Mathematica Pannonica
7/2 (1996), 215 – 222

∞ – REGULAR LANGUAGES DEFINED BY A LIMIT OPERATOR

Karel **Mikulášek**

Department of Mathematics, Technical University, 61669 Brno, Technická 2, Czech Republic

Received: May 1995

MSC 1991: 68 Q 45

Keywords: ∞-regular language, ∞-acceptor, transition graph, limit operator.

Abstract: Finite deterministic ∞-acceptor accepting both finite and infinite words over a finite alphabet is introduced. It is shown that ∞-regular languages can be defined as sets of ∞-words accepted by an ∞-acceptor. A limit operator on regular languages is used to define a special class of ∞-regular languages. In ∞-acceptors accepting these languages incidence relations between the sets used for acceptance are determined.

1. Introduction

The paper deals with special classes of ∞-regular languages. If a finite alphabet Σ is given, then by an ∞-regular language over Σ we understand the union of a regular and an ω-regular language over Σ. In this paper we show that it is possible to define such languages by means of a single finite-state device. A deterministic machine capable of constructing both finite and infinite sequences is first introduced in [7]. In [4] the structure of the sets constructed in [7] is investigated and in [5] a non-deterministic version of such machines is shown. Another generalization can be found in [3] where the notion of a k-machine is introduced. In [6] generalized non-deterministic acceptors accepting sequences of both finite and infinite length are introduced. Limit operators on regular languages are usefull tools for investigating relations between regular and ω-regular languages. In [8] and [9] they are also used for studying topological properties of ω-languages. The operator lim used in this paper appears e.g. in [1], [2], [6] and [9]. A generalization of some of the results of [6] gives rise to a question of how the structure of ∞-regular languages defined by inclusions between a limit closure of their words and their ω-words is reflected in the incidence of sets used for accepting words and ω-words.

2. Preliminaries

In this paper, all the numbers are non-negative integers if not specified otherwise. By ω we denote the least infinite ordinal number. A non-empty finite set Σ is called *alphabet* and its elements *letters*. Σ^* denotes the set of all finite sequences of Σ. The empty word is denoted by λ. Its elements are called *words*. For a $u \in \Sigma^*$, we write $u = u_1u_2\dots u_n$. For two non-empty words $u = u_1u_2\dots u_m$, $v = v_1v_2\dots v_k$, we define their *catenation* $u.v = u_1u_2\dots u_mv_1v_2\dots v_k$. We put: $\lambda.u = = u.\lambda = u$, $u \in \Sigma^*$. The catenation of k identical words $w \in \Sigma^*$ is denoted by w^k.

Σ^ω denotes the set of all infinite sequences over Σ. If $w \in \Sigma^\omega$, then we write $w = w_1w_2\dots$. These sequences are called *ω-words*. For $w \in \Sigma^*$, w^ω denotes the ω-word $w.w.\dots$. We put $\Sigma^\infty = \Sigma^* \cup \Sigma^\omega$. The elements of Σ^∞ are called *∞-words*. We define the catenation of two ∞-words u, v by using the above definition for $u, v \in \Sigma^*$ and putting $u.v = u_1u_2\dots u_nv_1v_2\dots$, if $u \in \Sigma^*, v \in \Sigma^\omega$. For $u \in \Sigma^\omega$, the catenation is not defined.

For a word $w = w_1w_2\dots w_n$, $n \geq 1$, we define its *length* as $\mid w \mid= n$ and for $w \in \Sigma^\omega$ we put $\mid w \mid= \omega$. Finally we put $\mid \lambda \mid= 0$.

For a $w \in \Sigma^\omega$ we define the set of *left fractions* of w:

$\mathrm{lf}\,(w) = \{u \in \Sigma^* \mid w = u.v,\ v \in \Sigma^\omega\}$.

A subset of Σ^* is called a *language*, a subset of Σ^ω an *ω-language* and a subset of Σ^∞ an *∞-language*.

For an infinite sequence $q = q_0, q_1, q_2, \dots$ of elements of a finite set Q we define $\mathrm{In}(q)$ as the set of all the elements of Q which occur infinitely many times in q.

