Groupoids of pseudoautomata

By F. FERENCI

Introduction

It is known [7] that to a unary universal algebra (universal algebra [6] with unary operations only) there corresponds a monoid (semigroup with identity). An automaton without outputs [4], or shortly automaton, can be obtained from a unary universal algebra by selecting an element and a subset from its base set, for the initial state and the final state set of automaton, respectively [8]. The above mentioned monoid is associated with this automaton, as well [5].

The concept of a tree automaton [1] is such a generalization of that of an automaton, when the corresponding universal algebra is not necessarily unary [12], [11]. The purpose of this paper is to show that there is an other way of generalization obtained by replacing the monoid by an arbitrary groupoid. Then the notions corresponding to those of the unary universal algebra and the automaton are the pseudoalgebra and the pseudoautomaton (which is a kind of tree automata, as well), respectively (see Conclusion at the end of the paper). These notions are introduced in the paper [10].

The method used here for representation of formal languages [9] by a set of trees is analogous to that in the author's paper [2].

1. Trees and algebraic structures

By ω we denote the set of all nonnegative integers, i.e., $\omega=\{0, 1, 2, \dots\}$, and by π the set of two parentheses (and), i.e., $\pi=\{(,)\}$. Furthermore, we suppose that π is disjoint from each other sets used here. For a set A, A^* is the set of all finite strings on A including the empty string λ and $A^+=A^*-\{\lambda\}$. If p is a finite string on A, then $\lg(p)$ denotes the length of p, i.e., the number of occurences of symbols from A in p. An *alphabet* is allways a finite nonempty set.

Let V be an alphabet and X a set of symbols disjoint from V (X may be infinite, finite, nonempty or empty). *The set of trees of type V over the set X*, in notation $[V, X]$, is a subset of $(V\cup X\cup\pi)^+$ defined as follows:

1.1. (1) $x\in[V, X]$ for all $x\in X$;

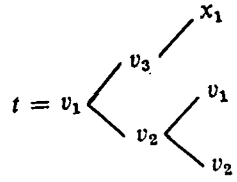
(2) if $v\in V$, $k\in\omega$ and $t_i\in[V, X]$ for $1\leqq i\leqq k$; then $v(t_1)(t_2)\dots(t_k)\in[V, X]$ (in the special case $k=0$, $v\in[V, X]$);

(3) the elements of $[V, X]$ are those and only those which we get from (1) and (2) in a finite number of steps.

It can be proved that for every $t=v(t_1)(t_2)\dots(t_k)\in[V, X]$, the components $v\in V$, $k\in\omega$ and $t_i\in[V, X]$ $(1\leqq i\leqq k)$ are uniquely determined.

If $X=\emptyset$ then we write $[V]$ for $[V, X]$, i.e., $[V]=[V, \emptyset]$.

Example 1.1. If $V=\{v_1, v_2, v_3, v_4\}$ and $X=\{x_1, x_2\}$, then $t=v_1(v_3(x_1))(v_2(v_1)(v_2))$ is an element of $[V, X]$. This tree is represented graphically in the next figure:



(It can be noticed from our definition of trees, that the use of parentheses differ from usually manner, see [10], [12], [1], etc. Our method is taken from [3].)

The *word function* W: $[V]\to V^+$ we define in the next way:

1.2. (1) if $t=v$ for some $v\in V$, then $W(t)=v$;

(2) if $t=v(t_1)(t_2)\dots(t_k)$ where $v\in V$, $k\geqq 1$, $t_i\in[V]$, $1\leqq i\leqq k$, then $W(t)=vW(t_1)W(t_2)\dots W(t_k)$.

It can be seen that for a $t\in[V]$, $W(t)$ is the word over the alphabet V [9] which is obtained from t by erasing all parentheses.

Example 1.2. If $V=\{v_1, v_2, v_3\}$ and $t=v_1(v_2)(v_1(v_2)(v_2))$, then $W(t)=v_1v_2v_1v_2v_2$.

For $T\subseteq[V]$, $W(T)$ is the set $\{W(t)\,|\,t\in T\}$ and it is a λ-free language over V [9].

