Theoretical Computer Science 548 (2014) 14–24

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Extensions of rich words

Jetro Vesti

University of Turku, Department of Mathematics and Statistics, 20014 Turku, Finland

ARTICLE INFO

Article history:
Received 18 December 2013
Received in revised form 11 June 2014
Accepted 25 June 2014
Available online 1 July 2014
Communicated by M. Crochemore

Keywords:
Combinatorics on words
Palindromes
Rich words
Sturmian words
Defect
Two-dimensional words

ABSTRACT

A word w is *rich* if it has $|w|+1$ many distinct palindromic factors, including the empty word. This article contains several results about rich words, particularly related to extending them. A word w can be *eventually extended richly* in n ways if there exists a finite word u and n distinct letters $a \in \mathrm{Alph}(w)$ such that wua is rich. We will prove that every (non-unary) rich word can be eventually extended richly in at least two different ways, but not always in three or more ways. We will also prove that every rich word can be extended to both periodic and aperiodic infinite rich words.

The *defect* of a finite word w is defined by $\mathrm{D}(w)=|w|+1-|\mathrm{Pal}(w)|$. This concept has been studied in various papers. Here, we will define a new concept, *infinite defect*. For a finite word w the definition is $\mathrm{D}_\infty(w)=\min\{\mathrm{D}(z) \mid z$ is an infinite word which has factor $w\}$. We will show that the infinite defect of a finite word is always finite and give some upper bounds for it. The difference between defect and infinite defect is also investigated.

We will also give an upper and a lower bound for the number of rich words. A new class of words, two-dimensional rich words, is also introduced.

1. Introduction

In [11], it was proved that every word w has at most $|w|+1$ many distinct palindromic factors, including the empty word. The class of words which achieve this limit was introduced in [6] with the term *full* words. The authors of [13] studied these words thoroughly and named them *rich* (in palindromes). This class of words has been studied in several other papers from various points of view, for example in [1,8–10] and [18].

In Section 2 we will prove several results about extending rich words. A rich word w can be *extended richly* with a word $u \in \mathrm{Alph}(w)^+$ if wu is rich. In [13] it was proved that every rich word can be extended with at least one letter. We will prove that every rich word w can be extended richly with at least two different letters, after it has been extended with a word of length at most $2|w|$. This fact will be used in several places. We will also show that every rich word can be extended to both an infinite aperiodic and infinite periodic rich word. Also, all Sturmian words can be extended richly in two ways.

In Section 3 we will define a new concept, the *infinite defect* of a finite word. The *defect* of a finite word w is defined by $\mathrm{D}(w)=|w|+1-|\mathrm{Pal}(w)|$. We can also study how many defects a finite word must have if it has to be extended to an infinite word. Hence, we define the *infinite defect* of a finite word w with $\mathrm{D}_\infty(w)=\min\{\mathrm{D}(z) \mid z$ is an infinite word which has factor $w\}$, where we suppose $\mathrm{Alph}(z)\subseteq \mathrm{Alph}(w)$. We will show that this number is always finite and give some upper bounds for it. We will also study how the defect and the infinite defect can differ.

E-mail address: jejove@utu.fi.

http://dx.doi.org/10.1016/j.tcs.2014.06.033

In Section 4 we will give upper and lower bounds for the number of rich words of length n. For the lower bound we will use the fact that every rich word can be extended in at least two different ways after a limited extension. There have been no previous studies investigating the number of these words.

In Section 5 we will shortly introduce and study two-dimensional rich words and their extensions.

In Section 6 we will give some open problems from the previous sections.

1.1. Definitions and notation

An *alphabet* A is a non-empty finite set of symbols, called *letters*. A *word* is a finite sequence of letters from A. The *empty* word ϵ is the empty sequence. The set A^* of all finite words over A is a *free monoid* under the operation of concatenation. The *free semigroup* $A^+ = A^* \setminus \{\epsilon\}$ is the set of non-empty words over A.

A *right* (resp. *left*) *infinite word* is a sequence indexed by $\mathbb{Z}_+$ (resp. $\mathbb{Z}_-$) with values in A. A *two-way infinite word* is a sequence indexed by $\mathbb{Z}$. We denote the set of all infinite words over A by A^ω and define $A^\infty = A^* \cup A^\omega$. A right infinite word is *ultimately periodic* if it can be written as $uv^\infty = uvvv\cdots$, for some $u, v \in A^*$, $v \neq \epsilon$. If $u = \epsilon$, then we say the infinite word is *periodic*. An infinite word that is not ultimately periodic is *aperiodic*.

