Univ. u Novom Sadu
Zb. Rad. Prirod.–Mat. Fak.
Ser. Mat. 23, 2 (1993), 239 -244

Review of Research
Faculty of Science
Mathematics Series

THE SET OF ALL THE WORDS OF LENGTH n OVER ANY ALPHABET WITH A FORBIDDEN GOOD SUBWORD

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

Abstract

In the paper the set of all words of length n is constructed and enumerated, over any alphabet A with a forbidden fixed subword $a_1a_2...a_k$ which is a good word i.e. $a_1a_2...a_s \neq a_{k-s+1}a_{k-s+2}...a_k$ for each natural number $s < k$. This number of words is counted in two different ways, which gives new combinatorial identities.

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

1. Definitions and notation

The set of first m natural numbers is denoted by N_m. Hence $N_m = \{1, 2, ..., m\}$. The set $A = \{\alpha_1, \alpha_2, ...\alpha_m\}$ is called an alphabet if for all $i \in N_m$, α_i is an arbitrary symbols. The elements of A are then the letters of the alphabet and A is a $k-$letter alphabet.

If $x \in A^n$ i.e. if $x = (x_1, x_2, ...x_n)$ is an ordered $n-$tuple with components from A, we say that x is a word of length n over the alphabet A. For the sake of brevity we shall write $(x_1, x_2, ...x_n)$ as $x_1x_2...x_n$.

A subword of length k of the word $x_1x_2...x_n$ is any word $x_sx_{s+1}...x_{s+k-1}$ where $s \in N_{n-k+1}$ and $k \in N_n$.

A subword (word) $y_1y_2...y_k$ is good iff $y_1y_2...y_s \neq y_{k-s+1}y_{k-s+2}...y_k$ for each natural number $s < k$.

The only element of X^0 is the empty string i.e. the string of length 0.

The set of all the words of finite length over the alphabet A will be denoted by A^* i.e.

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

The number of ways in which a subword y occurs in a word $x \in A^*$ is denoted by $l_y(x)$. In particular, the number of ways in which the letter $\alpha \in A$ occurs in the word $x \in A^*$ is $l_\alpha(x)$.

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

$$[z] = \begin{cases} \lfloor z \rfloor & \text{if } |\lfloor z \rfloor - z| \leq 0,5 \\ \lceil z \rceil & \text{if } |\lceil z \rceil - z| < 0,5 \end{cases}$$

2. Results and discussion

Theorem 1.

$$L_r(k,m,n) = A_r(k,m,n) = (-1)^r \sum_{i \equiv r}^{\lfloor \frac{n}{k} \rfloor} (-1)^i \binom{n - ki + i}{i} m^{n-ki}$$

where $A_r(k,m,n) = \{x | x = x_1x_2...x_n \in A^n,\ \forall i \in N_{n-k+1}\ x_ix_{i+1}...x_{i+k-1} \neq p,\ l_p(x) \geq r\}$ *and* $p = a_1a_2 \ldots a_k$ *is a fixed good subword of length* k *of the word* x.

Proof. The set $A_r(k,m,n)$ is the set of all the words of length n over the alphabet $A = \{\alpha_1, \alpha_2, \ldots \alpha_m\}$ in which the number of appearances of the good subword p of length n, is at least r.

We shall give the proof by induction on $r = q, q-1, \ldots 1$ where $q = \lfloor \frac{n}{k} \rfloor$. We shall first prove that the theorem is valid for $r = q$ i.e.

$$|A_q(k,m,n)| = \binom{n - kq + q}{q} m^{n-kq}.$$

Since the subword p is good and $q = \lfloor \frac{n}{k} \rfloor$ where k is the length of the subword p, we can write only q subwords p into the n empty places of the word x. It can be done in

$$\frac{(n-kq+q)!}{(n-kq)!q!} = \binom{n-kq+q}{q}$$

different ways, because we permute these q subwords and $n-kq$ empty places. Now we can write the letters from alphabet A on $n-kq$ empty places in an arbitrary way, that is, we can make variations of m elements of class $n-kq$. There are m^{n-kq} such variations, and now it is obvious that

$$L_q(k,m,n) = |A_q(k,m,n)| = \binom{n-kq+q}{q} m^{n-kq}$$

because all of the words obtained in this way are different. Notice, that all the words obtained in the same way for $r = q-1$ are not different and because of that

$$\begin{aligned} L_{q-1}(k,m,n) &= \binom{n-k(q-1)+q-1}{q-1} m^{n-k(q-1)} - \\ &- \binom{n-kq+q}{q} m^{n-kq}. \end{aligned}$$

Let us assume that the assertion is valid for r. We shall prove, then, that it is also valid for $r-1$. Now,we begin constructing an arbitrary word from the set $A_{r-1}(k,m,n)$. First we write $r-1$ subwords p into the n empty places of the word x. It can be done in

$$\binom{n-k(r-1)+r-1}{r-1}$$

different ways, because we permute these $r-1$ subwords p and $n-k(r-1)$ empty places. Then there still remains to write $n-k(r-1)$ letters from the m -letter alphabet. It can be done in $m^{n-k(r-1)}$ different ways. In this way we obtain

$$\binom{n-k(r-1)+r-1}{r-1} m^{n-k(r-1)}$$

words, but not all of them are different. It is easy to see that some words are repeated and there are $L_r(k,m,n)$ repeated ones and because of that and on the basis of the inductive hypothesis it follows that

$$L_{r-1}(k,m,n) = \binom{n-k(r-1)+r-1}{r-1} m^{n-k(r-1)} - L_r(k,m,n) =$$

$$= \binom{n-k(r-1)+r-1}{r-1} m^{n-k(r-1)} -$$

$$-(-1)^r \sum_{i\equiv r}^{\lfloor \frac{n}{k} \rfloor} (-1)^i \binom{n-ki+i}{i} m^{n-ki}$$

$$= (-1)^{r-1} \sum_{i\equiv r-1}^{\lfloor \frac{n}{k} \rfloor} (-1)^i \binom{n-ki+i}{i} m^{n-ki} \qquad \square$$

