ZBORNIK RADOVA
Prirodno-matematičkog fakulteta
Univerziteta u Novom Sadu
Serija za matematiku, 14,2 (1984)

REVIEW OF RESEARCH
Faculty of Science
University of Novi Sad
Mathematics Series, 14, 2 (1984)

THE SET OF ALL WORDS OVER ALPHABET {0,1} OF LENGTH n WITH THE FORBIDDEN SUBWORD $\underbrace{11\ldots1}_{k}$

Doroslovački Rade
Fakultet tehničkih nauka. Institut za primenjene osnovne discipline, 21000 Novi Sad, V. Vlahovića 3, Jugoslavija

ABSTRACT

We enumerate the number of those words of length n over the alphabet {0,1} in which the subword consisting of k consecutive 1's is forbidden. This number of words is counted in two different ways, which gives some new combinatorical identities.

1. DEFINITIONS AND DENOTATIONS

Let X denote a finite and nonempty set of symbols, X is called an alphabet. By X^n we shall denote the set of all strings of the length n over the alphabet X, i.e. $X^n = \{x_1x_2\ldots x_n | x_1, x_2,\ldots,x_n \in X\}$, the only element of X^0 is the empty string, i.e. the string of lenght 0. The set of all finite strings over the alphabet X is $X^* = \bigcup_{i\geq 0} X^i$. If S is a set, then $|S|$ is the cardinality of S. By $\lceil n \rceil$ and $\lfloor n \rfloor$ we denote the smallest integer $\geq n$ and the greatest integer $\leq n$, respectively. By $\ell_i(a)$ we denote the number of i's in the string $a \in X^*$ for $i \in X$, $\binom{n}{k} = 0$ iff $n < k$, $N_n = \{1,2,\ldots,n\}$ and

---

AMS Mathematics Subject Classification (1980): Primary 05A15.
Key words and phrases: Word, forbidden subword.

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

2. RESULTS AND DISCUSSION

THEOREM.

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

$$\binom{n-i_{k-1}+1}{i_{k-1}-i_{k-2}}\binom{i_{k-1}-i_{k-2}}{i_{k-2}-i_{k-3}} \cdots \binom{i_3-i_2}{i_2-i_1}\binom{i_2-i_1}{i_1}, \; k \geq 3$$

where

$$A_k(n) = \{x \mid x = x_1x_2 \ldots x_n \in \{0,1\}^n,$$

$$\forall i \in N_{n-k+1} x_i x_{i+1} \ldots x_{i+k-1} \neq \underbrace{11\ldots1}_{k}\} .$$

PROOF: We shall introduce the denotation $L_k(n)=|A_k(n)|$. Obviously $L_1(n) = 1$. It is known that:

$$L_2(n) = \sum_{i_1=0}^{\lceil n/2 \rceil} \binom{n-i_1+1}{i_1} = \left[\frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^{n+2}\right]$$

(Fibonacci numbers).

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

$$L_3(n) = \sum_{i_2=0}^{\lceil 2n/3 \rceil} \sum_{i_1=0}^{\lfloor i_2/2 \rfloor} \binom{n-i_2+1}{i_2-i_1}\binom{i_2-i_1}{i_1} .$$

We make a partition of the set $A_3(n)$ into subsets $A_3^{i_2}(n)$, where

$A_3^{i_2}(n)$ is the set of all those words of length n over the alphabet {0,1} which contain exactly i_2 1's each, and do not contain the subword 111 i.e.

$$A_3^{i_2}(n) = \{x \mid x = x_1x_2 \ldots x_n \in \{0,1\}^n,$$

$$\forall i \in N_{n-2}\, x_ix_{i+1}x_{i+2} \neq 111,\ \ell_1(x) = i_2\} .$$

Let us construct the words from the set $A_3^{i_2}(n)$. We write one of the letters "0" or "λ" in each of the i_2-1 places between i_2 1's, that is, we make some words of lenght i_2-1 over the alphabet {0,λ}. The letter "λ" denotes the empty letter, i.e. if the letter "λ" is written between two 1's, then, actually, nothing is written. Since the subword 111 is forbidden in the words of the set $A_3^{i_2}(n)$, it follows that the set Q of the words of length i_2-1 over the alphabet {0,λ} must satisfy the property that the subword λλ is forbidden in the words of Q. Consequently

