NOVI SAD J. MATH.
VOL. 29, NO. 1, 1999, 27-32

ON BINARY n-WORDS WITH FORBIDDEN 4-SUBWORDS

Doroslovački Rade[1]

Abstract. The set of all words of length n over alphabet $\{0,1\}$ with a fixed forbidden subword of length 4 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 and Subword.

1. Definitions and notations

Let $X = \{0,1\}$ be the 2-letter alphabet and 0 and 1 are its letter.

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 $x_1x_2\ldots x_n$.

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 greatest integer $\leq x$, respectively. By $\ell_q(p)$ we denote the number of subwords q in the word $p \in X^*$, where X^* is the set of all finite strings over the alphabet X i.e.

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

In the special case, by $\ell_0(p)$ and $\ell_1(p)$ we denote the number of zeros and ones, respectively in the string $p \in X^*$. The set N is the set of natural numbers. $N_n = \{1, 2, \ldots, n\}$, $N_n = \emptyset$ iff $n \leq 0$, the binomial coefficient $\binom{n}{k} = 0$ iff $n < k$ and

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

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

[1]Department of Mathematics, Faculty of Engineering, University of Novi Sad, Trg Dositeja Obradovića 6, Yugoslavia

2. Results and discussion

There are 16 cases for a forbidden subword over the alphabet $\{0.1\}$ of the length 4 and the set of all those forbidden subwords we denote by

$$S = \{x_1x_2x_3x_4 | x_1, x_2, x_3, x_4 \in X\}$$

Now we define relation ρ in the set S in the following way:

$$a_1a_2a_3a_4 \ \rho \ b_1b_2b_3b_4 \iff S_1 = S_2 \quad \text{where}$$

$S_1 = \{\mathbf{x_n} | \mathbf{x_n} = x_1, x_2 \dots x_n \in X^n \wedge (\forall i \in N_{n-2})(x_ix_{i+1}x_{i+2} \neq a_1a_2a_3a_4)\}$ and

$S_2 = \{\mathbf{x_n} | \mathbf{x_n} = x_1, x_2 \dots x_n \in X^n \wedge (\forall i \in N_{n-2})(x_ix_{i+1}x_{i+2} \neq b_1b_2b_3b_4)\}$.

We shall prove that ρ is an equivalence relation and that there are only four equivalence classes:

$$S_A = \{0000, 1111\}, \ \ S_B = \{1001, 0110, 1101, 1011, 0010, 0100\},$$

$$S_C = \{1000, 0001, 1110, 0111, 0011, 1100\} \text{ and } S_D = \{1010, 0101\}.$$

Theorem 1. *If n is a natural number, then*

$$\sum_{i_3=0}^{\lceil \frac{3n}{4} \rceil} \sum_{i_2=0}^{\lfloor \frac{2i_3}{3} \rfloor} \sum_{i_1=0}^{\lfloor \frac{i_2}{2} \rfloor} \binom{n-i_3+1}{i_3-i_2} \binom{i_3-i_2}{i_2-i_1} \binom{i_2-i_1}{i_1} = \left[\frac{2\alpha^3+2\alpha^2+2\alpha+1}{\alpha^3+2\alpha^2+3\alpha+4} \alpha^n \right]$$

where $\alpha \approx 1,927561975482925303$ is a real root of $x^4 - x^3 - x^2 - x - 1 = 0$.

Proof. In the paper [3] we have

$$a_n = |A(n)| = \sum_{i_3=0}^{\lceil \frac{3n}{4} \rceil} \sum_{i_2=0}^{\lfloor \frac{2i_3}{3} \rfloor} \sum_{i_1=0}^{\lfloor \frac{i_2}{2} \rfloor} \binom{n-i_3+1}{i_3-i_2} \binom{i_3-i_2}{i_2-i_1} \binom{i_2-i_1}{i_1} \text{ where}$$

$A(n) = \{\mathbf{x_n} | \mathbf{x_n} = x_1, x_2 \dots x_n \in X^n \wedge \ (\forall i \in N_{n-3}) \ (x_ix_{i+1}x_{i+2}x_{i+3} \neq 1111)\}$.