Theorem 2.

$$L(k,m,n) = |A(k,m,n)| = \sum_{i\equiv 0}^{\lfloor \frac{n}{k} \rfloor} (-1)^i \binom{n-ki+i}{i} m^{n-ki}$$

where $A(k,m,n) = \{x | x = x_1 x_2 \dots x_n \in A^n,\ \forall i \in N_{n-k+1}\ x_i x_{i+1} \dots x_{i+k-1} \neq p\}$ *and* p *is a fixed good subword of the word* x.

Proof. The set $A(k,m,n)$ is the set of all the words of length n over the alphabet $A = \{\alpha_1, \alpha_2, \dots \alpha_m\}$ in which the forbidden good subword is p of length k.

It is easy to see that

$$L(k,m,n) = m^n - L_1(k,m,n).$$

Now, using Theorem 1 we have

$$L(k,m,n) = m^n - (-1)^1 \sum_{i\equiv 1}^{\lfloor \frac{n}{k} \rfloor} (-1)^i \binom{n-ki+i}{i} m^{n-ki} \text{ i.e.}$$

$$L(k,m,n) = \sum_{i\equiv 0}^{\lfloor \frac{n}{k} \rfloor} (-1)^i \binom{n-ki+i}{i} m^{n-ki}. \quad \square$$

On the other hand, we can make recursive relations for $L(k,m,n)$ for some k and m and find its explicit formula. For example:

$$L(1,m,n) = \sum_{i\equiv 0}^{n} (-1)^i \binom{n}{i} m^{n-i} = (m-1)^n$$

$$L(2,2,n)=\sum_{i\equiv 0}^{\lfloor\frac{n}{2}\rfloor}(-1)^i\binom{n-i}{i}2^{n-2i}=n+1$$

$$L(2,3,n)=\sum_{i\equiv 0}^{\lfloor\frac{n}{2}\rfloor}(-1)^i\binom{n-i}{i}3^{n-2i}=\left[\frac{5+3\sqrt{10}}{10}\left(\frac{3+5\sqrt{5}}{2}\right)^n\right]$$

$$L(2,4,n)=\sum_{i\equiv 0}^{\lfloor\frac{n}{2}\rfloor}(-1)^i\binom{n-i}{i}4^{n-2i}=\left[\frac{3+2\sqrt{3}}{6}(2+\sqrt{3})^n\right]$$

$$L(2,m,n)=\sum_{i\equiv 0}^{\lfloor\frac{n}{2}\rfloor}(-1)^i\binom{n-i}{i}m^{n-2i}=C_1X_1^n+C_2X_2^n$$

where $x_{1,2}=\frac{m+\sqrt{m^2-4}}{2}$ and C_1, C_2 are determined from the initial conditions.

$$L(3,2,n)=\sum_{i\equiv 0}^{\lfloor\frac{n}{3}\rfloor}(-1)^i\binom{n-2i}{i}2^{n-3i}=-1+\left[\frac{5+2\sqrt{5}}{5}\left(\frac{1+\sqrt{5}}{2}\right)^n\right]$$

$$L(3,3,n)=\sum_{i\equiv 0}^{\lfloor\frac{n}{3}\rfloor}(-1)^i\binom{n-2i}{i}3^{n-3i}=\left[\frac{9\alpha^2-\alpha-3}{\alpha^3-2}\alpha^{n-1}\right]$$

$$\alpha=1+2\cos\frac{\pi}{9}$$

References

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

[2] Cvetković, D., The generating function for variations with restrictions and paths of the graph and self-complementary graphs, Univ. Beog., Publ. Electrotehnical Faculty, serija Mat. (1970), 320-328.

[3] Doroslovački, R., The set of all the words over the alphabet {0.1} of length n with the forbidden subword 11...1, Univ. u Novom Sadu Zb. Rad. Prirod. - Mat. Fak. Ser. Mat. 14,2 (1984), 167-173.

[4] Doroslovački, R., The set of all the words over any alphabet with the forbidden subword *aa...a* where letter *a* is fixed, Univ. u Novom Sadu Zb. Rad. Prirod. - Mat. Fak. Ser. Mat. 23,1 (1993), 227 - 234.

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

REZIME

SKUP SVIH REČI DUŽINE n NAD PROIZVOLJNOM AZBUKOM SA ZABRANJENOM DOBROM PODREČI

U radu je konstruisan i prebrojan skup svih reči dužine n nad azbukom A sa zabranjenom fiksnom podreči $a_1a_2\ldots a_k$, koja je "dobra podreč" tj. $a_1a_2\ldots a_s \neq a_{k-s+1}a_{k-s+2}\ldots a_k$ za sve prirodne brojeve s manje od k. Ovaj skup reči je prebrojan na dva različita načina, što daje nove kombinatorne identitete.

Received by the editors July 25, 1990