the formula in the problem 46

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

Then a coefficient of x^7 equals to $256^7 - 3\cdot 256^2\cdot 255 - 256^3 = 256^7 - 4\cdot 256^3 + 3\cdot 256^2$.
2) The forbidden word has the form abca and at least two distinct letters. In that case $C_n = x^{3n+1}$.

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

A coefficient of x^7 equals to $256^7 - 4\cdot 256^3 + 1$
3) The forbidden word is of the form abab. In that case $C_n = x^{2(n+1)}$.

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

A coefficient of x^7 equals to $256^7 - 4\cdot 256^3 + 2\cdot 256$.
4) The forbidden word is free. This case was analyzed in the problem 18.
49. From the 46 it follows that

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

The left part of this equality is a dimension series of the set of inadmissible words. The right part of this equality is an alternative sum of dimension series of langugages $L\cdot C_n\cdot F_A$. One can define a map from these labguages to the set of inadmissible words and vice versa, every inadmissible word can be decomposed as gc_nu, where g is admissible, c_1 is a forbidden word. But one can decompose an inadmissible word as gc_nu in several ways. Denote the number of such decompositions of the word w as w_n. So then the sum of the numbers $(w_1-w_2+w_3-w_4+\dots)$ for all inadmissible words of the length k equals to a coefficient of x^k in the right part of the equality behind – and therefore in the left part, that is equal to the amount of inadmissible words of the length k.

Note that from the decomposition of the word w as gc_nu one can obtain another decomposition, in which the chain length is decreased by one, by moving a tail of c_n to u. So two these decompositions will be cancel in the sum $(w_1-w_2+w_3-w_4+\dots)$. In other words the sum $(w_1-w_2+w_3-w_4+\dots)$ equals to the quantity of decompositions of the word w as gcu where subword c is a maximal chain of odd length.

Consider a decomposition of the word w as gcu where c is a maximal chain with the most right [1]arc. By the problem 45 in the word gc there is a maximal chain of the length one or it can be decomposed as $g'c'$ where g' is admissible and c' is a chain which length differs by one from the length of the chain c. In both cases there is a maximal chain of odd length in the word w. So we showed that $(w_1-w_2+w_3-w_4+\dots)$ is at least one. But the sum of such quantities for all inadmissible words equals to their number. So then every such quantity equals to one and in every inadmssible word there is only one maximal chain of odd length.
50. Apply the formula from the problem 46.

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

51. Apply the formula from the problem 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. Admissible words of the language M are the chains of the language L. Chains of the length n consist of $n+1$ letters. Therefore $C_n(-x) = (-1)^{n+1}C_n(x)$. Thus we obtain $M(-x) = 1-Nx+C_1(x)-C_2(x)+\dots$, i.e. $L(x)M(-x) = 1$.

8 Additional problems

53. The existence of a free subset under the assumption $m \le k^d(d-1)^{d-1}$ is established in Problems 37 and 38. It remains to show that no such set exists for $m > k^d(d-1)^{d-1}$.

[1]this means that any arc of any maximal chain is not on the right from the last arc of the chain c

a) If we are given $m > n^2/4$ words of length two, which form a free set S, then the first letter of any of these words cannot coincide with the last letter of another word, i. e., there are two disjoint subsets of letters, P and Q, whose elements may only serve as the first letter or the last letter, respectively, for a word in S. Let $r = |P| + |Q| \le n$, and let $s = |P| \cdot |Q| \ge |S| = m > n^2/4$. By the Viet theorem, the numbers $|P|$ $|Q|$ are the roots of quadratic equation $x^2 - rx + s = 0$, whose discriminant $D = r^2 - 4s$ is negative under the above constrains on r and s; hence we get a contradiction.
b) Let B be a free set consisting of m words of length 3. Since no letter can be both the first and the last letter for words of the same free set, the alphabet A contains two disjoint subsets X and Y, whose element can serve as the first letter or the last letter, respectively, for a word in B. If there are letters that does not occur as the first or the last letter of a word in B, we add each of them to one of X and Y. Without loss of generality, we assume that the number s of elements of the set $X = \{x_1, \dots, x_s\}$ is no greater than the number t of elements of $Y = \{y_1, \dots, y_t\}$. Each element B has the form $x_{i_1}x_{i_2}y_{j_1}$ or $x_{i_1}y_{j_1}y_{j_2}$, and no final subword xy of an element of the first type can be the beginning of an element of the second type. Let us change the set B by replacing each word of the first type with a word of the second type according to the rule $x_ixy \mapsto xyy_i$, where $1 \le i \le s \le t$. It is readily seen that such transformation maps no element of B to another element of B, no distinct elements are mapped to the same one, and the resulting set B' remains free. All elements of B' are of the form $x_{i_1}y_{j_1}y_{j_2}$. Therefore, the number of the elements of B' (which is still m) does not exceed st^2, whence

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

as required.
c) The proof uses the following analytical
Lemma. Let $R(x) = 1 + a_1x + a_2x^2 + \dots$ be a series with positive integer coefficients. Suppose that $R(x) = 1/p(x)$ for a polynomial $p(x)$ with constant term 1. Let $R_n(x) = 1 + a_1x + a_2x^2 + \dots + a_nx^n$. If we have $p(x) \ge m > 0$ for all $x \in [0, x_0]$, where $x_0 > 0$, then inequality $R_n(x_0) \le 1/m$ holds for each $n > 0$.

