Univ. u Novom Sadu
Zb. Rad. Prirod.–Mat. Fak.
Ser. Mat. 23, 1 (1993), 227 - 234

Review of Research
Faculty of Science
Mathematics Series

THE SET OF ALL THE WORDS OF LENGTH n OVER ANY ALPHABET WITH THE FORBIDDEN SUBWORD $aa\dots a$ WHERE THE LETTER a IS FIXED

Rade Doroslovački
Faculty of Engeenering, University of Novi Sad,
Trg Dositeja Obradovića 6, 21000 Novi Sad, Yugoslavia

Abstract

The paper gives a special costruction of those words of length n over any alphabet $A = \{a_1, a_2, \dots, a_m\}$ in which the subword consisiting of k consecutive a's is forbidden, where letter a is fixed from alphabet A. This construction gives the number of all these words. This number of words is counted in two different ways, which gives some new families of combinatorial identities.

AMS Mathematics Subject Classification (1991): 05A15
Key words and phrases: Forbidden subword.

1. Definitions and denotations

Let $A = \{a_1, a_2, \dots, a_m\}$ denote a finite and nonempty set of symbols. A is called an alphabet. By A^n we shall denote the set of all the strings of length n over alphabet A. $A^n = \{u_1u_2\dots u_n | u_1, u_2, \dots, u_n \in A\}$, the only element of A^0 is the empty string, i.e. the string of length 0. The set of all the finite strings over alphabet A is

$$A^* = \bigcup_{i \geq 0} A^i.$$

If S is a set, then $|S|$ is the cardinality of S. By $\lceil x \rceil$ and $\lfloor x \rfloor$ we denote the smallest integer $\geq x$ and the greatest integer $\leq x$, respectively. By $l_i(p)$

we denote the number of i's in the string $p \in A^*$, for $i \in A$, $\binom{n}{k} = 0$ iff $n < k$, $N_n = \{1, 2, \ldots, n\}$ and

$$[x] = \begin{cases} \lfloor x \rfloor & |\lfloor x \rfloor - x| \le 0.5 \\ \lceil x \rceil & |\lceil x \rceil - x| < 0.5 \end{cases}$$

2. Results and discussion

Theorem 1. *If k,m,n are natural numbers and $k \ge 3$, then*

$$|S(k,m,n)| = \sum_{i_{k-1}=0}^{\lceil \frac{(k-1)n}{k} \rceil} \sum_{i_{k-2}=0}^{\lfloor \frac{(k-2)i_{k-1}}{k-1} \rfloor} \cdots \sum_{i_2=0}^{\lfloor \frac{2i_3}{3} \rfloor} \sum_{i_1=0}^{\lfloor \frac{i_2}{2} \rfloor}$$

$$\binom{n - i_{k-1} + 1}{i_{k-1} - i_{k-2}} \binom{i_{k-1} - i_{k-2}}{i_{k-2} - i_{k-3}} \cdots \binom{i_2 - i_1}{i_1} (m-1)^{n - i_{k-1}}$$

where

$$S(k,m,n) = \{u | u = u_1 u_2 \ldots u_n \in A^n,\ (\forall i \in N_{n-k+1})$$
$$(u_i u_{i+1} \ldots u_{i+k-1} \ne \underbrace{aa \ldots a}_{k})\}$$

and a is a fixed letter from A.

Proof. It is easy to see that:

$$|S(1,m,n)| = (m-1)^n$$

and it is known that:

$$|S(2,m,n)| = \sum_{i_1=0}^{\lfloor \frac{n}{2} \rfloor} \binom{n - i_1 + 1}{i_1} (m-1)^{n-i_1}$$

(for $m = 2$ we have Fibonacci numbers).

We shall give the proof by induction on $k (k \ge 3)$ for each $m, n \in N$. We shall first prove that the theorem is valid for $k = 3$ i.e.

$$|S(3,m,n)| = \sum_{i_2=0}^{\lceil \frac{2n}{3} \rceil} \sum_{i_1=0}^{\lfloor \frac{i_2}{2} \rfloor} \binom{n - i_2 + 1}{i_2 - i_1} \binom{i_2 - i_1}{i_1} (m-1)^{n-i_2} \quad m, n \in N$$

We make a partition of the set $S(3,m,n)$ into subsets $S_{i_2}(3,m,n)$, where $S_{i_2}(3,m,n)$ is the set of all those words of length n over alphabet A which contain exactly i_2 a's ($p \in S_{i_2}(3,m,n) \Rightarrow l_a(p) = i_2$) and do not contain the subword aaa i.e.