Words $\mathbf{x_n} \in A(n)$ are obtained from other words $\mathbf{x_{n-1}} \in A(n-1)$ by appending 0 or 1 in front of them. Let $\mathbf{x_{n-1}} \in A(n-1)$, $\mathbf{x_{n-2}} \in A(n-2)$ $\mathbf{x_{n-3}} \in A(n-3)$ and $\mathbf{x_{n-4}} \in A(n-4)$. Then $0\mathbf{x_{n-1}} \in A(n)$ and $1\mathbf{x_{n-1}} \in A(n)$ if and only if $\mathbf{x_{n-1}}$ not begins with the letters 111. Since $10\mathbf{x_{n-2}} \in A(n)$, $110\mathbf{x_{n-3}} \in A(n)$, $1110\mathbf{x_{n-4}} \in A(n)$ this implies the recurrence relation

$$a_n = a_{n-1} + a_{n-2} + a_{n-3} + a_{n-4}$$

whose characteristic equation is $x^4 - x^3 - x^2 - x - 1 = 0$ and whose roots are $\alpha \approx 1.927561975482925$, $\beta \approx -0.77480411321543$, $\gamma + i\delta$, $\gamma - i\delta$. The explicit formula for a_n is

$$a_n = C_1\alpha^n + C_2\beta^n + C_3(\gamma + i\delta)^n + C_4(\gamma - i\delta)^n.$$

Since $a_0 = 1$, $a_1 = 2$, $a_2 = 4$, $a_3 = 8$, $|\beta| < 1$, $|\gamma \pm i\delta| = \sqrt{\gamma^2 + \delta^2} = \sqrt{\frac{-1}{\alpha\beta}} < 1$, $\lim_{n\to\infty} \beta^n = 0$ and $\lim_{n\to\infty}(\gamma \pm i\delta)^n = 0$, the proof is completed. □

Corollary 1.

$$\lim_{n\to\infty} \frac{\sum_{i_3=0}^{\lceil \frac{3n}{4} \rceil} \sum_{i_2=0}^{\lfloor \frac{2i_3}{3} \rfloor} \sum_{i_1=0}^{\lfloor \frac{i_2}{2} \rfloor} \binom{n-i_3+1}{i_3-i_2}\binom{i_3-i_2}{i_2-i_1}\binom{i_2-i_1}{i_1}}{\alpha^n} = \frac{2\alpha^3+2\alpha^2+2\alpha+1}{\alpha^3+2\alpha^2+3\alpha+4}$$

Theorem 2.

$$|B(n)| = 1 + \sum_{i=1}^{n}\sum_{j=0}^{n-i}\sum_{k=0}^{\lfloor \frac{n-i-j}{3} \rfloor} \binom{i-1}{j}\binom{i-1-j}{k}\binom{n-i-j-2k+1}{k+1}$$

$$\textit{and} \quad b(n) = |B(n)| = \left[\frac{2\alpha^3+1}{2\alpha^3-3\alpha+4}\alpha^n\right] \qquad \textit{where}$$

$\alpha = \frac{1}{2}\left(1 + \sqrt{3+2\sqrt{5}}\right) \approx 1,86676039917386$ *is a root of* $x^4 - 2x^3 + x - 1 = 0$

and $B(n) = \{\mathbf{x_n} \mid \mathbf{x_n} = x_1x_2...x_n \in X^n, \ (\forall i \in N_{n-3})(x_ix_{i+1}x_{i+2}x_{i+3} \neq 0110)\}$.

The proof is given in [6]. The subwords from the set S_B are equivalent because we have the same recurrence relations for all forbidden subwords from the set S_B.

Corollary 2.

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n}\sum_{j=0}^{n-i}\sum_{k=0}^{\lfloor \frac{n-i-j}{3} \rfloor} \binom{i-1}{j}\binom{i-1-j}{k}\binom{n-i-j-2k+1}{k+1}}{\alpha^n} = \frac{2\alpha^3+1}{2\alpha^3-3\alpha+4}$$

Definition 1. *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* $s < k$.

It is obvious that all subwords (words) from S_C are good subwords and because of that [4] they are equivalent.

Theorem 3.

$$c_n = |C(n)| = \sum_{i=0}^{\lfloor \frac{n}{4} \rfloor}(-1)^i\binom{n-3i}{i}2^{n-4i} = \left[\left(1 + \frac{3}{2(\alpha^2+\alpha-1)}\right)\alpha^n - \frac{1}{2}\right]$$

where $C(n) = \{\mathbf{x_n} | \mathbf{x_n} = x_1x_2\ldots x_n \in X \wedge (\forall i \in N_{n-3})x_ix_{i+1}x_{i+2}x_{i+3} \neq p\}$, p *is a* **good** *subword of word* $\mathbf{x_n}$ *from set* S_C *and*

$$\alpha = \frac{1}{3}\left(1 + \sqrt[3]{19+3\sqrt{33}} + \sqrt[3]{19-3\sqrt{33}}\right) \approx 1.83928675521416$$

