Univ. u Novom Sadu
Zb. Rad. Prirod.–Mat. Fak.
Ser. Mat. 25, 2 (1995), 111-115

Review of Research
Faculty of Science
Mathematics Series

THE SET OF ALL THE WORDS OF LENGTH n OVER ALPHABET $\{0, 1\}$ WITH ANY FORBIDDEN SUBWORD OF LENGTH THREE

Doroslovački Rade
Department of Mathematics, Faculty of Engineering
University of Novi Sad, Trg Dositeja Obradovića 6, Yugoslavia.

Abstract

The set of all words of length n over the alphabet $\{0, 1\}$ with a fixed forbidden subword of length 3 is enumerated and constructed. The number of words is counted in two different ways, which gives some new combinatorial identities.

AMS *Mathematics Subject Classification:* 05A15
Key words and Phrases: word, subword.

1. Definitions and notations

Let $X = \{0, 1\}$ be an alphabet. The elements of X are the letters of the alphabet and X is a 2-letter alphabet.

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

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 largest integer $\leq x$. By $\ell_0(p)$ and $\ell_1(p)$ we denote the number of zeros and ones, respectively, in the string $p \in X^*$, where X^* is the set of all finite strings over the alphabet X i.e.

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

$N_n = \{1,2,\ldots,n\}$, $N_n = \emptyset$ iff $n \le 0$, the binomial coeficient $\binom{n}{k} = 0$ iff $n < k$ and

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

i.e. [x] is the nearest integer to x.

2 . Results and discussion

There are 8 cases for the forbidden subword over the alphabet $\{0,1\}$ of length 3: 000, 111, 010, 101, 100, 001, 011 and 110. The cases 000 and 111 are obviously equivalent. In [3] we have

Theorem 1.

$$|A_n| = \sum_{i_2=0}^{\lceil \frac{2n}{3} \rceil} \sum_{i_1=0}^{\lfloor \frac{i_2}{2} \rfloor} \binom{n-i_2+1}{i_2-i_1}\binom{i_2-i_1}{i_1} = \left[\frac{\alpha^{n+3}}{3\alpha^2-2\alpha-1}\right], \quad where$$

$$\alpha = \frac{1}{3}\left(1+\sqrt[3]{19+3\sqrt{33}}+\sqrt[3]{19-3\sqrt{33}}\right) \quad and$$

$$A_n = \{\mathbf{x_n} | \mathbf{x_n} = x_1, x_2 \ldots x_n \in X^n \wedge (\forall i \in N_{n-2})(x_i x_{i+1} x_{i+2} \ne 111)\}.$$

The set A_n is the set of all words of length n with forbidden subword 111. Cases 010, 101 are equivalent, too. In these cases we have

Theorem 2.

$$b_n = |B_n| = \sum_{i=0}^{n} |B_n^i| = 1 + \sum_{i=1}^{n}\sum_{j=0}^{i-1}\binom{i-1}{j}\binom{n-2i+j+2}{i-j}, \quad where$$

$$B_n = \{\mathbf{x_n} | \mathbf{x_n} = x_1 x_2 \ldots x_n \in \{0,1\}^n \wedge (\forall k \in N_{n-2})(x_k x_{k+1} x_{k+2} \ne 010)\}.$$

Proof. Now we shall construct words from the set B_n, where B_n is the set of all words of length n over the alphabet X with the forbidden subword 010. First we make a partition of the set B_n into subsets B_n^i, where B_n^i is the set of all those words of length n over the alphabet X which contain exactly i zeros and do not contain the subword 010. This is a partition because

$$(1) \qquad B_n = \bigcup_{i=0}^{n} B^i \quad and \quad i \ne j \Rightarrow B_n^i \bigcap B_n^j = \emptyset \text{ for all } i,j \in N_n.$$

Let us construct the words from the set B_n^i. We write i zeros and then one of the letters "α" and "λ" in the $i-1$ ($i \geq 1$) places between i zeros. The letter "λ" denotes the empty letter, i.e. if the letter λ is written between two zeros, then, actually nothing is written and the letter α is the subword 11 i.e. $\alpha = 11$. Now we are sure that between two zeros there is not exactly one letter 1. This we can do in

$$\sum_{j=0}^{i-1} \binom{i-1}{j} \tag{2}$$

different ways, where j is the number of appearances of the letter λ in words which we are constructing. There remains to write $n - i - 2(i-1-j) = n - 3i + 2j + 2$ letters 1 on $i - 1 - j$ regions which already contain 11, as well as into the regions in front of and behind the word, that is into $i - 1 - j + 2 = i - j + 1$ regions in all. We can make this arrangement by forming a string consisting of $i - j$ partition lines and $n - 3i + 2j + 2$ letters 1. The number of these arrangements, i.e. permutations, is

$$\binom{n-2i+j+2}{i-j}. \tag{3}$$

Thus from (1), (2) and (3) Theorem 2 follows. □

Theorem 3.

