Discrete Applied Mathematics 206 (2016) 115–121

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

On del-robust primitive words

Amit Kumar Srivastava [a,*], Ananda Chandra Nayak [b], Kalpesh Kapoor [b]

[a] Department of Computer Science and Engineering, Indian Institute of Technology Guwahati, Guwahati, 781039, India

[b] Department of Mathematics, Indian Institute of Technology Guwahati, Guwahati, 781039, India

ARTICLE INFO

Article history:
Received 16 May 2015
Received in revised form 10 December 2015
Accepted 26 January 2016
Available online 21 February 2016

Keywords:
Combinatorics on words
Primitive words
Del-robust
Reflective
Context-free language

ABSTRACT

A word is said to be primitive if it cannot be expressed as non-trivial power of another word. We characterize a class of primitive words, referred as del-robust primitive words, which remain primitive on deletion of any letter. It is also shown that the language of primitive words that are not del-robust is not context-free. Finally, we present a linear time algorithm to recognize del-robust primitive words and give a lower bound on the number of n-length del-robust primitive words.

1. Introduction

Let V be a finite alphabet. A word or string is a sequence of symbols or letters drawn from V. Combinatorics on words is the study of mathematical and computational problems related to words and it plays an important role in several areas including formal languages and automata theory [1], coding theory [2], string algorithms [4], computational biology [8], and DNA-computing [12].

A word is said to be primitive if it cannot be expressed as a non-trivial power of another word. Formally, a word w is primitive if there does not exist any word v such that $w = v^n$ with $n \geq 2$. As the definition is meaningful only when the alphabet has at least two letters, we assume throughout that V is a non-trivial alphabet with at least two distinct symbols. Primitive words have been extensively studied in the literature, see for example [14,19,22,18,6,5].

The relation between the language of primitive words and other formal languages has also been thoroughly explored [14,6,18]. It is a long standing important open problem whether the language of primitive words is a context-free language [21]. A linear time algorithm to test whether a given word is primitive is given in [7].

In [19,17], the language of primitive words is explored with respect to point mutations and homomorphism. A primitive word, u, is said to be del-robust if the word remains primitive on deletion of any letter in u. In this paper, we investigate the language of del-robust primitive words. In particular, our contribution is as follows.

(a) We characterize del-robust primitive words and identify several properties.
(b) We show that the language of primitive words that are not del-robust, is not context-free.
(c) We give a linear time algorithm to test if a word is del-robust primitive.
(d) We give a lower bound on the number of del-robust primitive words of a given length.

* Corresponding author.
E-mail addresses: amit.srivastava@iitg.ernet.in (A.K. Srivastava), n.ananda@iitg.ernet.in (A.C. Nayak), kalpesh@iitg.ernet.in (K. Kapoor).

http://dx.doi.org/10.1016/j.dam.2016.01.031

The paper is organized as follows. The next section reviews the basic concepts on words and some existing results on robustness of primitive words. In Section 3 we introduce del-robust primitive words and study several of their properties. In Section 4 we give a linear time algorithm to recognize a del-robust primitive word. In Section 5, we give a lower bound on the number of del-robust primitive words of a given length. Finally, conclusions and some open problems are presented in Section 6.

2. Preliminaries

Let V be a non-trivial alphabet having at least two distinct elements. The elements of V are referred to as letters or symbols. A word is a sequence of letters and it may be finite or infinite. We use word or string interchangeably. In this paper we only deal with finite words. The length of a word u is denoted by $|u|$. The notation $|u|_s$ is used for the number of occurrences of a symbol $s \in V$ in a word u. The string with length zero (also referred to as empty string) is denoted as λ.

The set of all words of length n over V is denoted by V^n. We define $V^* = \bigcup_{n \in N} V^n$, where $V^0 = \{\lambda\}$ and, $V^+ = V^* \backslash \{\lambda\}$. A language L over V is a subset of V^*. Let u and v be two words. The *concatenation* of any two words u and v is denoted as $u \,.\, v$ or simply as uv.