Let V be an alphabet. A *pseudoalgebra of type* V is a system $\mathbf{A}=\alpha\langle V, A\rangle$ where A is a nonvoid set disjoint from V, the *base set* of $\mathbf{A}$, and α is an operator which for each v from V determines a mapping αv: $A^*\to A$ [10]. The pseudoalgebra $\mathbf{A}$ is *finite* iff A is finite. A pseudoalgebra $\mathbf{B}=\beta\langle V, B\rangle$ of the same type V is called a *subpseudoalgebra* of $\mathbf{A}$ iff $B\subseteq A$ and $\beta v(p)=\alpha v(p)\in B$ for every $v\in V$ and $p\in B^*$. Let $\mathbf{C}=\gamma\langle V, C\rangle$ be a pseudoalgebra and h a mapping of C into A. When for arbitrary $v\in V$ and $c_1c_2\dots c_k\in C^*$ $(k\in\omega,\ c_i\in C,\ 1\leqq i\leqq k)$ $h(\gamma v(c_1c_2\dots c_k))=\alpha v(h(c_1)h(c_2)\dots h(c_k))$ holds, then h is a *homomorphism* of $\mathbf{C}$ into $\mathbf{A}$. If in addition h is *onto* then $\mathbf{A}$ is a *homomorphic image* of $\mathbf{C}$. Moreover, if h is an onto and one-to-one homomorphism then it is an *isomorphism*, and $\mathbf{A}$ and $\mathbf{C}$ are called *isomorphic* pseudoalgebras.

Let us consider a pseudoalgebra $\mathbf{A}=\alpha\langle V, A\rangle$ and the set of trees $[V, A]$. By α we define a mapping $\bar\alpha$ from $[V, A]$ into A in the following way:

1.3. (1) if $t=a$ for an $a\in A$, then $\bar\alpha(t)=a$;

(2) if $t=v(t_1)(t_2)\dots(t_k)$ where $v\in V$ and $t_i\in[V, A]$ for $1\leqq i\leqq k$, then $\bar\alpha(t)=\alpha v(\bar\alpha(t_1)\bar\alpha(t_2)\dots\bar\alpha(t_k))$.

The next lemma expresses a property of homomorphisms.

Lemma 1.1. If $\mathbf{A}=\alpha\langle V, A\rangle$ and $\mathbf{B}=\beta\langle V, B\rangle$ are pseudoalgebras and $h: B\to A$ a homomorphism of **B** into **A**, then $h(\bar{\beta}(t))=\bar{\alpha}(t)$ holds for every t from $[V]$.

Proof. The proof is by induction on $\lg(t)$. First, it can be shown that if $t=v$ for some $v\in V$, the assertion is true. Then, supposing that the assertion is true for t_i, $1\leqq i\leqq k$, $k\geqq 1$, one can prove, that it is true for $t=v(t_1)(t_2)\dots(t_k)$.

A pseudoalgebra $\mathbf{A}=\alpha\langle V, A\rangle$ is *connected* iff $\bar{\alpha}([V])=A$.

The next statement is a consequence of the previous lemma.

Lemma 1.2. If **A** and **B** are connected pseudoalgebras of the same type and there exists a homomorphism h of **B** into **A** then h is uniquelly determined and **A** is a homomorphic image of **B**.

A nonvoid set A with a binary operation "multiplication" defined on A is a *groupoid* $\tilde{A}$. The result of the multiplication of two elements a_1 and a_2 from A (their "product") will be denoted by $(a_1\cdot a_2)$, but expressions obtained by a successive application of multiplication can be simplified in the known way, i.e., by erasing the outer parentheses. For example, instead of $(a\cdot b)$ and $(a\cdot(b\cdot c))$ we shall write $a\cdot b$ and $a\cdot(b\cdot c)$, respectively.

If to a groupoid $\tilde{A}$ we add an alphabet V disjoint from A and introduce a mapping $\xi\colon V\to A$, we get a *designed groupoid* $\mathcal{A}$, in symbols, $\mathcal{A}=\langle\tilde{A}, V, \xi\rangle$. The designed groupoid $\mathcal{A}$ is *finite* iff $\tilde{A}$ (i.e. A) is finite. It is *connected* iff $\tilde{A}$ is generated by the set $\xi(V)=\{\xi(v)|v\in V\}$. If $\mathcal{B}=\langle\tilde{B}, V, \eta\rangle$ is a designed groupoid too, then the mapping $h: B\to A$ is termed *homomorphism* of $\mathcal{B}$ into $\mathcal{A}$ iff it is a homomorphism of $\tilde{B}$ into $\tilde{A}$ and $h(\eta(v))=\xi(v)$ holds for arbitrary $v\in V$. "Onto" homomorphism and isomorphism are defined in the natural way.

