Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Advances in Applied Mathematics 33 (2004) 413–415
ADVANCES IN Applied Mathematics
www.elsevier.com/locate/yaama

On the theorem of Fredricksen and Maiorana about de Bruijn sequences

Eduardo Moreno [a,b,1]

[a] *Institut Gaspard Monge, Université de Marne-la-Vallée, Champs-sur-Marne, 77454 Marne-la-Vallée cedex 2, France*
[b] *Departamento de Ingeniería Matemática, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Centro de Modelamiento Matemático, UMR 2071, UCHILE-CNRS, Casilla 170-3, Correo 3, Santiago, Chile*

Received 20 September 2003; accepted 26 October 2003

Available online 25 March 2004

Abstract

This work gives an alternative proof for the theorem of Fredricksen and Maiorana [Discrete Math. 23 (1978) 207–210] about constructing a de Bruijn sequence by concatenation of the Lyndon words in lexicographic order. This proof gives the exact position of all the words in the sequence, and allows us to extend this result to the concatenation of any number of the last Lyndon words in increasing order.

Keywords: De Bruijn sequence; Lyndon words; Necklaces

1. Introduction

Let A be a finite set with a linear order $<$. A *word* on the alphabet A is a finite sequence of elements of A. The length of a word $w \in A^*$ is denoted by $|w|$. A word p is said to be a *prefix* of a word w if there exists a word u such that $w = pu$. The prefix p is proper if $p \neq w$. The definition for a *suffix* is symmetrical.

The set A^* of all the words on the alphabet A is linearly ordered by the lexicographic order induced by $<$. By definition, $x < y$ if either x is a prefix of y or if $x = uav$, $y = ubw$

E-mail address: emoreno@dim.uchile.cl.
[1] Partially supported by ECOS C00E03 (French-Chilean Cooperation), Programa Iniciativa Científica Milenio P01-005, Proyecto MECESUP UCH0009 and CONICYT PhD Fellowship.

0196-8858/$ – see front matter
doi:10.1016/j.aam.2003.10.002

with $u, v, w \in A^*$, $a, b \in A$ and $a < b$. A basic property of the lexicographic order is the following: if $x < y$ and if x is not a prefix of y, then $xu < yv$ for all words u, v.

Two words x, y are *conjugated* if there exist words u, v in A^* such that $x = uv$ and $y = vu$. Conjugacy is an equivalence relation in A^*. A word is said to be *minimal* if it is the smallest in its conjugacy class. A word is *primitive* if it is not a proper power, i.e., if it is not of the form u^n for $u \in A^*$ and $n \geqslant 2$. A *Lyndon word* is a word that is both primitive and minimal.

A de Bruijn sequence of span n is a string B^n of length $|A|^n$ such that all the words of length n are present as substrings of B^n exactly once. A very important question about de Bruijn sequence is how to generate them efficiently (see [1] for a survey in this subject). One of the most efficient and elegant solutions to this problem is given in [2], and it is achieved by the concatenation in lexicographic order of the Lyndon words whose lengths divide n.

In this work, we give an alternative proof of this theorem giving the exact position of all words in the sequence, and presenting some conditions for the existence of a de Bruijn sequence by concatenation of Lyndon words. These conditions allow us to conclude on the construction of a de Bruijn sequence by concatenating any number of the last Lyndon words in the lexicographic order.

2. The theorem of Fredricksen and Maiorana

Theorem 1. *For a given n, the lexicographic concatenation of Lyndon words of length dividing n generates a de Bruijn sequence of span n.*

Proof. Let a and z be the minimal and the maximal letters in the alphabet A, let σ be the usual shift operator and let B^n be the de Bruijn sequence of span n.

We will prove that for any minimal word w of length n, all its conjugated words $\sigma^i(w)$, $i = 0 \ldots n-1$, are substrings of B^n.

Let $w = w_1 \ldots w_j z^{n-j}$ be a minimal word, with $w_j < z$. First, we will show that the last $n-j$ conjugated words are substrings of B^n. Note that these words have the form $z^i w_1 \ldots w_j z^{n-j-i}$, for $i = 1 \ldots n-j$.

Let v be the minimal Lyndon word with prefix $w_1 \ldots w_j z^{n-j-i}$. Then, the previous minimal word in the lexicographic order has the form $u = u_1 \ldots u_{n-i} z^i$ with $u_1 \ldots u_{n-i} < w_1 \ldots w_j z^{n-j-i}$. Hence, the Lyndon word before v has suffix z^i, and then $z^i w_1 \ldots w_j z^{n-j-i}$ is a substring of B^n.

