Theoretical Computer Science 593 (2015) 106–116

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

The sequence of return words of the Fibonacci sequence

Yuke Huang, Zhiying Wen *

Department of Mathematical Sciences, Tsinghua University, Beijing, 100084, PR China

ARTICLE INFO

Article history:
Received 30 January 2014
Received in revised form 8 February 2015
Accepted 30 May 2015
Available online 10 June 2015
Communicated by J. Karhumäki

Keywords:
Return words
Fibonacci sequence
Singular kernel
Singular decomposition
Spectrum

ABSTRACT

Let ω be a factor of the Fibonacci sequence $F_\infty = x_1x_2\cdots$, then it occurs in the sequence infinitely many times. Let ω_p be the p-th occurrence of ω and $r_p(\omega)$ be the p-th return word over ω. In this paper, we study the structure of the sequence of return words $\{r_p(\omega)\}_{p\geq 1}$. We first introduce the singular kernel word $sk(\omega)$ for any factor ω of F_∞ and give a decomposition of ω with respect to $sk(\omega)$. Using the singular kernel and the decomposition, we prove that the sequence of return words over the alphabet $\{r_1(\omega), r_2(\omega)\}$ is still a Fibonacci sequence. We also determine the expressions of return words completely for each factor. Finally we introduce the spectrum for studying some combinatorial properties, such as power, overlap and separate of factors.

1. Introduction

As a classical example over a binary alphabet, the Fibonacci sequence, having many remarkable properties, appears in many aspects of mathematics and computer science etc., we refer to Lothaire [13,14], Allouche and Shallit [1], Berstel [3,4].

F. Durand [10] introduced the return words and proved that a sequence is primitive substitutive if and only if the set of its return words is finite. L. Vuillon [20] proved that an infinite word τ is a Sturmian sequence if and only if each nonempty factor $\omega \prec \tau$ has exactly two distinct return words. But they have not studied the expressions of the return words and the properties of the sequence composed by the return words [2].

Wen and Wen [19] introduced the singular words of the Fibonacci sequence and established two decompositions of the Fibonacci sequence with respect to the singular words, and studied some applications. A. de Luca [9] has considered these words by another way and proved that all singular words are palindromes. Cao and Wen [6] generalized the singular words to Sturmian sequences. Tan and Wen [18] generalized the singular words to Tribonacci sequence. The singular words have some applications in some aspects, such as Lyndon words [15,16], palindromes [11], smooth words [5], position of factors [7,8], Padé approximation [12,17], etc.

This paper can be seen as a continuation of the paper by Wen and Wen [19]. We extend the results from singular words to general words ω of the Fibonacci sequence, and study the structure of the sequence of return words. **The main result in this paper is: for any factor ω, the sequence of return words $\{r_p(\omega)\}_{p\geq 1}$ is the Fibonacci sequence over the alphabet $\{r_1(\omega), r_2(\omega)\}$.**

This paper is organized as follows.

Section 1 is devoted to the introduction and preliminaries. In Section 2, we prove the main result that: for any factor ω, the sequence of return words is the Fibonacci sequence. In Section 3, we determine the expressions of all return words completely. In Section 4, we will define and determine the spectrums of some combinatorics properties of factors.

* Corresponding author.
E-mail addresses: hyg03ster@163.com (Y. Huang), wenzy@tsinghua.edu.cn (Z. Wen).

http://dx.doi.org/10.1016/j.tcs.2015.05.048
0304-3975/

1.1. Basic notations

Let $\mathcal{A} = \{a, b\}$ be a binary alphabet. A word is a finite string of elements in $\mathcal{A}$. The set of all finite words on $\mathcal{A}$ is denoted by $\mathcal{A}^*$, which is a free monoid generated by $\mathcal{A}$ with the concatenation operation. The concatenation of two words $\nu = x_1x_2\cdots x_r$ and $\omega = y_1y_2\cdots y_n$ is the word $x_1x_2\cdots x_r y_1y_2\cdots y_n$, denoted by $\nu * \omega$ or $\nu\omega$. This operation is associative and has a unit element, the empty word ε. The set $\mathcal{A}^*$ is thus endowed with the structure of a monoid, and is called the free monoid generated by $\mathcal{A}$. Let Ξ denote the free group generated by $\mathcal{A}$. Let $\mathcal{A}^{\mathbb{N}}$ be the set of one-sided infinite words.

For a finite word $\omega = x_1x_2\cdots x_n$, the length of ω is equal to n and denoted by $|\omega|$. We denote by $|\omega|_a$ (resp. $|\omega|_b$) the number of letters of a (resp. b) occurring in ω. The i-th conjugation of ω is the word $C_i(\omega) := x_{i+1}\cdots x_n x_1\cdots x_i$ where $0 \le i \le n-1$. The mirror word $\overleftarrow{\omega}$ of ω is defined to be $\overleftarrow{\omega} = x_n\cdots x_2x_1$. A word ω is called a palindrome if $\omega = \overleftarrow{\omega}$.