Using a designed groupoid $\mathcal{A}=\langle\tilde{A}, V, \xi\rangle$ one can consruct a pseudoalgebra which is denoted by ind $\mathcal{A}$ *(pseudoalgebra induced by $\mathcal{A}$)* in the next way: ind $\mathcal{A}=\alpha\langle V, A\rangle$, i.e. it is of type V, its base set is A, and for every $v\in V$, $p\in A^*$ and $a\in A$ holds

1.4. (1) $\alpha v(\lambda)=\xi(v)$;

(2) $\alpha v(pa)=\alpha v(p)\cdot a$.

In other terms, for $p=a_1a_2\dots a_k$ $(a_i\in A,\ 1\leqq i\leqq k)$ $\alpha v(p)=(\dots((\xi(v)\cdot a_1)\cdot a_2)\cdot\dots)\cdot a_k$.

The following lemma will be useful in some proofs.

Lemma 1.3. Let $\mathcal{A}$ be a designed groupoid. Then ind $\mathcal{A}$ is connected iff $\mathcal{A}$ is connected.

Proof. Let $\mathcal{A}=\langle\tilde{A}, V, \xi\rangle$ and ind $\mathcal{A}=\alpha\langle V, A\rangle$. We shall show that an $a\in A$ is a product of some elements from $\xi(V)$ iff there exists a $t\in[V]$ with $\bar{\alpha}(t)=a$. We proceed by induction.

When the number of factors in the product is one, i.e. $a=\xi(v)$ for some $v\in V$, then it is equivalent to $a=\bar{\alpha}(t)$ for $t=v\in[V]$. If we suppose that the element a_i from A is a product of elements from $\xi(V)$ and $a_i=\bar{\alpha}(t_i)$ holds for some $t_i\in[V]$, where $1\leqq i\leqq k$, $k\geqq 1$, then $a=(\dots((\xi(v)\cdot a_1)\cdot a_2)\cdot\dots)\cdot a_k$ iff $a=\alpha v(a_1a_2\dots a_k)=\bar{\alpha}(t)$ for $t=v(t_1)(t_2)\dots(t_k)\in[V]$.

2. The groupoid of a pseudoalgebra

Let θ be a binary relation on a set X ($\theta \subseteq X^2$). Then we write $x_1 \theta x_2$ iff $(x_1, x_2) \in \theta$. If θ is an equivalence relation, then X/θ denote the partition of X induced by θ, i.e. the set of all equivalence classes modulo θ. For an $x \in X$ the equivalence class containing x will be denoted by $\theta(x)$.

Let $[V, A]$ be the set of trees of type V over a nonempty set A. We define the subsets $[V, A]'$ and $[V, A]''$ of this set in the following way:

2.1. $$[V, A]' = [V, A] - A$$

and

2.2.
$$[V, A]'' = \{t \mid t \in [V, A],\ t = v(a_1)(a_2)\dots(a_k) \text{ for } v \in V \text{ and } a_1 a_2 \dots a_k \in A^* \ (a_i \in A,\ 1 \leqq i \leqq k)\}.$$

If $t \in [V, A]'$ and $p = a_1 a_2 \dots a_k$ is an element from A^* ($a_i \in A$, $1 \leqq i \leqq k$), then let $t\vec{p} = t(a_1)(a_2)\dots(a_k)$ ($\in [V, A]'$).

Let us suppose that $\mathbf{A} = \alpha\langle V, A\rangle$ is a pseudoalgebra of type V. On the set $[V, A]'$ we define the relation $\varrho' \subseteq ([V, A]')^2$ in the next way:

2.3. for arbitrary trees t_1 and t_2 from $[V, A]'$ $t_1 \varrho' t_2$ holds iff $\bar{\alpha}(t_1\vec{p}) = \bar{\alpha}(t_2\vec{p})$ is satisfied for each $p \in A^*$.