Omitting the proof of the Lemma, we pass to the solution of the Problem. Let $s = mk^{-d} - (d-1)^{(d-1)}$; we need to show that no free set exists if $s > 0$. Assume the contrary. Then, by Problem 39 a), the series $1/p(x)$, where $p(x) = 1 - dkx + mx^d$, has positive integer coefficients (there can be no zero coefficients, since the series is infinite by Problem 33). Note that the polynomial $p(x)$ is positive on the segment $[0, 1]$ whenever $s > 0$ (proof: the minimum of this polynomial on $[0, 1]$ is achieved either at an end of this segment, where $p(x)$ is positive, or at a point x_0 such that $p'(x_0) = 0$, i. e., at $x_0 = \frac{1}{k(d-1)}$; then $p(x_0) = sx_0^d > 0$). It means that there is a number $m > 0$ such that $p(x) \ge m$ for $x \in [0, 1]$. By Lemma, it follows that, for all n, the number of words of length at most n, which is $L_n(1)$, is bounded by the constant $1/m$.

54. a) By definition, the forbidden words of $L^!$ are all two-letter words that are not forbidden in L. Hence the forbidden words of $(L^!)^!$ are all two-letter words that are not forbidden in $L^!$, i. e. exactly all forbidden words of L. Thus the alphabets and the sets of forbidden words for languages L and $(L^!)^!$ are the same, hence the languages are equal.
b) Since the set of forbidden words for the language $M = (L_1 + L_2)^!$ is the union of the sets of forbidden words for the languages $L_1^!$ and $L_2^!$, and the alphabet of M is the union of their (disjoined) alphabets, the language M is the free product of $L_1^!$ and $L_2^!$ (see definition in Problem 51).
c) The forbidden words for $(L_1 \cdot L_2)^!$ are the admissible two-letter words of the languages L_1 and L_2 and all words of the form aB, where a is a letter of the alphabet of L_1 and B is a letter of the alphabet of L_2. Hence

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

55. Let w be a word of length nk (where $k \ge 1$) over the alphabet of L, and let $w^{(n)}$ be the corresponding word of $L^{(n)}$. Let us break w into subwords $w = w_1 \dots w_k$ each of which corresponds to a letter of the language $L^{(n)}$. It is readily seen that w has a forbidden subword u (which consists, by definition, of at most d letters) if and only if there is a subword $w' = w_p \dots w_{p+m-1}$ such that each w_i either is contained in the word u or overlaps with it, so that the number m of n-letter pieces in w' satisfies the inequality $m \le s$, where

$$s = 2 + \left[\frac{d-2}{n}\right].$$

Thus any non-admissible word $w^{(n)}$ of $L^{(n)}$ contains a non-admissible word of length at most s, hence the language $L^{(n)}$ can be defined by a finite set of forbidden words, and the length of each forbidden word is no greater than s. This proves part a) of the problem.
b) Answer: not always.

Let us prove that the lengths of forbidden words of $L^{(n)}$ are less than d if $d \ge 3$ and $n \ge 2$; in particular, this language is not d-defined, which gives a negative answer to b). It suffices to prove the inequality $s < d$, or

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

The last inequality is clearly equivalent to inequality $(d-2)(1-1/n) > 0$, which is obvious under the given constrictions on d and n.

c) Answer: $n = d-1$.

By the above, the language $L^{(n)}$ is either quadratic or free (i. e., the lengths of forbidden words do not exceed 2) under assumption $s \leq 2$, which is equivalent to inequality $2 + \frac{d-2}{n} < 3$, or $n > d-2$, i. e., $n \geq d-1$. On the other hand, if $n \leq d-2$, then there exists a d-defined language L such that the language $L^{(n)}$ has forbidden words of more than three letters: for example, we can take the language L over the three-letter alphabet $\{a, b, c\}$ with one forbidden word abc^{d-2}.