$$|Q| = |A_2(i_2-1)| = \sum_{i_1=0}^{\lceil \frac{i_2-1}{2} \rceil} \binom{i_2-1-i_1+1}{i_1} = \sum_{i_1=0}^{\lfloor \frac{i_2}{2} \rfloor} \binom{i_2-i_1}{i_1}$$

where i_1 denotes the number of the letter λ in the word from the set Q. Since i_2-1-i_1 o's is already written between these i_2 1's, there remains to write $n-i_2-(i_2-1-i_1) = n-2i_2+i_1+1$ o's. These o's may be written into some of the i_2-1-i_1 regions which already contain one zero 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 $n-2i_2+i_1+1$ o's into i_2-i_1+1 regions by putting i_2-i_1 compartments among these o's. The number of permutations of these compartments and o's equals the number of arrangements of these o's into these regions, that is,

$$\binom{n-i_2+1}{i_2-i_1},$$

Thus

$$|A_3^{i_2}(n)| = \sum_{i_1=0}^{\lfloor \frac{i_2}{2} \rfloor} \binom{n-i_2+1}{i_2-i_1}\binom{i_2-i_1}{i_1} .$$

It is obvious from the definition of these sets $A_3^{i_2}(n)$ that

$$A_3(n) = \bigcup_{i_2>0} A_3^{i_2}(n), \qquad i_2' \neq i_2'' \Rightarrow A_3^{i_2'}(n) \cap A_3^{i_2''} = \emptyset$$

and

$$i_2 > \left\lceil \frac{2n}{3} \right\rceil \Rightarrow A_3^{i_2}(n) = \emptyset .$$

Thus

$$|A_3(n)| = \sum_{i_2=0}^{\lceil \frac{2n}{3} \rceil} |A_3^{i_2}(n)| = \sum_{i_2=0}^{\lceil \frac{2n}{3} \rceil} \sum_{i_1=0}^{\lfloor \frac{i_2}{2} \rfloor} \binom{n-i_2+1}{i_2-i_1}\binom{i_2-i_1}{i_1} .$$

Let us assume that the assertion is valid for k. We shall prove, then, that it is also valid for k+1. We make a partition of the set $A_{k+1}(n)$ into the subsets

$$A_{k+1}^{i_k}(n) = \{x \mid x = x_1x_2 \ldots x_n \in \{0,1\}^n,$$

$$\forall i \in N_{n-k+1} x_i x_{i+1} \ldots x_{i+k-1} \neq \underbrace{11\ldots1}_{k+1}, \ell_1(x)=i_k\}.$$

The set $A_{k+1}^{i_k}(n)$ is the set of all those words of length n over the alphabet {0,1}, which contain exactly i_k 1's and which do not contain the subword $\underbrace{11\ldots1}_{k+1}$. We proceed with the construction of the set $A_{k+1}^{i_k}(n)$. We write one of the letters "0" or "λ", into each of the i_k-1 places between i_k 1's, that is, we make the words of length i_k-1 over the alphabet $\{0,\lambda\}$. The letter λ denotes the empty letter, i.e., if the letter λ is written between two 1's, then the effect is the same as if nothing is written. Since the word $\underbrace{11\ldots1}_{k+1}$ is forbidden in the words of the set $A_{k+1}^{i_k}(n)$, it follows that the subword $\underbrace{\lambda\lambda\ldots\lambda}_{k}$ is forbidden in the set of words that we are making over the alphabet $\{0,\lambda\}$.

If we denote this set of words of length i_k-1 over the alphabet $\{0,\lambda\}$, which do not contain the subword $\underbrace{\lambda\lambda\ldots\lambda}_{k}$ by Q_1, then it is obvious, on the basis of the inductive hypothesis, that

$$|Q_1| = |A_k(i_k-1)| = \sum_{i_{k-1}=0}^{\left\lceil\frac{(k-1)(i_k-1)}{k}\right\rceil} \sum_{i_{k-2}=0}^{\left\lfloor\frac{(k-2)i_{k-1}}{k-1}\right\rfloor} \cdots \sum_{i_2=0}^{\left\lfloor\frac{2i_3}{3}\right\rfloor} \sum_{i_1=0}^{\left\lfloor\frac{i_2}{2}\right\rfloor} \binom{i_k-1-i_{k-1}+1}{i_{k-1}-i_{k-2}}\binom{i_{k-1}-i_{k-2}}{i_{k-2}-i_{k-3}} \cdots \binom{i_3-i_2}{i_2-i_1}\binom{i_2-i_1}{i_1}$$

i.e.

