ON LANGUAGES GENERATED BY SEMIGROPS

Ladislav Janiga and Václav Koubek
Charles University
Malostranské náměstí 25
Praha 1
Czechoslovakia

1. Abstract

In this paper we investigate properties of classes of languages generated by semigroups in the following sense:

Let $S$ be a semigroup, $s$ its element, $A$ an alphabet and $f$ a mapping which assigns to each $a \in A$ a finite subset $f(a)$ of $S$. A word $w = a_1 \ldots a_n \in A^+$ belongs to the language generated by $S$, $f$, $s$ and $A$ iff the following condition holds:

$$\exists s_1 \in f(a_1) \ldots \exists s_n \in f(a_n) \quad s = s_1 \ldots s_n.$$

We investigate the languages generated in this sense by certain types of semigroups. It is important to emphasize that we do not consider the empty word, thus all languages are subsets of free semigroups of the form $A^+$, similarly all homomorphisms we investigate are semigroup homomorphisms only. Since the proofs of our results are based on rather complicated semigroup constructions we sketch them out or omit at all. The detailed proofs are to be presented in [10].

2. Introduction

The aim of this paper is to present an other way in which formal languages can be characterized. Let us mention three important approaches to characterization of formal languages:

/i/ Characterization of languages by classes of machines which recognize the languages.

/ii/ Characterization via a class of grammars which generate the languages.

/iii/ Characterization by a collection of operations by which the languages can be obtained from a basic class of languages.
Our approach is based on the following idea:

Suppose we are given a class $C$ of objects over which it is possible to effectuate some kind of computation /in the sequel it will be a semigroup together with its operation/. Let us select one object $s$ from $C$. Now we can, when characterizing some language $L$, proceed as follows:

/i/ Any word over the alphabet over which we consider $L$ is related with an expression over $C$.

/ii/ For every word is evaluated the expression and if the resulting object is equal to $s$ then the word is accepted.

This idea is formalized in the following definition:

Definition 1

Let $S$ be a semigroup, let $s \in S$ and let $A$ be an alphabet.

/i/ An assignement is any function $f$ which assignes to every $a \in A$ a finite and nonempty subset $f(a)$ of $S$.

/ii/ We say that a word $w = a_1 \ldots a_n \in A^+$ belongs to the language generated by $S, A, f$ and $s$ (denoted $L(A, S, f, s)$) iff

$$\exists s_1 \in f(a_1) \ldots \exists s_n \in f(a_n) \quad s_1 \cdot \ldots \cdot s_n = s.$$ 

/iii/ A language $L$ is said to be generated by the semigroup $S$ iff it is of the form $L(A, S, f, s)$ for appropriate $A, f$ and $s \in S$.

Note

Let us denote by $P(S)$ the global of $S$, i.e. the following set:

$$\{S' \mid S' \subseteq S \land S' \text{ is finite}\}$$

together with the operation

$$S_1 \cdot S_2 = \{s_1 \cdot s_2 \mid s_1 \in S_1 \& s_2 \in S_2\}.$$ It follows that an assignement $f$ is a function of the form:

$$f : A \rightarrow P(S)$$

This function has a unique homomorphism extension $f^+: A^+ \rightarrow P(S)$ and it is easy to see that

$$L(A, S, f, s) = \{w \in A^+ \mid s \in f^+(w)\}.$$ 

Definition 1 relates the language $L(A, S, f, s)$ with the following computational process /nondeterministic in its nature/:

The input is a word $w = a_1 \ldots a_n$

/1/ for any $i$ select some $s_i$ from the set which is assigned by $f$ to the $i$-th letter of $w$,

/2/ compute $s_1 \cdot \ldots \cdot s_n = e$,

/3/ if $s = e$ then accept.

As we shall see in the sequel the assumption that the assignement $f$