The relation ϱ' is evidently an equivalence relation on the set $[V, A]'$. Into the set $[V, A]'/\varrho'$ of equivalence classes modulo ϱ' one can introduce a binary operation — multiplication — in the next manner: for arbitrary t_1 and t_2 from $[V, A]'$

2.4. $$\varrho'(t_1) \cdot \varrho'(t_2) = \varrho'(t) \quad \text{where} \quad t = t_1(t_2).$$

One can easily prove that this operation is well defined.

Now, we have a groupoid $\widetilde{[V, A]'/\varrho'}$ whose multiplication is defined by 2.4. The designed groupoid $\langle \widetilde{[V, A]'/\varrho'}, V, \alpha'\rangle$ where $\alpha'(v) = \varrho'(v)$ holds for each $v \in V$, we call the *groupoid of the pseudoalgebra* $\mathbf{A}$ and denote it by $\mathcal{G}(\mathbf{A})$.

Since $\varrho'' = \varrho' \cap ([V, A]'')^2$ is an equivalence relation too, however now on the set $[V, A]''$ ($\subseteq [V, A]'$), and each equivalence class modulo ϱ' contains elements from $[V, A]''$ (if $t_1 = v(r_1)(r_2)\dots(r_k)$ and $\bar{\alpha}(r_i) = a_i \in A$, $1 \leqq i \leqq k$, then $t_1 \varrho' t_2$, where $t_2 = v(a_1)(a_2)\dots(a_k) \in [V, A]''$). If on the set $[V, A]''/\varrho''$ we define the multiplication by

2.5. $$\varrho''(t_1) \cdot \varrho''(t_2) = \varrho''(t), \quad \text{where} \quad t = t_1(\bar{\alpha}(t_2)),$$

we get a groupoid $\widetilde{[V, A]''/\varrho''}$ isomorphic to $\widetilde{[V, A]'/\varrho'}$. Namely, it can be easily shown that the mapping $f\colon [V, A]''/\varrho'' \to [V, A]'/\varrho'$, which is defined by $f((\varrho''(t))) = \varrho'(t)$, is an isomorphism between these groupoids. For this reason the designed groupoid $\langle \widetilde{[V, A]''/\varrho''}, V, \alpha''\rangle$ where $\alpha''(v) = \varrho''(v)$ holds for each $v \in V$, is isomorphic to $\mathcal{G}(\mathbf{A})$. That means that we can consider this designed groupoid to be equal to $\mathcal{G}(\mathbf{A})$.

We investigate now the nature of the elements of $\mathcal{G}(\mathbf{A})$. For this purpose, for every $t \in [V, A]'$ let us introduce a mapping $\tilde{\alpha}t\colon A^* \to A$ in the next way: for arbitrary $p \in A^*$ let $\tilde{\alpha}t(p) = \bar{\alpha}(t\vec{p})$. Since for $t = v$ ($v \in V$), $\tilde{\alpha}t(p) = \alpha v(p)$ holds $\tilde{\alpha}$ is an extension of α. This mapping $\tilde{\alpha}t$ can be called the *mapping induced by t in* $\mathbf{A}$. From 2.3 we conclude now that *the set of elements of* $\mathcal{G}(\mathbf{A})$, up to notation, *is the same as the set of all different mappins induced by trees from* $[V, A]'$ *(or* $[V, A]''$*) in* $\mathbf{A}$.

In a similar way as we obtained the relation ϱ'' we can get the relation $\dot{\varrho}=\varrho'\cap\cap([V])^2$. The set $[V]/\dot{\varrho}$ of equivalence classes modulo $\dot{\varrho}$ equals to the set of all mappings induced by trees from $[V]$ in $\mathbf{A}$, and it is a subset of all mappings induced by trees from $[V, A]'$. This set forms a groupoid with the multiplication defined as in 2.4, and this groupoid is the subgroupoid of $\widetilde{[V, A]'/\varrho'}$ generated by the set $\{\varrho'(v)|v\in V\}$. (For each t from $[V]$, $\varrho'(t)$ is a product of elements from $\{\varrho'(v)|v\in V\}$. This can be proved by induction on $\lg(t)$.) If $\mathbf{A}$ is connected then $\widetilde{[V, A]'/\varrho'}$ is generated by the set $\{\varrho'(v)|v\in V\}$ (since then for every t from $[V, A]'$ there is an r from $[V]$ for which $r\varrho' t$ is valid; r can be getted from t by substituting each $a\in A$ with an r_a from $[V]$ for which $\bar{\alpha}(r_a)=a$). Now the next lemma, by the Lemma 1.3, follows from the fact that the set $\{\varrho'(v)|v\in V\}$ equals to the set $\alpha'(V)$.