Definition 1 (*Reflective Language [19]*). A language L is called reflective if $uv \in L$ implies $vu \in L$, for all $u, v \in V^*$.

Let $w = uv$ be a word. Then, the words u and v are said to be *prefix* and *suffix* of the word w, respectively. A word y is said to be a *factor* of a word w if w can be written as xyz, where $y \in V^+$ and $x, z \in V^*$. The word y is said to be a proper factor if $x \neq \lambda$ or $z \neq \lambda$. A prefix (suffix) with length k of a word u is denoted by $pref(u, k)$ (suff(u, k), respectively), where $k \in \{0, 1, \ldots, |u|\}$ and $pref(u, 0) = suff(u, 0) = \lambda$.

We use the notation $pref(u)$ to specify the set of all non-empty prefixes of a word u. Let $w = a_1 \ldots a_n$ be a word, where $a_i \in V$ for $i \in \{1, \ldots, n\}$. The reverse of the word w is $rev(w) = a_n a_{n-1} \ldots a_2 a_1$. The factor $a_i a_{i+1} \ldots a_j$ is denoted by $w[i..j]$, where $i \leq j$. The cardinality of a set X is denoted by $|X|$. For elementary notions and results in formal language theory, we refer to [9]. A word z is a p-power if $z = x^p$ for some nonempty word x and $p \geq 1$ [20].

The set V^* is a free monoid under the concatenation operation and $V^+ = V^* \backslash \{\lambda\}$ is the semigroup and λ is the identity. A well studied theme on words is about their periodicity and primitivity properties [14,22,18,6,19] and associated counting and sampling problems [13].

The period p of a word $w = a_1 \ldots a_n$, where each $a_i \in V$, is defined as the smallest integer such that $a_i = a_{i+p}$ for $1 \leq i \leq n - p$. The ratio $e = \frac{|w|}{p}$ is called the exponent of the word w. A word w is said to be a repetition if and only if $e \geq 2$.

A maximal repetition at a position i in a word is a factor $w(i, j)$ which is a repetition such that its extension by one letter to the right or to the left yields a word with a larger period, that is,

- $per(w(i, j)) < per(w(i, j + 1))$
- $per(w(i, j)) < per(w(i - 1, j))$

where $per(w)$ is period of a word w [10,11].

For example, the factor *ababa* in the word $w = abaababaabaab$, is a maximal repetition at fourth position with period 2, while the factor *abab* is not a maximal repetition at this position.

A word w is said to be a conjugate of a word x if w is a cyclic shift of x, that is, if $w = uv$ and $x = vu$ for some $u, v \in V^*$ [22]. The following theorem relates two conjugates if they are powers of words.

Theorem 1 ([20]). *Let w and x be conjugates. Then w is a power if and only if x is a power. Furthermore, if $w = y^k$, $k \geq 2$, then $x = z^k$ where z is a conjugate of y.*

A word $w \in V^+$ is said to be primitive if w cannot be written as an integer power of a shorter word. Formally, w is primitive if $w = v^n$ implies $w = v$ and $n = 1$. The languages of primitive and non-primitive words are denoted by Q and Z, respectively [14]. We denote the set of primitive words of length n as $Q(n)$ and the set of non-primitive words of length n as $Z(n)$. Several facts are known about the languages Q and Z. We mention some of them below which will be used later in the paper.

Lemma 2 ([19]). *The languages Q and Z are reflective.*

The next theorem is about an equation in words and identifies a sufficient condition under which three words are powers of a common word.

Theorem 3 ([16]). *If $u^m v^n = w^k \neq \lambda$ for words $u,\ v,\ w \in V^*$ and natural numbers $m,\ n,\ k \geq 2$, then u, v and w are powers of a common word.*

The following lemma is a consequence of Theorem 3 which states that a word obtained by concatenating powers of two distinct primitive words is also primitive.

Lemma 4 ([16]). *If $p, q \in Q$ with $p \neq q$, then $p^i q^j \in Q$ for all $i, j \geq 2$.*