In the sequel, we will use the following evident assertion: *For an arbitrary sequence $s_1, s_2, \dots, s_k$, $k \geq 1$ of elements of $\mathrm{In}(q)$, there are indices $i_1 < i_2 < \dots < i_k$ such that $s_j = q_{i_j}$, $1 \leq j \leq k$.*

A *graph* is a triple $D = (U, \tau, H)$ where U is a finite set of *nodes*, H is a finite set of *arcs* and $\tau : H \to U \times U$ is an *incidence mapping.* A graph $D' = (U', \tau', H')$ is a subgraph of a graph $D = (U, \tau, H)$ if $U' \subseteq U, H' \subseteq H$ and $\tau' : H' \to U' \times U'$ is a restriction of τ on H. For $V \subseteq U$, we say that $D_V = (V, \tau', H')$ is the subgraph of D *induced by* V if H' is a maximum subset of H such that $\tau(h) \in V \times V$ for every $h \in H'$ and τ' is a restriction of τ on H'.

A *trace* in a graph $D = (U, \tau, H)$ is an alternating sequence of nodes and arcs $u_0, a_1, u_1, a_2, \ldots, a_n, u_n$, $n \geq 1$ where $\tau(a_i) = (u_{i-1}, u_i)$, $1 \leq i \leq n$. We say that a graph $D = (U, \tau, H)$ is *strong* if, for every $u, v \in U$, there is a trace from u to v.

In proofs we will use the following obvious assertion: $D = (U, \tau, H)$ *is strong iff, for every* $u, v \in U$, *there is a trace in* D *starting in* u, *ending in* v *and containing all nodes of* U.

3. ∞-regular languages

Definition 3.1. We say that a five-tuple $A = (Q, \Sigma, \delta, q_0, F)$ is a *finite deterministic acceptor* or, shortly, an *acceptor* if Q is a finite set of states, Σ is an alphabet, $\delta : Q \times \Sigma \to Q$ is a transition function, $q_0 \in Q$ is the initial state and $F \subseteq Q$ is a set of final states. For $w \in \in \Sigma^*, w = a_1.a_2 \ldots a_n$, $n \geq 0$, a sequence $q(w) = q_0, q_1, q_2, \ldots, q_n$ is called the *run* of A on w if $q_i = \delta(q_{i-1}, a_i)$, $1 \leq i \leq n$. We put $\delta^*(w) = = q_n$. If $\delta^*(w) \in F$ then we say that A *accepts* w. The set of all words accepted by A is denoted by $\mathcal{L}(A)$. A language $L \subseteq \Sigma^*$ is called *regular* if $L = \mathcal{L}(A)$ for an acceptor A.

Definition 3.2. We say that a five-tuple $A = (Q, \Sigma, \delta, q_0, \Phi)$ is a *finite deterministic ω-acceptor* or, shortly, an *ω-acceptor* if Q is a finite set of states, Σ is an alphabet, $\delta : Q \times \Sigma \to Q$ is a transition function, $q_0 \in Q$ is the initial state and $\Phi \subseteq 2^Q$ is a system of subsets of infinitely many times occurring states. For $w \in \Sigma^\omega$, $w = a_1.a_2 \ldots$ a sequence $q(w) = q_0, q_1, q_2, \ldots$ is called the *run* of A on w if $q_i = = \delta(q_{i-1}, a_i)$, $1 \leq i$. We put $\delta^\omega(w) = \mathrm{In}(q(w))$. If $\delta^\omega(w) \in \Phi$ then we say that A *accepts* w. The set of all words accepted by A is denoted by $\mathcal{L}(A)$. An ω-language $L \subseteq \Sigma^\omega$ is called *ω-regular* if $L = \mathcal{L}(A)$ for an ω-acceptor A.

Note 3.3. Clearly, for a $w \in \Sigma^\omega$ the run of an acceptor on w is also defined and so is the run of an ω-acceptor on a $w \in \Sigma^*$.