Proof. It is obvious that all words from S_C are equivalent. Because of that, we can use the subword 1000 for counting the words in $C(n)$. The left side of the identity follows from [4]. 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 in front of them. Let $\mathbf{x_{n-1}} \in C(n-1)$, and $\mathbf{x_{n-4}} \in C(n-4)$. Then $0\mathbf{x_{n-1}} \in C(n)$ and $1\mathbf{x_{n-1}} \in C(n)$ if and only if $\mathbf{x_{n-1}}$ not begins with the letters 000. Since $000\mathbf{x_{n-4}} \in C(n-1)$ this implies the recurrence relation

$$c_n = 2c_{n-1} - c_{n-4}$$

whose characteristic equation is $x^4 - 2x^3 + 1 = 0$ and whose roots are 1, $\alpha \approx 1.83928675521416$, $\beta + i\gamma$ and $\beta - i\gamma$. The explicit formula for c_n is

$$c_n = C_1 + C_2\alpha^n + C_3(\beta + i\gamma)^n + C_4(\beta - i\gamma)^n.$$

Since $c_0 = 1$, $c_1 = 2$, $c_3 = 4$, $c_4 = 8$, $|\beta \pm i\gamma| = \sqrt{\beta^2 + \gamma^2} = \sqrt{\frac{1}{\alpha}} < 1$ and $\lim_{n\to\infty}(\beta \pm i\gamma)^n = 0$, the proof is completed. □

Corollary 3.

$$\lim_{n\to\infty} \frac{\sum_{i=0}^{\lfloor \frac{n}{4} \rfloor} (-1)^i \binom{n-3i}{i} 2^{n-4i}}{\alpha^n} = 1 + \frac{3}{2\alpha^2 + 2\alpha - 2}$$

Theorem 4.

$$d_n = |D(n)| = n+1+\sum_{k=1}^{\lfloor \frac{n+1}{3} \rfloor} \sum_{i_1+i_2+\ldots+i_{k+1}=n-2k+2} i_1(i_2+1)(i_3+1)\ldots(i_k+1)i_{k+1}$$

where $i_1, i_2, \ldots, i_{k+1} \in N$ *and*

$$D(n) = \{\mathbf{x_n} | \mathbf{x_n} = x_1x_2\ldots x_n \in X^n \wedge (\forall i \in N_{n-3}) x_ix_{i+1}x_{i+2}x_{i+3} \neq 1010\}$$

Proof.

Let us count the number of all strings of the length n over the alphabet X with the forbidden substring 1010 i.e. the number of strings in the set $D(n)$. We partition the set $D(n)$ into the subsets $D^k(n)$, where $D^k(n)$ is the set of all those words of length n over alphabet X which contain exactly k substrings 10 ($\mathbf{x_n} \in D^k(n) \Rightarrow l_{10}(\mathbf{x_n}) = k$) and do not contain the subword 1010, i.e.

$$D^k(n) = \{\mathbf{x_n} | \mathbf{x_n} = x_1x_2\ldots x_n \in D(n),\ l_{10}(\mathbf{x_n}) = k\}.$$

Let us construct the words from the set $D^k(n)$. First we write k substrings 10. Then we write nonempty substrings, which are letters from the alphabet X, on the $k-1$ places between k substrings 10 and we write substrings from the same alphabet on the places in front and behind the string, that is into $k+1$ regions in all, and the number of letters in these regions, from left to

right, are $m_1, m_2, \ldots, m_{k+1}$ respectively. It is clear that $m_1, m_2 \in N \cup \{0\}$ and $m_2, m_3, \ldots, m_k \in N$. These substrings must satisfy the property that the substring 10 is forbidden in them. Now we have that

$$m_1 + m_2 + m_3 + \ldots + m_k + m_{k+1} = n - 2k \text{ and } |D(n)| = \sum_{k=1}^{\lfloor \frac{n+1}{3} \rfloor} |D^k(n)|.$$

From Theorem 1 we have that the number of substrings in the regions with m_i letters is $m_i + 1$, because these substrings are with the forbidden substring 10. It follows that

$$d_n = |D(n)| = \sum_{k=0}^{\lfloor \frac{n+1}{3} \rfloor} \sum_{m_1+m_2+\ldots+m_{k+1}=n-2k} (m_1+1)(m_2+1)\ldots(m_{k+1}+1)$$

$$\text{where } m_1, m_2 \in N \cup \{0\} \text{ and } m_2, m_3, \ldots, m_k \in N.$$

If we substitute $m_1 = i_1 - 1, m_2 = i_2, \ldots, m_k = i_k$ and $m_{k+1} = i_{k+1} - 1$, then follows Theorem 4. □

Theorem 5.

$$d_n = |D(n)| = \left[\frac{2\alpha^3 + 2\alpha - 1}{2\alpha^3 - 2\alpha^2 + 6\alpha - 4} \alpha^n \right] \quad \textit{where}$$

$$\alpha = \frac{1 + \sqrt{2} + \sqrt{2\sqrt{2} - 1}}{2} \approx 1.8832035059135.$$