$$S_{i_2}(3,m,n) = \{p | p = u_1 u_2 \ldots u_n \in A^n,\ l_a(p) = i_2\ , \\ \forall i \in N_{n-2} \quad u_i u_{i+1} u_{i+2} \neq aaa\}$$

Let us construct the words from the set $S_{i_2}(3,m,n)$. We write one of the letters "α" and "λ" of the $i_2 - 1$ places between i_2 a's, that is, we make some words of length $i_2 - 1$ over the alphabet $\{\alpha; \lambda\}$. The letter "λ" denotes the empty letter, i.e. if the letter "λ" is written between two a's, then, actually, nothing is written and letter "α" denotes the empty place for some words from

$$\bigcup_{i \in N} (A \setminus \{a\})^i.$$

Since the subword aaa is forbidden in words of the set $S_{i_2}(3,m,n)$, it follows that the set F of words of length $i_2 - 1$ over the alphabet $\{\alpha; \lambda\}$ must satisfy the property that the subword $\lambda\lambda$ is forbidden in the words of F. Consequently,

$$(1) |F| = |S(2,2,i_2-1)| = \sum_{i_1=0}^{\lceil \frac{i_2-1}{2} \rceil} \binom{i_2 - 1 - i_1 + 1}{i_1} = \sum_{i_1=0}^{\lfloor \frac{i_2}{2} \rfloor} \binom{i_2 - i_1}{i_1},$$

where i_1 denotes the number of appearances of letter λ in words from set F. If β is any letter from $A \setminus \{a\}$, then we must first write exactly one letter β on $i_2 - 1 - i_1$ places which determined by letter α. There remains to write $n - i_2 - (i_2 - 1 - i_1) = n - 2i_2 + i_1 + 1$ letters β on $i_2 - 1 - i_1$ regions which already contain one β each, as well as into the regions in front of and behind the word, that is into $i_2 - 1 - i_1 + 2 = i_2 - i_1 + 1$ regions in all. We make this arrangement of $n - 2i_2 + i_1 + 1$ β's into $i_2 - i_1 + 1$ regions by putting $i_2 - i_1$ compartments among these β's. The number of permutations of these compartments and β's equals the number of arrangements of these β's into these regions, that is,

$$(2) \qquad \binom{n - i_2 + 1}{i_2 - i_1}$$

Now, we must substitute letter β with some letters from $A \setminus \{a\}$. It can be done in

$$(3) \qquad (m-1)^{n-i_1}$$

different ways.
Thus from (1), (2) and (3) follows

$$|S(3,m,n) = \sum_{i_2=0}^{\lceil\frac{2n}{3}\rceil} |F| \binom{n-i_2+1}{i_2-i_1} (m-1)^{n-i_2}$$
$$= \sum_{i_2=0}^{\lceil\frac{2n}{3}\rceil} \sum_{i_1=0}^{\lfloor\frac{i_2}{2}\rfloor} \binom{n-i_2+1}{i_2-i_1} \binom{i_2-i_1}{i_1} (m-1)^{n-i_2}$$

because

$$i_2' \neq i_2'' \Rightarrow S_{i_2'}(3,m,n) \cap S_{i_2''}(3,m,n) = \emptyset$$
$$S(3,m,n) = \bigcup_{i_2 \geq 0} S_{i_2}(3,m,n)$$
$$i_2 > \lceil\frac{2n}{3}\rceil \Rightarrow S_{i_2}(3,m,n) = \emptyset.$$

Let us assume that the assertion is valid for some $k \geq 3$. We shall prove, then, that it is also valid for $k+1$. We make a partition of the set $S(k+1,m,n)$ into the subsets

$$S_{i_k}(k+1,m,n) = \{u| \ u = u_1u_2\ldots u_n \in A^n, (\forall \in N_{n-k+1}) \ (u_iu_{i+1}\ldots u_{i+k-1} \neq \underbrace{aa\ldots a}_{k+1}), \ \ l_a(u) = i_k\}$$

The set $S_{i_k}(k+1,m,n)$ is the set of all those words of length n over alphabet A, which contain exactly i_k a's and which do not contain the subword $\underbrace{aa\ldots a}_{k+1}$. We proceed with the construction of the set $S_{i_k}(k+1,m,n)$.We write one of the letters "α" or "λ" into each of the $i_k - 1$ places between i_k a's, that is, we make the words of length $i_k - 1$ over the alphabet $\{\alpha, \lambda\}$. The leter "λ" denotes the empty letter, i.e., if letter λ is written between two a's, then the effect is the same as if nothing is written and letter "α" denotes empty places for some word from

$$\bigcup_{i \in N} (A \setminus \{a\})^i.$$

Since the subword $\underbrace{aa\ldots a}_{k+1}$ is forbidden in the words of the set $S_{i_{k+1}}(k+1,m,n)$, it follows that the set of E of words of length $i_k - 1$ over the alphabet