Definition 3.4. An ∞-language $L \subseteq \Sigma^\infty$ is called ∞-*regular* if a regular language $L_F \subseteq \Sigma^*$ and an ω-regular language $L_\omega \subseteq \Sigma^\omega$ exist such that $L = L_F \cup L_\omega$.

Definition 3.5. We say that a six-tuple $A = (Q, \Sigma, \delta, q_0, F, \Phi)$ is a *finite deterministic* ∞-*acceptor* or, shortly, an ∞-*acceptor* if Q is a finite set of states, Σ is an alphabet, $\delta : Q \times \Sigma \to Q$ is a transition function, $q_0 \in Q$ is the initial state, $F \subseteq Q$ is a set of final states and $\Phi \subseteq 2^Q$ is a system of subsets of infinitely many times occurring states.

For $w \in \Sigma^\infty$ we define the *run* of A on w using either Definition 3.1 if $w \in \Sigma^*$ or Definition 3.2 if $w \in \Sigma^\omega$. Similarly, we put $\delta^\infty(w) = \delta^*(w)$ or $\delta^\infty(w) = \delta^\omega(w)$ depending on whether $w \in \Sigma^*$ or $w \in \Sigma^\omega$. If $\delta^\infty(w) \in F \cup \Phi$, then we say that A *accepts* w. The set of all ∞-words accepted by A is denoted by $\mathcal{L}(A)$, the set of all ω-words accepted by A is denoted by $\mathcal{L}^\omega(A)$ and the set of all words accepted by A is denoted by $\mathcal{L}^F(A)$.

Definition 3.6. Let $A = (Q, \Sigma, \delta, q_0, F, \Phi)$ be an ∞-acceptor. We say that $G(A) = (Q, \tau, H)$ is the *transition graph* of A if

$$H = \big\{(q_i, a, q_j)\big|q_i, q_j \in Q, a \in \Sigma, q_j = \delta(q_i, a)\big\}, \quad \tau(q_i, a, q_j) = (q_i, q_j).$$

For a $P \subseteq Q$, we denote by $G(A, P)$ the subgraph of the transition graph of A induced by P.

Theorem 3.7. *Let $L \subseteq \Sigma^\infty$ be an ∞-regular language. Then there exists an ∞-acceptor A such that $L = \mathcal{L}^\infty(A)$.*

Proof. We have $L = L_F \cup L_\omega$ where $L_F \subseteq \Sigma^*$ is regular and $L_\omega \subseteq \Sigma^\omega$ is ω-regular. This means that $L_F = \mathcal{L}(B), L_\omega = \mathcal{L}(C)$ where $B = (P, \Sigma, \epsilon, p_0, G)$ is an acceptor and $C = (R, \Sigma, \phi, r_0, \Psi)$ is an ω-acceptor. Let us construct an ∞-acceptor $A = (Q, \Sigma, \delta, q_0, F, \Phi)$ by putting $Q = P \times R$, $q_0 = (p_0, r_0)$, $\delta((p, r), a) = (\epsilon(p, a), \phi(r, a))$ and $F = \{(p, r) \in Q \mid p \in G\}$. To construct Φ let us consider $w \in L_\omega$. Let

$$p(w) = p_0, p_1, p_2, \dots \tag{1}$$

be the run of B on w and

$$r(w) = r_0, r_1, r_2, \dots \tag{2}$$

the run of C on w. Put

$$q(w) = (p_0, r_0), (p_1, r_1), (p_2, r_2), \dots \tag{3}$$

and define $\Phi = \{\mathrm{In}(q(w)) \mid w \in L_\omega\}$. Φ is finite since, for every $w \in L_\omega$, $\mathrm{In}(q(w)) \subseteq Q$. Let $w \in L$. For a $w \in L_F$ we have $\epsilon^*(w) \in G$, which means that $\delta^\infty(w) \in F$ and thus $w \in \mathcal{L}^F(A)$. If $w \in L_\omega$, then, using the notation (1), (2) and (3), we see that $\mathrm{In}(q(w)) \in \Phi$ and thus, since clearly $\delta^\infty(w) = \mathrm{In}(q(w))$, we get $w \in \mathcal{L}^\omega(A)$.

