Univ. u Novom Sadu
Zb.Rad. Prirod.-Mat.Fak.
Ser.Mat. 19,2, 53-60 (1989)

REVIEW OF RESEARCH
FACULTY OF SCIENCE
MATHEMATICS SERIES

COUNTING, GENERATION AND RECOGNITION OF S-SEQUENCES

Ratko Tošić and Ivica Bošnjak

Institute of Mathematics, Dr Ilije Duričića 4, 21000 Novi Sad, Yugoslavia

Abstract

S-words (also called *S*-sequences) are of importance in the theory of n-ary structures. In this paper it is proved that the number of different *S*-words of the length n is the $(n-1)$-th Catalan number. A simple way of generating all the different *S*-words of the length n is given. Also an algorithm is given for the recognition of *S*-words. The complexity of this algorithm is $O(n)$, where n is the length of the word.

1. Some definitions and notation

The following scheme will be called the *S*-scheme:

0
11
221, 122
3321, 2331, 2222, 2222, 1332, 1233

It consists of the sequences (words) of non-negative integers generated in the following way. Let a be a word in the i-th row of the *S*-scheme. Replacing a number k of the word a by the subword $(k+1)(k+1)$ (we denote such a replacement by $k\to(k+1)(k+1)$), we obtain a word in the $(i+1)$-th row of the *S*-scheme. Since the only word of the first row is 0, we conclude that all

AMS Mathematics Subject Classification: (1980): 05A15.
Key words and phrases: S-word, Catalan numbers.

the words of the i-th row have the same length i. In such a way each word of the i-th row produces i words of the $(i+1)$-th row. So, the number of words in the i-th row is $i!$ They are not all different, however. For example, the word 2222 appears twice in the fourth row.

We call the words of the S-scheme S-words or S-sequences.

Let α be an integer. By replacing each number k in the S-scheme with $k+\alpha$, we obtain the so-called $(S+\alpha)$-scheme. For example,

2
33
443, 344
5543, 4553, 4444, 4444, 3554, 3455
.
.
.

is an $(S+2)$-scheme.

A word from the n-th row of the S-scheme has a length n. We also say that it is an S-(n) word.

In [1] some properties of S-words are examined. It is proved, for example, that in the n-th row of the S-scheme there are not two words $a = a_1a_2\ldots a_n$ and $b = b_1b_2\ldots b_n$ such that $a_i < b_i$ for each $i = 1,2,\ldots n$. The problem is posed by Cupona. The investigation of S-words was inspired by their applications in the theory of n-ary structures.

We define the weigth $w(a)$ of an S-word $a=a_1a_2\ldots a_n$ in the following way:

$$w(a) = \sum_{i=1}^{n} a_i .$$

In [1] it is also proved that for an S-(n) word the maximal weigth is

$$\binom{n+1}{2} - 1$$

and the minimal weigth is

$$2(n - 2^{[\log n]}) + n\,[\log n],$$

where $\log n = \log_2 n$ and $[x]$ is the greatest integer $\le x$.

We shall also use the following notation:

$\|a\|$ – the length of the word a;

x_k^a – the number of appearances of the number k in the word a;

y_k^a – the number of applications of the replacement k-1 ⟶ kk during the process of producing the word a.

2. Counting of S-(n) words

First, we shall prove an auxiliary statement.

Theorem 1. *If a and b are two S-words such that* $\|a\| < \|b\|$, *then there is a non-negative integer k such that* $x_k^a > x_k^b$.

Proof. For an arbitrary *S*-word *c* we have:

$$\|c\| = 1 + \sum_{i\geq 1} y_i^c$$

and

(1)
$$x_j^c = 2y_j^c - y_{j+1}^c ,$$

for j=1,2,..

If $\|a\| < \|b\|$, then

(2)
$$\sum_{i\geq 1} y_i^a < \sum_{i\geq 1} y_i^b .$$

Suppose that, for each i, $y_i^a \geq y_i^b$.

Then,

$$\sum_{i\geq 1} y_i^a \geq \sum_{i\geq 1} y_i^b ,$$

which is in contradiction with (2).