$\{\alpha, \lambda\}$ must satisfy the property that the subword $\underbrace{\lambda\lambda\ldots\lambda}_{k}$ is forbidden in the words of E. It is obvious, on the basis of the inductive hypothesis, that

$$|E| = |S(k,2,i_k-1)| = \sum_{i_{k-1}=0}^{\lceil\frac{(k-1)(i_k-1)}{k}\rceil} \sum_{i_{k-2}=0}^{\lfloor\frac{(k-2)i_{k-1}}{k-1}\rfloor} \cdots \sum_{i_2=0}^{\lfloor\frac{2i_3}{3}\rfloor} \sum_{i_1=0}^{\lfloor\frac{i_2}{2}\rfloor}$$

$$\binom{i_k-1-i_{k-1}+1}{i_{k-1}-i_{k-2}}\binom{i_{k-1}-i_{k-2}}{i_{k-2}-i_{k-3}}\cdots\binom{i_2-i_1}{i_1} \text{ i.e.}$$

$$(4) \qquad |E| = \sum_{i_{k-1}=0}^{\lfloor\frac{(k-1)i_k}{k}\rfloor} \sum_{i_{k-2}=0}^{\lfloor\frac{(k-2)i_{k-1}}{k-1}\rfloor} \cdots \sum_{i_2=0}^{\lfloor\frac{2i_3}{3}\rfloor} \sum_{i_1=0}^{\lfloor\frac{i_2}{2}\rfloor}$$

$$\binom{i_k-i_{k-1}}{i_{k-1}-i_{k-2}}\binom{i_{k-1}-i_{k-2}}{i_{k-2}-i_{k-3}}\cdots\binom{i_2-i_1}{i_1}$$

where i_{k-1} denotes the total number of appearances of letter λ in the words of the set E. If β is any letter from A, then we must first write exactly one letter β on $i_k - 1 - i_{k-1}$ places, which are determined by letters α. There remains to write $n - i_k - (i_k - 1 - i_{k-1}) = n - 2i_k + i_{k-1} + 1$ letters β on $i_k - 1 - i_{k-1}$ regions which already contain one β each, as well as into the regions in front of and behind the word, that is into $i_k - 1 - i_{k-1} + 2 = i_k - i_{k-1} + 1$ regions in all. We make this arrangement of $n - 2i_k + i_{k-1} + 1$ β's into $i_k - i_{k-1} + 1$ regions by putting $i_k - i_{k-1}$ compartments among these β's. The number of permutations of these compartments and β's equals the number of arrangements of these β's into these regions, that is,

$$(5) \qquad \binom{n-i_k+1}{i_k-i_{k-1}}.$$

Now we must substitute letter β with some letters from $A \setminus \{a\}$. This can be done in

$$(6) \qquad (m-1)^{n-i_k}$$

different ways.
From (4), (5) and (6) we have

$$|S(k+1,m,n)| = \sum_{i_k=0}^{\lceil\frac{kn}{k+1}\rceil} |S_{i_k}(k+1,m,n)|$$

$$= \sum_{i_k=0}^{\lceil \frac{kn}{k+1} \rceil} |E| \binom{n-i_k+1}{i_k-i_{k-1}} (m-1)^{n-i_k+1}$$

$$= \sum_{i_k=0}^{\lceil \frac{kn}{k+1} \rceil} \sum_{i_{k-1}=0}^{\lfloor \frac{(k-1)i_k}{k} \rfloor} \cdots \sum_{i_2=0}^{\lfloor \frac{2i_3}{3} \rfloor} \sum_{i_1=0}^{\lfloor \frac{i_2}{2} \rfloor} \binom{n-i_k+1}{i_k-i_{k-1}} \cdot$$

$$\binom{i_k-i_{k-1}}{i_{k-1}-i_{k-2}} \cdots \binom{i_2-i_1}{i_1} (m-1)^{n-i_k}$$