To prove the inverse inclusion $\mathcal{L}^\infty(A) \subseteq L$ let us first take a $w \in \mathcal{L}^F(A)$. This gives us a run of A on w: $(p_0, r_0), (p_1, r_1), (p_2, r_2), \ldots, (p_k, r_k)$, $k \geq 0$, where $p_k \in G$. Then, of course, $p_0, p_1, p_2, \ldots, p_k$ is a run of B on w and $w \in \mathcal{L}(B) = L_F$. If, on the other hand, $w \in \mathcal{L}^\omega(A)$, we get the run of B on w: $p(w) = p_0, p_1, p_2, \ldots$, the run of C on w: $r(w) = r_0, r_1, r_2, \ldots$ and the run of A on w: $q(w) = (p_0, r_0), (p_1, r_1), (p_2, r_2), \ldots$ with $\delta^\infty(w) \in \Phi$. This means that there is a $w' \in \mathcal{L}(C)$ such that

$$\mathrm{In}(q(w)) = \mathrm{In}(q(w')) \tag{4}$$

where again $p(w') = p_0, p'_1, p'_2, \ldots$ is the run of B on w, $r(w') = r_0, r'_1, r'_2, \ldots$ is the run of C on w and $q(w') = (p_0, r_0), (p'_1, r'_1), (p'_2, r'_2), \ldots$ is the run of A on w. Let $s \in \mathrm{In}(r(w))$. This means that s occurs infinitely many times in $r(w)$ and thus a p_i exists such that (p_i, s) occurs infinitely many times in $q(w)$ or $(p_i, s) \in \mathrm{In}(q(w))$ and (4) gives us $(p_i, s) \in \mathrm{In}(q(w'))$. Then s must occur infinitely many times in $r(w')$, which implies $s \in \mathrm{In}(r(w'))$ and $\mathrm{In}(r(w)) \subseteq \mathrm{In}(r(w'))$. However $\mathrm{In}(r(w')) \subseteq \mathrm{In}(r(w))$ can be proved in much the same way. The equality $\mathrm{In}(r(w)) = \mathrm{In}(r(w'))$ then implies $w \in \mathcal{L}(C) = L_\omega$. $\Diamond$

4. Limit ∞-regular languages

Definition 4.1. Let $L \subseteq \Sigma^*$. Put $\lim L = \{w \in \Sigma^\omega \mid \mathrm{card}(\mathrm{lf}(w) \cap L) = \omega\}$. The operator lim maps the set of all languages over Σ into the set of all ω-languages over Σ.

Note 4.2. In [1] this operator is denoted by L^δ and in [8] it is called *δ-limit*. We use the notation as introduced in [2] and [6].

Lemma 4.3. *Let $w \in \Sigma^\omega$ and $L \subseteq \Sigma^*$. Then $w \in \lim L$ iff a sequence $w_1, w_2, \ldots$ of words from L exists such that*

$$\mid w_i \mid < \mid w_j \mid, i < j, w_i \in \mathrm{lf}(w), i \geq 1.$$

Proof. The proof follows directly from Def. 4.1. $\Diamond$

Definition 4.4. We say that an ∞-acceptor $A = (Q, \Sigma, \delta, q_0, F, \Phi)$ is *concise* if $\mathcal{L}^F(A) \neq \emptyset, \mathcal{L}^\omega(A) \neq \emptyset$ and, for any $A' = (Q, \Sigma, \delta, q_0, F', \Phi)$, $A'' = (Q, \Sigma, \delta, q_0, F, \Phi'')$ where $F' \subset F$, $\Phi'' \subset \Phi$, we have $\mathcal{L}^F(A') \subset \mathcal{L}^F(A)$, $\mathcal{L}^\omega(A'') \subset \mathcal{L}^\omega(A)$.