$$|Q_1| = \sum_{i_{k-1}=0}^{\left\lfloor\frac{(k-1)i_k}{k}\right\rfloor} \sum_{i_{k-2}=0}^{\left\lfloor\frac{(k-2)i_{k-1}}{k-1}\right\rfloor} \cdots \sum_{i_2=0}^{\left\lfloor\frac{2i_3}{3}\right\rfloor} \sum_{i_1=0}^{\left\lfloor\frac{i_2}{2}\right\rfloor} \binom{i_k-i_{k-1}}{i_{k-1}-i_{k-2}}\binom{i_{k-1}-i_{k-2}}{i_{k-2}-i_{k-3}} \cdots \binom{i_3-i_2}{i_2-i_1}\binom{i_2-i_1}{i_1},$$

where i_{k-1} denotes the total number of appearances of the letter λ in the words of the set Q_1. Since i_k-1-i_{k-1} 0's is already written among these i_k 1's, there remains still to write $n-i_k-(i_k-1-i_{k-1}) = n-2i_k+i_{k-1}+1$ 0's. These 0's can be arbitrarily written into the i_k-1-i_{k-1} regions which already contain one zero 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. This arrangement of $n-2i_k+i_{k-1}+1$ 0's into $i_k-i_{k-1}+1$ regions can be done in

$$\binom{n-i_k+1}{i_k-i_{k-1}} \quad \text{different ways.}$$

Thus

$$|A_{k+1}^{i_k}(n)| = \binom{n-i_k+1}{i_k-i_{k-1}}\cdot|Q_1|.$$

Since the definition of the sets $A_{k+1}^{i_k}(n)$ and $A_{k+1}(n)$ obviously implies

$$A_{k+1}(n) = \bigcup_{i_k \geq 0} A_k^{i_k}(n), \quad i_k' \neq i_k'' \Rightarrow A_{k+1}^{i_k'}(n) \cap A_{k+1}^{i_k''} = \emptyset$$

and

$$i_k > \left\lceil \frac{kn}{k+1} \right\rceil \Rightarrow A_{k+1}^{i_k}(n) = \emptyset,$$

it follows

$$|A_{k+1}(n)| = \sum_{i_k=0}^{\left\lceil \frac{kn}{k+1} \right\rceil} |A_{k+1}^{i_k}(n)| = \sum_{i_k=0}^{\left\lceil \frac{kn}{k+1} \right\rceil} \binom{n-i_k+1}{i_k-i_{k-1}} |Q_1|,$$

which completes the proof.

Since

$$L_1(n) = 1 = 2^0 \cdot 1^n$$

$$L_2(n) = \left[\frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2} \right)^{n+2} \right] = [(1,17\ldots)(1,6\ldots)^n].$$

It can be counted that

$$L_3(n) = \left[\frac{\alpha^{n+3}}{3\alpha^2-2\alpha-1} \right] = [(1,1\ldots)(1,8\ldots)^n]$$

where α is the real root of the equation $x^3-x^2-x-1 = 0$ i.e.

$$\alpha = \frac{1}{3}(1 + \sqrt[3]{19 + 3\sqrt{33}} + \sqrt[3]{19-3\sqrt{33}}).$$

Also

$$L_k(n) = C_1\alpha_1^n + C_2\alpha_2^n + \ldots + C_k\alpha_k^n$$

where α_i are the roots of the equation $x^k-x^{k-1}- \ldots - x^2 - x - 1 = 0$ for each $i = 1,2,\ldots,k$. The constants $C_1,\ldots,C_k$ are determined from the initial conditions.

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. No 320 - No 328 (1970), 27 - 34.*

[2] Einbu, J.M., *The enumeration of bit sequences that satisfy local criteria, Publications de l'Institut Mathématique Beograd, tome 27(41) (1980) pp. 51-56.*

Received by the editors November 6,1984.

REZIME

SKUP SVIH REČI NAD AZBUKOM {0,1} DUŽINE n SA ZABRANJENOM PODREČI $\underbrace{11\ldots1}_{k}$

U ovom radu se izračunava broj svih reči dužine n nad azbukom {0,1}, u kojima je zabranjena podreč k uzastopnih jedinica. Ovaj broj se dobija na dva različita načina što daje neke nove kombinatorne identitete.