(because

$$i'_k \neq i''_k \Rightarrow S_{i'_k}(k+1,m,n) \bigcap S_{i''_k}(k+1,m,n) = \emptyset$$

$$S(k+1,m,n) = \cup_{i_k \geq 0} S_{i_k}(k+1,m,n)$$

$$i_k > \lceil \frac{kn}{k+1} \rceil \Rightarrow S_{i_k}(k+1,m,n) = \emptyset)$$

The proof is completed. On the other hand, the recursive formula for $|S(k,m,n)|$ is

$$|S(k,m,n)| = (m-1) \sum_{i=n-k}^{n-1} |S(k,m,i)|$$

whose explicite formula is

$$|S(k,m,n)| = C_1 x_1^n + C_2 x_2^n + \cdots + C_k x_k^n,$$

where x_i are roots of equation

$$x^k - (m-1)x^{k-1} - (m-1)x^{k-2} - \cdots - (m-1)x - (m-1) = 0$$

for $i = 1, 2, \ldots, k$. The constants $C_1, C_2, \ldots, C_k$ are determined from the initial conditions. Now, we obtain the following identites:

$$|S(2,m,n)| = \left[\frac{m+1+\sqrt{m^2+2m-3}}{2\sqrt{m^2+2m-3}} \left(\frac{m-1+\sqrt{m^2+2m-3}}{2} \right)^n \right]$$

$$= \sum_{i_1=0}^{\lceil \frac{n}{2} \rceil} \binom{n-i_1+1}{i_1} (m-1)^{n-i_1}$$

$$|S(3,2,n)| = \left[\frac{\alpha^{n+3}}{3\alpha^2 - 2\alpha - 1}\right] = (1,1\ldots)(1,8\ldots)^n$$
$$= \sum_{i_2=0}^{\lceil \frac{2n}{3} \rceil} \sum_{i_1=0}^{\lfloor \frac{i_2}{2} \rfloor} \binom{n - i_2 + 1}{i_2 - i_1} \binom{i_2 - i_1}{i_1},$$

where α is a real root of $x^3 - x^2 - x - 1 = 0$, i.e. $\alpha = \frac{1}{3}(1 + \sqrt[3]{19 + 3\sqrt{33}} + \sqrt[3]{19 - 3\sqrt{33}})$

$$|S(3,3,n)| = \left[\frac{\gamma^{n+3}}{3\gamma^2 - 8\gamma - 4}\right] = [(1,046\ldots)(2,919\ldots)^n]$$
$$= \sum_{i_2=0}^{\lceil \frac{2n}{3} \rceil} \sum_{i_1=0}^{\lfloor \frac{i_2}{2} \rfloor} \binom{n - i_2 + 1}{i_2 - i_1} \binom{i_2 - i_1}{i_1} 2^{n-i_2},$$

where γ is a real root of $x^3 - 2x^2 - 2x - 2 = 0$, i.e. $\gamma = \frac{1}{3}(2 + \sqrt[3]{53 + \sqrt{1809}} + \sqrt[3]{53 - \sqrt{1809}})$

$$|S(3,m,n)| = \frac{\alpha^{n+3}}{3(m-1)\alpha^2 - 2\alpha(m-1)^2 - (m-1)^2} + 2\mathrm{Re}\frac{\beta^{n+3}}{3(m-1)^2\beta^2 - 2\beta(m-1)^2 - (m-1)^2},$$

where α is a real root and β is a nonreal root of $x^3 - (m-1)x^2 - (m-1)x - (m-1) = 0$ and many other identities.

References

[1] Cvetković, D., The generating function for variations with restrictions and paths of the graph and self complementary graphs, Univ. Beograd, Publ. Elektrotehnički fakultet, serija Mat. Fiz. 320–328 (1970), 27–34.

[2] Einb, J. M., The enumeration of bit sequences that satisfy local criteria, Publ. Inst. Math. (Beograd), 27(41)(1980), 51–56.

[3] Doroslovački, R., The set of all words over alphabet $\{0,1\}$ of lenght n with the forbidden subword $11\ldots1$, Univ. u Novom Sadu Zb. Rad. Prirod. - Mat. Fak. Ser. Mat. 14,2 (1984), 167–173.

[4] Austin, R., Guy, R., Binary sequences without isolated ones, The Fibonacci Quarterly, 16,1 (1978), 84–86.

REZIME

SKUP SVIH REČI DUŽINE n NAD PROIZVOLJNOM AZBUKOM SA ZABRANJENOM PODREČI $aa\dots a$ GDE JE SLOVO a FIKSNO

Rad daje specijalnu konstrukciju ovih reči dužine n nad azbukom $A = \{a_1, a_2, \dots, a_m\}$ u kojima je zabranjena podreč koja se sastoji iz k uzastopnih slova a, gde je a fiksno slovo iz azbuke A. Ova konstrukcija daje broj svih ovih reči. Broj reči je dobijen na dva različita načina, što daje neke nove familije kombinatornih identiteta.

Received by the editors May 18, 1989