Hence, it follows that for some $k \geq 0$,

(3)
$$y_{k+1}^a < y_{k+1}^b .$$

If $y_1^a < y_1^b$, then $x_0^a > x_0^b$. Otherwise, take the least such integer $k \geq 1$. Then,

(4)
$$y_k^a \geq y_k^b .$$

Now, from (1) it follows that

$$x_k^a = 2y_k^a - y_{k+1}^a , \quad x_k^b = 2y_k^b - y_{k+1}^b ,$$

and taking into account (3) and (4) we obtain

$$x_k^a > x_k^b . \quad \square$$

Corollary. *If* $a = a_1 a_2 \ldots a_n$ *is an S-word, then for any* $m = 1,2,\ldots,n-1$, $a' = a_1 a_2 \ldots a_m$ *is not an S-word.*

Theorem 2. *The number of different S-(n)words is*

$$\frac{1}{n}\binom{2n-2}{n-1} .$$

Proof. Denote the number of different S-(n)words by $f(n)$. It is clear that at the same time $f(n)$ is the number of different $(S+\alpha)$-(n) words, for any integer α .

Let a be an arbitrary S-(n)word. This word can be represented as a concatenation of some two $(S+1)$-words b and c, i.e. $a = bc$. From the Corollary of Theorem 1, it follows that such a representation is unique. On the other hand, it is clear that the concatenation of any two $(S+1)$-words is an S-word.

Hence, it follows that the number $f(n)$ satisfies the recurrence relation

$$f(n) = \sum_{k=1}^{n-1} f(k)\, f(n-k).$$

Since obviously $f(1) = 1$, the solution of this recurrence relation is the very well known series of Catalan numbers (see e.g. [2]), i.e.

$$f(n) = C_{n-1} = \frac{1}{n}\binom{2n-2}{n-1} . \tag{5}$$

$\square$

Now, we are going to determine the number of S-word of special types.

Theorem 3. *The number of different S-(n)words of maximal weigth is* 2^{n-1}, *for* $n \geq 2$.

Proof. An S-(n)word has the maximal weigth iff in the process of generation of this word we apply each replacement to the maximal number. This means that at the k-th step a number k-1 will be replaced be kk. In such a word, the maximal number always appears exactly twice in two adjacent places. Hence, there are two possibilities at each step, except at the first one when we have only one possibility: $0 \to 11$. It follows that in the n-th row of the S-scheme we obtain 2^{n-2} words of maximal weigth. They are all different. Indeed, it is obvious for n=2. Suppose that it is true for n=k. In each S-(k) word the maximal element appears in exactly two adjacent places. Now, any two S-(k+1) words obtained from two different S-(k) words of maximal weigth are obviously different. On the other hand, two S-(k+1) words of maximal weigth obtained from the same S-(k) word of maximal weigth differ in places in which the first maximal numbers appear. □

Theorem 4. *The number of different S-(n) words of minimal weigth is*

$$\binom{2^{[\log n]}}{n - 2^{[\log n]}},$$

for $n \geq 1$.

Proof. It is obvious that for $n=2^k$ ($k \in N$), the only S-(n) word of minimal weigth is of the form

(6) $$kk\ldots k\ .$$

So, the number of such words is

$$1 = \binom{2^k}{2^k - 2^k} = \binom{2^{[\log n]}}{n - 2^{[\log n]}}.$$

Now, suppose that $n = 2^k + r$, where $1 \leq r < 2^k$. Then, any S-(n) word of minimal weigth has $2^k - r$ letters (numbers) k and $2r = 2(n-2^k)$ letters (k+1). Such a word is obtained from word (6) using $r = n-2^k$ replacements of the form $k \to (k+1)(k+1)$. A choice of r letters k to be replaced can be made in

$$\binom{2^k}{r} = \binom{2^k}{n - 2^k} = \binom{2^{[\log n]}}{n - 2^{[\log n]}}$$

ways. Hence follows the statement. □

3. Recognition of S-words

The series 1, 1, 2, 5, 14, 42, 132,... i.e. $C_n = \frac{1}{n+1}\binom{2n}{n}$, for $n=0,1,2,\ldots$ occurs very often in counting problems. Some form of Catalan numbers appears every time we find a recurrence relationship of type (5). This series appears also as the solution of the following combinatorial problem [2].