Now we will prove this for the first $j-1$ rotations. If w is not a Lyndon word (is not primitive), let $\bar{w}$ be the primitive root of w and let l be its length. Note that $\bar{w}$ has the form $\bar{w}_1 \ldots \bar{w}_{j'} z^{l-j'}$ with $l - j' = n - j$. If $\bar{w} \neq z$ then the next Lyndon word in lexicographic order x has the form $x = \bar{w}^{n/l-1} w_1 \ldots w_{j'-1}(w_{j'}+1) b_{j'+1} \ldots b_l$, so $\sigma^i(w)$ is a substring of $\bar{w}x$ for $i = 0 \ldots j-1$. The case $\bar{w} = z$ is trivial.

If w is primitive, let x be the next minimal word in lexicographic order (not necessarily primitive). Therefore x has the form $x_1 \ldots x_{j-1}(x_j+1) b_{j+1} \ldots b_n$ and in this case $\sigma^i(w)$ is a substring of wx for $i = 0 \ldots j-1$. If x is primitive, then wx is a substring of B^n and we are done, otherwise, by the previous argument, x is a prefix of $\bar{x}y$ where y is the next

Lyndon word in lexicographic order, hence wx is a substring of $w\bar{x}y$ and therefore it is a substring of B^n. □

Given a set $\{L_i\}_{i\in I}$ of Lyndon words whose lengths divide n, we can naturally associate a language to this set composed by words of lengths n corresponding to all conjugated words of L_i, if $|L_i| = n$, and all conjugated words of $(L_i)^r$, if $|L_i| = n/r$. A de Bruijn sequence generating this language will be called a "partial"-de Bruijn sequence.

Those languages are interesting because given a set $\mathcal{F}$ of forbidden blocks, the language of circular words not having a substring in $\mathcal{F}$ is exactly the language associated to Lyndon circular words not having a substring in $\mathcal{F}$. This kind of language is know as *subshift of finite type* and is the fundamental concept of the symbolic dynamic.

Corollary 2. *Let $L_1, \ldots, L_m$ be the Lyndon words of A^n of length dividing n ordered in lexicographic order, then for any $s < m$, $L_s L_{s+1} \ldots L_m$ is a "partial"-de Bruijn sequence.*

Proof. We will prove that all rotations of a minimal word w are substrings of the sequence.

If w is not a Lyndon word (i.e., if it is a power of a Lyndon word L_k with $|L_k| < n$) then by the previous proof we know that all rotations of w are contained in $L_{k-1}L_kL_{k+1}$. Hence, we only have to check the case $i = s$, but in this case, for $m > 2$ we have that z^n is a suffix of the sequence, and then we have enough z letters to obtain all the rotations of w beginning with z.

Let $w = w_1 \ldots w_j z^{n-j}$ be a Lyndon word L_k with $w_j < z$. Again by the previous proof, we know that the first $j-1$ rotations of w are contained in $L_k L_{k+1} \ldots L_m$ and so we only need to check the last $n-j$ conjugated words, which have the form $z^i w_1 \ldots w_j z^{n-j-i}$ for $i = 1 \ldots n-j$.

By the previous proof, if the minimal Lyndon word having prefix $w_1 \ldots w_j$ is included in the sequence, then we know that all the rotations of w are substrings of the sequence, on the contrary we would have $w_1 \ldots w_j < L_s \leqslant w_1 \ldots w_j z^{n-j}$, which means that the first Lyndon word in the sequences has the form $L_s = w_1 \ldots w_j b_{j+1} \ldots b_n$ and therefore $z^{n-j-i} w_1 \ldots w_j$ is a substring of the sequence.

It remains to check the rotations of the form $z^i w_1 \ldots w_j z^{n-j-i}$ for $i = 1 \ldots n-j-1$. If the minimal Lyndon word having prefix $w_1 \ldots w_j z$ is included in the sequence then we are done. If not, $w_1 \ldots w_j z < L_s \leqslant w1 \ldots w_j z^{n-j}$, in which case $L_s = w_1 \ldots w_j z b_{j+2} \ldots b_n$. Therefore we conclude that $z^{n-j-1} w_1 \ldots w_j z$ is a substring of the sequence.

We can repeat this argument until have a minimal Lyndon word with prefix $w_1 \ldots w_j z^t$, in which case all the remaining rotation $z^i w_1 \ldots w_j z^{n-j-i}$ for $i = 1 \ldots n-j-t$ will be substrings of the sequence. We know that such a t exists because $L_s \leqslant w_1 \ldots w_j z^{n-j}$. □

References

[1] H. Fredricksen, A survey of full length nonlinear shift register cycle algorithms, SIAM Review 24 (2) (1982) 195–221.

[2] H. Fredricksen, J. Maiorana, Necklaces of beads in k colors and k-ary de Bruijn sequences, Discrete Math. 23 (1978) 207–210.