Lemma 4.5. *Let $A = (Q, \Sigma, \delta, q_0, F, \Phi)$ be a concise ∞-acceptor. Then the following conditions are equivalent:*

1. $\lim \mathcal{L}^F(A) \subseteq \mathcal{L}^\omega(A)$;
2. *if, for an $S \subseteq Q$, $G(A, S)$ is strong and $S \cap F \neq \emptyset$, then $S \in \Phi$.*

Proof. Let the first condition hold and let $S \subseteq Q$ be such that $G(A, S)$ is strong and $S \cap F \neq \emptyset$. Let $f \in S \cap F$. Since A is concise, there is a word $u \in \mathcal{L}^F(A)$ such that $\delta^\infty(u) = f$. $G(A, S)$ being strong and $f \in S$ implies that there is a trace in $G(A, S)$

$$c_1, (c_1, a_1, c_2), c_2, (c_2, a_2, c_3), \dots, (c_{k-1}, a_{k-1}, c_k), c_k, \quad k > 1$$

such that it contains all the nodes of $G(A, S)$ and $c_1 = c_k = f$. Let us now consider a sequence of words $w_1, w_2, \dots$ where $w_i = u.(a_1.a_2 \dots \dots a_{k-1})^i$ and an ω-word $w = u.(a_1 a_2 \dots a_{k-1})^\omega$. It is easy to see that $w_i \in \mathcal{L}^F(A), w_i \in \mathrm{lf}(w),\ i \geq 1$ and $| w_i | < | w_j |, i < j$, which, by Lemma 4.3, implies $w \in \lim \mathcal{L}^F(A)$. It is also obvious that $\{c_1, c_2, \dots \dots, c_k\}$ is exactly the set of states occurring infinitely many times in the run of A on w and thus $\delta^\infty(w) = S$. However, by the assumption, $w \in \mathcal{L}^\omega(A)$ and thus $S \in \Phi$.

Let the second condition hold and $w \in \lim \mathcal{L}^F(A)$. By Lemma 4.3, we get a sequence of words $w_1, w_2, \dots, w_i \in \mathrm{lf}(w),\ i \geq 1$ such that $| w_i | < | w_j |,\ i < j$. Let us consider an infinite sequence of states from F

$$f_1, f_2, \dots, \quad f_i = \delta^\infty(w_i), \quad i \geq 1. \tag{5}$$

Since F is finite, there is an f which occurs infinitely many times in (5). Denote by

$$q(w) = q_0, q_1, q_2, \dots \tag{6}$$

the run of A on w. Since, clearly, (5) is a subsequence of (6), we have $f \in \mathrm{In}(q(w))$. $G(A, \mathrm{In}(q(w))$ is strong and $| w_i | < | w_j |$ implies that it has at least one arc and so we get $\mathrm{In}(q(w)) \in \Phi$ by the assumption and finally $w \in \mathcal{L}^\omega(A)$. $\Diamond$

Lemma 4.6. *Let $A = (Q, \Sigma, \delta, q_0, F, \Phi)$ be a concise ∞-acceptor. Then the following conditions are equivalent:*

1. $\lim \mathcal{L}^F(A) \supseteq \mathcal{L}^\omega(A)$;
2. $F_i \cap F \neq \emptyset$ *for every* $F_i \in \Phi$.

Proof. Let the first condition hold and let $F_i \in \Phi$. As A is concise, there exists a $w \in \mathcal{L}^\omega(A)$ such that $\delta^\infty(w) = F_i$. By the assumption then $w \in \lim \mathcal{L}^F(A)$ and, by Lemma 4.3, there exists a sequence of words

$$w_1, w_2, \dots, w_i \in \mathrm{lf}(w), \quad w_i \in \mathcal{L}^F(A), \quad i \geq 1$$

such that $| w_i | < | w_j |, i < j$. Thus we have $\delta^\infty(w_i) \in F, i \geq 1$ and since F is finite, there must be an $f \in F$ which occurs infinitely many times in the sequence

$$\delta^\infty(w_1), \delta^\infty(w_2), \dots. \tag{7}$$

Let

(8) $$q(w) = q_0, q_1, q_2, \ldots$$
be the run of A on w. It is easy to see that (7) is a subsequence of (8), which means that $f \in \mathrm{In}((q(w)) = \delta^\infty(w) = F_i$. Therefore $F \cap F_i \neq \emptyset$ and the second condition holds.