56. See the solution of Problem 58.

57. Answer: yes.

For example, let A be an alphabet of $n \geq 2$ letters. Consider the language $L = F_A \cdot F_A^!$. Since F_A has exponential growth (in the statement of Problem 55 c) we can take $c_1 = n+1$ and $c_2 = n$), and since $2F_A(x) \geq L(x) \geq F_A(x)$, the language L also has exponential growth. By Problem 53 c), we have $L^! = (F_A^!)^! \cdot F_A^! = L$, thus both L and $L^!$ have exponential growth.

58. First we prove the following assertion (it is not necessary for solving Problem 56 only).

Lemma. Let $a = \{a_0, a_1, a_2, \dots\}$ be a sequence such that $a_0 = 1$ and the inequalities $a_1 \geq 2, \dots, a_N \geq 2$ for some positive integer N. Then the sequence a has polynomial (respectively, exponential) growth if and only if the corresponding inequalities in assertions b) and) hold for all a_k with $k \geq N$.

Proof of the Lemma. Let $M = \max\limits_{i \leq N}\{a_i\}$. It is clear that if, for some polynomials p, q of degree d, the inequalities $p(k) \geq a_k \geq q(k)$ hold for $k \geq N$, the inequalities $p(k) + M \geq a_k \geq q(k) - M$ hold for all k. The Lemma for the case of polynomial growth follows. Similarly, if $c_1^k \geq a_k \geq c_2^k$ for $k \geq N$, then $(M + c_1)^k \geq a_k \geq g^k$ for all k, which completes the proof of the Lemma.

Let us pass to the solution of the problem. Clearly, to get all admissible words of length $\geq d-1$, we can do the following. We start with the word at a vertex of the graph. Then we go along a path that starts at this vertex, and each time we read a letter on an edge that we pass, we add this letter to the right of our word. Clearly, different words are obtain from different paths. It is readily seen that the language is finite if and only if no path returns to the initial vertex, that is, the graph has no cycles (it proves assertion a)). It remains to consider the case when the language is finite and there is a cycle in the graph. In this case the number a_j of words of length $j \geq d$ is equal to the number of paths of length $j - d + 1$.

Assume that there are two intersecting cycles; let their lengths be d_1 and d_2, and let v be their common vertex such that the edges issuing from it when we go along the two cycles are distinct (and correspond, say, to letters x and y). The words that we read on edges when we walk by paths of length k that start at v and go along each of the cycles are distinct, hence $a_k \geq 2$ for all $k \geq 0$. Moreover, for each $j = (d-1) + q(d_1 + d_2) + r$, where $r < d_1 + d_2$ is the remainder of division of $j - d + 1$ by $d_1 + d_2$, there exist at least 2^q distinct paths of length $j - d + 1$ (on each of q steps we go along both cycles in an arbitrary order, and then make r steps in an arbitrary cycle), thus for $j \geq 2d$ we have $a_j \geq 2^q = 2^{\left[\frac{j-d+1}{d_1+d_2}\right]} \geq c^j$, where $c = 2^{1/2(d_1+d_2)}$. Since always $a_j \leq n^j$, the Lemma (for $N = 2g$) implies that the growth is exponential.

It remains to consider the case when the graph Γ_L has cycles, but they do not intersect each other. It suffices to verify the polynomiality condition for the number $b_k = a_{k+d-1}$ of paths of length k in the graph Γ_L (since if the corresponding inequalities hold for b_k, then they also hold for a_k for $k \geq d-1$ after the polynomials $p(x)$ and $q(x)$ are replaced with polynomials of the same degree $p_1(x) = p(x+d-1)$ and $q_1(x) = q(x+d-1)$). We will prove that each term of the sequence b_k is equal to the value of some polynomial $b(k)$ with positive highest coefficient (we say that such sequences are *polynomial*).

Let us consider another graph Γ'_L, whose vertices are the cycles of Γ_L and those vertices of Γ_L that belong to no cycle (the latter will be referred to as *isolated* vertices), and whose edges correspond to the edges that connect the corresponding components (cycles or isolated vertices) of Γ_L. It is clear that the graph Γ'_L has no cycles, i. e., the set of paths in it is finite. Let Q^v be the set of paths in Γ'_L that start at a given vertex v, and let q_k^v be the set of the corresponding paths of length k in Γ_L. Since $b_k = \sum_v q_k^v$, it suffices to show that the sequence $\{q_k^v\}$ is polynomial for each vertex v. We proceed by induction on the length $D = D(v)$ of a maximal path that starts at v. If $D = 0$, then either $q_k^v = 0$ for $k > 0$ (if v is an isolated vertex), or $q_k^v = 1$ for all k (if v is a cycle), thus the corresponding sequence is always polynomial. Let now v be an initial vertex of Γ'_L, from which r arrows $a_1, \dots, a_r$ issue to vertices $v_1, \dots, v_r$ (which are not necessarily distinct). By induction, we assume that $q_k^{v_i} = b_i(k)$ is a polynomial with positive highest coefficient. If v is an isolated vertex, then $q_k^v = \sum_{i=1}^r q_{k-1}^{v_i}$, hence this sequence is polynomial as the sum of polynomial sequences. On the other way, if v is a cycle, then, before we pass along one of the edges $a_1, \dots, a_r$ a word of any length is possible in the cycle, hence $q_k^v = \sum_{i=1}^r \left(\sum_{j=1}^k q_{k-j}^{v_i}\right) = \sum_{i=1}^r \left(\sum_{j=1}^k b_i(k-j)\right)$ is the sum of polynomials with positive highest coefficients. This completes the proof.