$$|B_n| = \left[\frac{2\alpha^2+1}{2\alpha^2-2\alpha+3}\alpha^n\right].$$

Proof. We can make a recurrence relation for $b_n = |B_n|$. The words $\mathbf{x_n} \in B_n$ are obtained from other words $\mathbf{x_{n-1}} \in B_{n-1}$ by appending 0 or 1 in front of them. Let $\mathbf{x_{n-1}} \in B_{n-1}$, $\mathbf{x_{n-2}} \in B_{n-2}$ and $\mathbf{x_{n-3}} \in B_{n-3}$. Then $1\mathbf{x_{n-1}} \in B_n$, $011\mathbf{x_{n-3}} \in B_{n-3}$, $010\mathbf{x_{n-3}} \notin B_n$ which means that $01\mathbf{x_{n-2}} \in B_n$ if and only if $\mathbf{x_{n-2}}$ begins with the letter 1. This implies the recurrence relation

$$b_n = 2b_{n-1} - b_{n-2} + b_{n-3}. \tag{4}$$

It is easy to see that $b_1 = 2$, $b_2 = 4$ and $b_3 = 7$. The characteristic equation for (4) is

$$x^3 - 2x^2 + x - 1 = 0. \tag{5}$$

The equation (5) has one real root

$$\alpha = \frac{1}{6}\left(4 + \sqrt[3]{100+4\sqrt{621}} + \sqrt[3]{100-4\sqrt{621}}\right) \approx 1.754877666247$$

and two complex roots $\beta \pm i\gamma$ whose module $\sqrt{\beta^2+\gamma^2} = \alpha - 1$ is less than 1. Now we have

$$b_n = r\alpha^n + (p+iq)(\beta+i\gamma)^n + (p-iq)(\beta-i\gamma)^n$$

where the constants $r, p+iq$ and $p-iq$ are determined from the initial conditions i.e. $r = \frac{2\alpha^2+1}{2\alpha^2-2\alpha+3}$ and

$$b_n = \left[\frac{2\alpha^2+1}{2\alpha^2-2\alpha+3}\alpha^n\right] \quad \text{because}$$

$$\lim_{n\to\infty}(\beta+i\gamma)^n = 0 \ \text{ and } \ \lim_{n\to\infty}(\beta-i\gamma)^n = 0. \ \square$$

Theorem 2 and Theorem 3 imply:

Corollary 1.

$$|B_n| = 1 + \sum_{i=1}^{n}\sum_{j=0}^{i-1}\binom{i-1}{j}\binom{n-2i+j+2}{i-j} = \left[\frac{2\alpha^2+1}{2\alpha^2-2\alpha+3}\alpha^n\right].$$

Cases 100, 001, 110, 011 are equivalent, and it was shown in [4] that

$$(6) \quad L(k,m,n) = |C(k,m,n)| = \sum_{i=0}^{\lfloor \frac{n}{k} \rfloor}(-1)^i\binom{n-ki+i}{i}m^{n-ki} \quad \textit{where}$$

$C(k,m,n)$ is the set of all words of length n over the alphabet $\{a_1, a_2, \ldots, a_m\}$ with the forbidden fixed *good* subword. The subword $a_1a_2\ldots a_k$ is a good subword iff $a_1a_2\ldots a_s \neq a_{k-s+1}a_{k-s+2}\ldots a_k$ for each natural number $s < k$.

Theorem 4.

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

Proof. The set C_n is the set of all words of length n with forbidden subword 100 and $c_n = |C_n|$. The words $\mathbf{x_n} \in C_n$ are obtained from other words $\mathbf{x_{n-1}} \in C_{n-1}$ by appending 0 or 1 behind of them. Let $\mathbf{x_{n-1}} \in C_{n-1}$ and $\mathbf{x_{n-2}} \in C_{n-2}$. Then $\mathbf{x_{n-1}}1 \in C_n$, $\mathbf{x_{n-2}}10 \in C_n$ and $\mathbf{x_{n-2}}00 \in C_n$ if and only if $\mathbf{x_{n-2}} = 00\ldots0$. This implies the recurrence relation

$$(7) \qquad c_n = c_{n-1} + c_{n-2} + 1.$$

A special case of (6) for $(k, m) = (3, 2)$ and (7) give the Theorem 4 because

$$\lim_{n\to\infty} (\frac{1-\sqrt{5}}{2})^n = 0. \qquad \square$$

Remark. It is easy to generalize the results of this paper by substituting the alphabet $\{0, 1\}$ by any alphabet.

References

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

[2] 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 320–328 (1970), 27–34.

[3] Doroslovački, R., The set of all words over alphabet $\{0, 1\}$ of lenght n with the forbidden subword $11\dots1$, Univ. u Novom Sadu Zb. Rad. Prir.-Mat. Fak. Ser. Mat. 14, 2 (1984), 167–173.

[4] Doroslovački, R., The set of all words of length n over any alphabet with a forbidden good subword, Univ. u Novom Sadu Zb. Rad. Prir.-Mat. Fak. Ser. Mat. (in print)

[5] Einb, J. M., The enumeration of bit sequences that satisfy local criteria, Publications de l'Institut Mathématique Beograd, 27(41) (1980), 51–56.

Received by the editors July 25, 1994.