Proof. It is obvious that the subwords 1010 and 0101 are equivalent and because of that we can use the subword 1010 in this counting. We call a word $\mathbf{x_n}$ *good* iff $\mathbf{x_n} \in D(n)$. Words $\mathbf{x_n} \in D(n)$ are obtained from words $\mathbf{x_{n-1}} \in D(n-1)$ by adding 0 or 1 in front of them, except that some not good words are also produced, namely those which begin in 1010, i.e. $0\mathbf{x_{n-1}} \in D(n)$ and $1\mathbf{x_{n-1}} \in D(n)$ iff $\mathbf{x_{n-1}}$ does not begin with string 010. First we subtract the number of all good words of the length $n-1$ which begin with the letter 0 (i.e. d_{n-2}) from $2d_{n-1}$, and after that we add the number of all good words of length $n-2$ which begin with either 11 or 01 or 00. The number of these words can be obtained by adding 0 or 1 in front of the words $\mathbf{x_{n-3}}$ (i.e. $2d_{n-3}$) and subtract all good words of the legth $n-3$ which begin with 0. So

$$d_n = 2d_{n-1} - d_{n-2} + 2d_{n-3} - d_{n-4}$$

whose characteristic equation is $x^4 - 2x^3 + x^2 - 2x + 1 = 0$ and whose roots are

$$\alpha = \frac{1 + \sqrt{2} + \sqrt{2\sqrt{2} - 1}}{2}, \qquad \beta = \frac{1 + \sqrt{2} - \sqrt{2\sqrt{2} - 1}}{2}$$

$$\gamma = \frac{1-\sqrt{2}+i\sqrt{2\sqrt{2}+1}}{2} \quad \text{and} \quad \delta = \frac{1-\sqrt{2}-i\sqrt{2\sqrt{2}+1}}{2}.$$

The explicit formula for d_n is

$$d_n = C_1\alpha^n + C_2\beta^n + 2R_e(C_3\gamma^n)$$

$$d_n = \frac{(2\alpha^3+2\alpha-1)\alpha^n}{2\alpha^3-2\alpha^2+6\alpha-4} + \frac{(2\beta^3+2\beta-1)\beta^n}{2\beta^3-2\beta^2+6\beta-4} + 2R_e\frac{(2\gamma^3+2\gamma-1)\gamma^n}{2\gamma^3-2\gamma^2+6\gamma-4}$$

where

$$C_1 = \frac{2\alpha^3+2\alpha-1}{2\alpha^3-2\alpha^2+6\alpha-4}, \quad C_2 = \frac{2\beta^3+2\beta-1}{2\beta^3-2\beta^2+6\beta-4}$$

$$C_3 = \frac{2\gamma^3+2\gamma-1}{2\gamma^3-2\gamma^2+6\gamma-4}, \quad \text{and} \quad C_4 = \frac{2\delta^3+2\delta-1}{2\delta^3-2\delta^2+6\delta-4}.$$

Since $|\beta| < 1$, $|\gamma| = |\delta| = 1$, and $|2R_e(C_3\gamma^n)| < 2|C_3||\gamma^n| < 2|C_3| < 0.05$, we obtain Theorem 5. □

Corollary 4

$$\lim_{n\to\infty}\frac{1}{\alpha^n}\sum_{k=1}^{\lfloor\frac{n+1}{3}\rfloor}\sum_{i_1+i_2+\ldots+i_{k+1}=n-2k+2} i_1(i_2+1)(i_3+1)\ldots(i_k+1)i_{k+1} =$$

$$= \frac{2\alpha^3+2\alpha-1}{2\alpha^3-2\alpha^2+6\alpha-4} \quad \textit{where} \quad i_1, i_2, \ldots, i_{k+1} \in N$$

References

[1] Austin Richard and Guy Richard, 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. No320–No328 1970, 27–34.

[3] Doroslovački, R., The set of all words over alphabet $\{0,1\}$ of length n with the forbidden subword $11\ldots1$, Rev. of Res., Fac. of Sci. math. ser., Novi Sad, Vol. 14, Num. 2, 1984, 167–173.

[4] Doroslovački, R., The set of all words of length n over any alphabet with a forbidden good subword, Rev. of Res., Fac. of Sci. math. ser.23, 2 (1993), 239-244 Novi Sad.

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

[6] Doroslovački, R., Binary sequences without $0\overbrace{11\ldots11}^{k-1}0$ for fixed k, Matematički vesnik 46 (1994),93-98, Beograd.

Received by the editors January 10, 1997.