We define an C-sequence as a sequence of integers $c_1c_2\ldots c_n$, such that

(7) $$1 \le c_1 \le c_2 \le \ldots \le c_n$$

and

(8) $$c_1 \le 1,\ c_2 \le 2, \ldots,\ c_n \le n\ .$$

It is very well known that the number of C-sequences of length n is

$$C_n = \frac{1}{n+1}\binom{2n}{n}\ .$$

Now, we are going to establish a bijection between the set of all S-$(n+1)$ words and the set of all the C-sequences of length n.

Let $c= c_1c_2\ldots c_n$ be a C-sequence. Then, starting form the S-word 0, we can generate an S-word $a=a_1a_2\ldots a_{n+1}$ in n steps as follows. The i-th step consists in replacing the c_i-th letter of the S-word. Condition (8) guaranties the possibility of correspondig replacements.

Example. The S-word corresponding to the C-sequence c = 113346 is generated in the following way:

0
11
221
2222
22332
223442
2234433 .

In the above example we first replace the first letter 0 by 11 (c_1=1), then the first letter 1 of word 11 by 22 (c_2=1), in the third step we replace the third letter 1 of word 221 by 22 obtaining the S-word 2222 etc.

The letters of a C-word can be considered as the sequence of instruction for the generation of the corresponding S-word. Instruction c_i means that the i-th replacement is applied to the i-th letter of the S-word.

We say that in this way an S-word is generated *from left to right*.

It is clear that different S-words correspond to different C-sequences. Since the number of different S-$(n+1)$ words is the same (C_n) as the number of different C-sequences of length n, it follows that each S-word can be generated from left to right uniquely.

This unique way of generation enables us to consturct an algorithm for recognizing any S-(n) word in at most n-1 steps.

The recognition algorithm for S-words

Let $a=a_1a_2\ldots a_n$ be a sequence of non-negative integers $(n \geq 2)$. Starting from the S-word 0, we first apply a_1 replacements of the first letter. In this way we obtain an S-word $a'=a_1a'_2\ldots a'_{a_1+1}$.

If $a'=a$, we conclude that a is an S-word. If $a_1+1 > n$ or $a_1+1 = n$ and $a'\neq a$, then a is not an S-word.

If $n > a_1+1$, we look for the first i such that $a_i\neq a'_i$. Such an i exists according to the Corollary of Theorem 1.

Now, if $a'_i > a_i$, this means that a is not an S-word. If $a'_i < a_i$, we apply $a_i - a'_i$ times the replacement of the i-th letter. In this way, we obtain a new S-word

$$a''= a_1a_2\ldots a_ia''_{i+1}\ldots a''_{a_1+a_i-a'_i} .$$

If, continuing in this way, after n-1 replacements we obtain the word a, this means that a is an S-word. Otherwise, after at most n-1 steps, we obtain a word $b= b_1b_2\ldots b_m$, such that either

$m < n$ and b is a prefix of a

or

for some $j \in \{1,2,\ldots,m\}$, $a_i = b_i$ if $i < j$ and $a_j < b_j$.

In that case we conclude that a is not an S-word.

References

1. **A. Samardžijski, R. Tošić, S. Markovski:** Slobodni (1,2)-grupoidi
2. **D.I.A. Cohen:** Basic techniques of combinatorial theory, John Wiley and Sons 1978.

Rezime

PREBROJANJE, GENERISANJE I PREPOZNAVANJE S-REČI

S-reči (*S*-nizovi) nalaze primenu u teoriji n-arnih struktura. U ovom radu dokazano je da je broj različitih *S*-reči dužine *n* jednak (*n*-1)-om broju Katalana. Dat je jedan postupak za generisanje svih različitih *S*-reči dužine *n*. Dat je, takođe, jedan algoritam složenosti $O(n)$ za prepoznavanje *S*-reči dužine *n*.

Received by the editors December 23, 1987.