The *length* of a word $w = a_1a_2\ldots a_n \in A^+$, with each $a_i \in A$, is denoted by $|w| = n$. The empty word ϵ is the unique word of length 0. By $|w|_a$ we denote the number of occurrences of a letter a in w. The *reversal* of w is denoted by $\tilde{w} = a_n \ldots a_2a_1$. A word w is called a *palindrome* if $w = \tilde{w}$. The empty word ϵ is assumed to be a palindrome.

A word x is a *factor* of a word $w \in A^\infty$ if $w = uxv$, for some $u, v \in A^\infty$. If $u = \epsilon$ ($v = \epsilon$) then we say that x is a *prefix* (resp. *suffix*) of w. A factor x of a word w is said to be *unioccurrent* in w if x has exactly one occurrence in w. Two occurrences of factor x are said to be *consecutive* if there is no occurrence of x between them. A factor of w having exactly two occurrences of a non-empty factor u, one as a prefix and the other as a suffix, is called a *complete return* to u in w.

If $w = uv \in A^+$, we use the notation $u^{-1}w = v$ or $wv^{-1} = u$ to mean the removal of a prefix or a suffix of w. The *right* (resp. *left*) *palindromic closure* of a word w is the unique shortest palindrome $w^{(+)}$ (resp. ${}^{(+)}w$) having w as a prefix (resp. suffix). If u is the (unique) longest palindromic suffix of $w = vu$ then $w^{(+)} = vu\tilde{v}$.

Let w be a finite or infinite word. The set $\mathrm{F}(w)$ is the set of all factors of w, the set $\mathrm{Alph}(w)$ is the set of all letters that occur in w and the set $\mathrm{Pal}(w)$ is the set of all palindromic factors of w. We say that a word w is *unary* if $|\mathrm{Alph}(w)| = 1$, *binary* if $|\mathrm{Alph}(w)| = 2$, *ternary* if $|\mathrm{Alph}(w)| = 3$ and *n-ary* if $|\mathrm{Alph}(w)| = n$

Other basic definitions and notation in combinatorics on words can be found from Lothaire's books [14] and [15].

1.2. Basic properties of rich words

In this subsection we provide some basic definitions and state some already known properties and characterizations of rich words.

Proposition 1.1. *(See [11, Proposition 2].) A word w has at most $|w| + 1$ distinct palindromic factors, including the empty word.*

Definition 1.2. A word w is *rich* if it has exactly $|w| + 1$ distinct palindromic factors.

Definition 1.3. An infinite word is *rich* if all of its factors are rich.

Proposition 1.4. *(See [13, Corollary 2.5].) A word w is rich if and only if all of its factors are rich.*

Proposition 1.5. *(See [13, Corollary 2.5].) If w is rich, then it has exactly one unioccurrent longest palindromic suffix (referred to later as* lps *or* $\mathrm{lps}(w)$*).*

From Corollary 2.5 in [13] we also get that if w is rich then $\tilde{w}$ is rich. From this we see that the above proposition holds for prefixes also and we refer to the unioccurrent longest palindromic prefix of w as lpp or $\mathrm{lpp}(w)$.

Proposition 1.6. *(See [13, Theorem 2.14].) A finite or infinite word w is rich if and only if all complete returns to any palindromic factor in w are themselves palindromes.*

Proposition 1.7. *(See [13, Proposition 2.8].) Suppose w is a rich word. Then there exist letters $x, z \in \mathrm{Alph}(w)$ such that wx and zw are rich.*

Proposition 1.8. *(See [13, Proposition 2.6].) Palindromic closure preserves richness.*

Let w be a word and $u \neq w$ its longest proper palindromic suffix. The *proper palindromic closure* of $w = vu$ is defined as $w^{(++)} = vu\tilde{v}$. From the proof of Proposition 2.8 in [13] we get that also the proper palindromic closure preserves richness using the fact that the longest proper palindromic suffix (referred to later as $\mathrm{lpps}(w)$ or lpps) can occur only in