If the second condition holds and $w \in \mathcal{L}^\omega(A)$, then $\delta^\infty(w) = F_i \in \in \Phi$. Since $F_i \cap F \neq \emptyset$, there exists an $f \in F_i \cap F$ which occurs infinitely many times in the run of A on w and thus there is a sequence
$$w_1, w_2, \ldots, \quad w_i \in \mathrm{lf}(w), w_i \in \mathcal{L}^F(A), \quad i \geq 1$$
such that $\mid w_i \mid < \mid w_j \mid, i < j$. Then, by Lemma 4.3, $w \in \lim \mathcal{L}^F(A)$. $\diamondsuit$

Definition 4.7. Let $D = (U, \tau, H)$ be a graph. For every $v \in U$, we define a system of subsets of U: $\sigma(D, v) = \{V \subseteq U \mid v \in V \wedge G(D, V)$ is strong$\}$. For a $W \subseteq U$, we put $\sigma(D, W) = \bigcup\limits_{v \in W} \sigma(D, v)$.

Theorem 4.8. *Let $A = (Q, \Sigma, \delta, F, \Phi)$ be a concise ∞-acceptor. Then the following conditions are equivalent:*

1. $\lim \mathcal{L}^F(A) = \mathcal{L}^\omega(A)$*;*
2. $\Phi = \sigma(G(A), F)$*.*

Proof. Let the first condition hold and $F_i \in \Phi$. Then $G(A, F_i)$ is strong since A is concise. By Lemma 4.6 we have $F_i \cap F \neq \emptyset$ with an $f \in F_i \cap F$ such that $F_i \in \sigma(G(A), f)$. This means that $F_i \in \sigma(G(A), F)$. For an $F_i \in \sigma(G(A), F)$, $G(A, F_i)$ is strong and $F_i \cap F \neq \emptyset$. Thus, by Lemma 4.5, we get $F_i \in \Phi$.

If $\Phi = \sigma(G(A), F)$, then, for every $S \subseteq Q$ such that $G(A, S)$ is strong and $S \cap F \neq \emptyset$, $S \in \sigma(G(A), F) = \Phi$ and, by Lemma 4.5, $\lim \mathcal{L}^F(A) \subseteq \mathcal{L}^\omega(A)$. Next if $F_i \in \Phi$, then $F_i \in \sigma(G(A), F)$ implies $F_i \cap F \neq \emptyset$ and, by Lemma 4.6, we get $\mathcal{L}^\omega(A) \subseteq \lim \mathcal{L}^F(A)$. $\diamondsuit$

References

[1] DAVIS, M.: Infinitary Games of Perfect Information, in: Advances in Game theory, Princeton Univ. Press, Princeton N. J. 1964, 89–101.

[2] EILENBERG, S.: Automata, Languages and Machines, Volume A, New York, Academic Press 1974.

[3] GRODZKI, Z.: The k-machines, *Bull. Acad. Polon. Sci. Ser. Sci. Math. Astronom. Phys.*, **18**/7 (1970), 399–402.

[4] KWASOWIEC, S.: Generable Sets, *Information and Control* **17** (1970), 257–264.

[5] MEZNÍK, I.: G-machines and Generable Sets, *Information and Control* **20** (1972), No 5, 499–509.

[6] NOVOTNÝ, M.: Sets Constructed by Acceptors, *Information and Control*, **26** (1974), 116–133.

[7] PAWLAK, Z.: Stored Program Computers, *Algorytmy* **10** (1969), 7–22.

[8] STEIGER, L.: Research in the Theory of ω-languages, *J. Inf. Process. Cybern. EIK* 8/9 **23** (1987), 415–439.

[9] STEIGER, L.: Hierarchies of Recursive ω-languages, *Elektron. Inf. Verarb. Kybern. EIK* 5/6 **22** (1986), 219–241.