Note. It is possible to define the growth of any regular set in a similar way. To this end, the corresponding finite automaton is used.

59. Let M be the set of admissible words. Assume that the language is d-defined. Let us prove that each word is M-equivalent to a word of no more than d letters.

Indeed if a word v is non-admissible, then, for any word w, the word vw is also non-admissible. Hence all non-admissible words are equivalent. in particular, any of them is equivalent to a forbidden word, which is of length at most d.

Assume that u is an admissible word of length greater than d. By v denote the subword of u consisting of its last d letters. Let w be an arbitrary word. If the word uw has a forbidden subword, then this subword is contained in vw, since the length of any forbidden word is at most d. Hence the words u and v are equivalent.

Let the alphabet have k letters. Then the number of words of length at most d is no greater than $(k+1)^d$. Let $n=(k+1)^d+1$. In any set of n words there exist two words that are M-equivalent to the same word of length at most d and, thus, to each other. Therefore, the set of admissible words is regular.

60. a) Let S be a maximal set of words in which no two words are M-equivalent. Then any other word is equivalent to one of S. Let us construct a finite automaton. Take S as the set of vertices of the graph. For all $s \in S$, $a \in A$, we draw an arrow marked by a from s to the vertex that is M-equivalent to sa. We say that the vertex which is M-equivalent to the empty word is the initial vertex of the automaton, and all elements of S which belong to M are the approving vertices of the automaton. It is easy to see that the automaton approves a word if and only if it belongs to M.
b) Any word determines a path along arrows of the finite automaton. Clearly, if to words determine paths ending at the same vertex, then these words are M-equivalent, where M is the set of all words approved by the automaton. Hence the number n in the definition of regular set can be taken to be one plus the number of the vertices of the automaton.

61. Consider a finite automaton (Γ, v_0, W) which approves the set M. For each vertex v of Γ, denote the set of words for which the corresponding paths in Γ end at v by T_v.

Further, for each vertex v and each letter a denote by $U(v,a)$ the set of all vertices u of Γ such that there is an arrow marked by a from u to v. Then the following relations hold:

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

and

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

for $v \neq v_0$.

Let us number the vertices of Γ, starting with v_0: $V = \{v_0, v_1, v_2 \ldots, v_k\}$. Note that each of relations (1), (2) can be viewed as an equation of the form

$$(1 + xP_i(x))T_{v_i}(x) = \sum_{j\neq i} xQ_{ij}(x)T_{v_j}(x) + R_i(x), \tag{3}$$

where $P_i(x), Q_{ij}(x), R_j(x)$ are some given polynomials, with respect to unknown series $T_{v_0}(x), \ldots, T_{v_k}(x)$.

Let us try to solve equations (3). We use the last equation to express $T_{v_k}(x)$ in terms of the rest series,

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

substitute this expression instead of $T_{v_k}(x)$ into the remaining equations, and multiply them by $(1+xP_k(x))$. Thus we obtain equations of the same form, but their number (which is also the number of unknowns) decreases by one. By doing the same for $T_{v_{k-1}}(x)$, $T_{v_{k-2}}(x)$, etc., we obtain at last an expression of $T_{v_0}(x)$ as a quotient of two polynomials. By substituting it into the expression for $T_{v_1}(x)$, we find that this series is also a quotient of two polynomials. In this way we obtain the same for all $T_{v_i}(x)$. It remains to note that $M(x) = \sum_{v\in W} T_v(x)$.

62. For each word v, denote by v^{opp} the word consisting of the same letters in the opposite order. For any set of words M, we write $M^{opp} = \{v^{opp} \mid v \in M\}$. Clearly, $M^{opp}(x) = M(x)$ for any M. If L is a language whose set of forbidden words is B, then by L^{opp} we denote the language whose set of forbidden words is B^{opp}.

Let us return to our problem. It is clear that the set M_w^{opp} consists of all admissible words of L^{opp} which start with a subword equal to w^{opp}. This set is regular (the proof is similar to the solution of problem 59). Hence the series $M_w(x) = M_w^{opp}$ can be represented as a quotient of two polynomials.

The set M_w is also regular, but the proof of this fact would take more place.