Let $\tau = x_1\cdots x_n$ be a finite word (or $\tau = x_1x_2\cdots$ be a sequence). For any $i \le j \le n$, define $\tau[i, j] := x_ix_{i+1}\cdots x_{j-1}x_j$. That means $\tau[i, j]$ is the factor of τ of length $j - i + 1$, starting from the i-th letter and ending to the j-th letter. By convention, we note $\tau[i] := \tau[i, i] = x_i$ and $\tau[i, i-1] := \varepsilon$. Word $\omega = \tau[i, j]$ is said to occur at position i in τ. We say also that ω is a factor of τ, denoted by $\omega \prec \tau$. When we say $\omega \prec \tau$, it can be that $\omega = \tau$.

We say that ν is a prefix (resp. suffix) of a word ω, and write $\nu \lhd \omega$ (resp. $\nu \rhd \omega$) if there exists $u \in \mathcal{A}^*$ such that $\omega = \nu u$ (resp. $\omega = u\nu$).

We denote by ω^{-1} the inverse word of ω, that is, $\omega^{-1} := x_n^{-1}\cdots x_2^{-1}x_1^{-1}$. This only makes sense in Ξ. But if ν is a prefix (resp. u is a suffix) of ω, we can write $\nu^{-1}\omega = u$ (resp. $\omega u^{-1} = \nu$), with $\omega = \nu u$. This makes sense in $\mathcal{A}^*$.

1.2. The Fibonacci sequence

Let $\mathcal{B}$ and $\mathcal{C}$ be two alphabets. A morphism is a map φ from $\mathcal{B}^*$ to $\mathcal{C}^*$ that the identity $\varphi(xy) = \varphi(x)\varphi(y)$ for all words $x, y \in \mathcal{B}^*$, see [1]. The Fibonacci sequence F_∞ is the fixed point beginning with a of the Fibonacci morphism $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ defined over $\mathcal{A}^*$ by $\sigma(a) = ab$ and $\sigma(b) = a$.

$$F_\infty = abaababaabaababaababaabaababaabaababaababaabaababaababa\cdots$$

For the details of the properties of the sequence, see [19].

Since σ is a morphism, $\sigma(ab) = \sigma(a)\sigma(b)$. The k-th iteration of σ is $\sigma^k(a) = \sigma^{k-1}(\sigma(a))$, $k \ge 2$ and we denote $F_k = \sigma^k(a)$. We define $\sigma^0(a) = a$ and $\sigma^0(b) = b$. Note that the length of $F_k = \sigma^k(a)$ is the k-th Fibonacci number f_k, given by the recursion formulas $f_{-1} = 1$, $f_0 = 1$, $f_1 = 2$, $f_{k+1} = f_k + f_{k-1}$ for $k \ge 0$. Let $\delta_k \in \{a, b\}$ be the last letter of F_k. It's easy to see that when k is even then $\delta_k = a$ and when k is odd then $\delta_k = b$.

It is well known that the Fibonacci sequence F_∞ is uniformly recurrent, i.e., each factor ω occurs infinitely often and with bounded gaps between consecutive occurrences [1]. We arrange them in the sequence $\{\omega_p\}_{p\ge1}$, where ω_p denote the p-th occurrence of ω.

Remark. We will distinguish the factors ω and ω_p. For the former, we only consider ω as a factor of F_∞ no matter where it occurs; for the latter, we stress ω_p is the p-th factor of ω occurring F_∞ (it depends on two "variables": ω and p). Although ω_p and ω_q have the same expression, the factors ω_p and ω_q ($p \ne q$) are distinct. For instance, take $\omega = aba$, then $\omega_1 = F_\infty[1, 3]$ and $\omega_2 = F_\infty[4, 6]$.

1.3. Return words and the sequence of return words

The definitions of return words and the sequence of return words below are from F. Durand [10].

Definition 1.1 *(Return words).* Let ω be a factor of F_∞. For $p \ge 1$ and let $\omega_p = x_{i+1}\cdots x_{i+n}$ and $\omega_{p+1} = x_{j+1}\cdots x_{j+n}$. The factor $x_{i+1}\cdots x_j$ is called the p-th return word of ω and denoted by $r_p(\omega)$. If no confusion happens, we denote by r_p for short.

We denote by $\mathcal{R}_\omega$ the set of return words over ω, i.e., $\mathcal{R}_\omega = \{r_p(\omega),\ p \ge 1\}$.

$r_0(\omega)$ is defined as the prefix of F_∞ before ω_1, it is not a return word, but for the convenience, we call it the 0-th return word of ω.

Let ω be a factor, by the definition of the return words, we get immediately a decomposition of F_∞ with respect to the return words.

Lemma 1.2 *(Decomposition w.r.t. return words). Let ω be a factor, the sequence F_∞ can be written in a unique way as*

$$F_\infty = r_0(\omega)r_1(\omega)r_2(\omega)r_3(\omega)\cdots r_p(\omega)\cdots$$

Definition 1.3 *(The sequence of return words).* The sequence $\{r_p(\omega)\}_{p\ge1}$ is called the sequences of the return words of factor ω.