Lemma 2.1. If the pseudoalgebra $\mathbf{A}$ is connected, then so is ind $\mathcal{G}(\mathbf{A})$.

The following theorem gives a connection between a pseudoalgebra and its groupoid.

Theorem 2.1. Let $\mathbf{A}=\alpha\langle V, A\rangle$ be a pseudoalgebra and $\mathcal{G}(\mathbf{A})$ its groupoid. Then the following assertions are valid:

1° There exists a homomorphism h from ind $\mathcal{G}(\mathbf{A})$ into $\mathbf{A}$.

2° If $\mathbf{A}$ is connected then

(a) the homomorphism h is completely determined and it is an *onto* homomorphism;

(b) if for some connected designed groupoid $\mathcal{B}=\langle\tilde{B}, V, \eta\rangle$ there exists a homomorphism from ind $\mathcal{B}$ into $\mathbf{A}$, then $\mathcal{G}(\mathbf{A})$ is a homomorphic image of $\mathcal{B}$ and ind $\mathcal{G}(\mathbf{A})$ is a homomorphic image of ind $\mathcal{B}$;

(c) if $\mathbf{A}=$ ind $\mathcal{A}$ for some designed groupoid $\mathcal{A}$, then $\mathcal{A}$ is isomorphic to $\mathcal{G}(\mathbf{A})$ and therefore, $\mathbf{A}$ is isomorphic to ind $\mathcal{G}(\mathbf{A})$.

Proof. 1° Since $\mathcal{G}(\mathbf{A})=\langle\widetilde{[V, A]'/\varrho'}, V, \alpha'\rangle$, the mapping h: $[V, A]'/\varrho'\to A$ for which $h(\varrho'(t))=\bar{\alpha}(t)$ holds, where t is an arbitrary element from $[V, A]'$, is well defined by 2.3. It can be easily shown that h is a homomorphism from ind $\mathcal{G}(\mathbf{A})$ into $\mathbf{A}$.

2° (a) It follows from 1°, and Lemmas 2.1 and 1.2.

(b) Let ind $\mathcal{B}=\beta\langle V, B\rangle$. By Lemma 1.3, ind $\mathcal{B}$ is connected and, by Lemma 1.2, $\mathbf{A}$ is a homomorphic image of ind $\mathcal{B}$ under the mapping h: $B\to A$ which is determined by Lemma 1.1. From these facts it follows that the mapping f: $B\to\to[V, A]'/\varrho'$, where $f(\bar{\beta}(t))=\varrho'(t)$ holds for arbitrary $t\in[V]$, is well defined. Indeed, if for $t_1, t_2\in[V]$, $\bar{\beta}(t_1)=\bar{\beta}(t_2)$ holds, then $\varrho'(t_1)=\varrho'(t_2)$ holds too. It can be checked as follows.

Since h is a homomorphism from ind $\mathcal{B}$ onto $\mathbf{A}$, then for $a_1, a_2, \dots, a_k$ from A there exists $b_1, b_2, \dots, b_k$ from B such, that $h(b_i)=a_i$ for $1\leqq i\leqq k$. Then

$$\bar{\alpha}(t_1(a_1)(a_2)\dots(a_k)) = h(\bar{\beta}(t_1(b_1)(b_2)\dots(b_k))) =$$
$$= h((\dots(\bar{\beta}(t_1)\cdot b_1)\cdot b_2)\cdot\dots)\cdot b_k) = h((\dots(\bar{\beta}(t_2)\cdot b_1)\cdot b_2)\cdot\dots)\cdot b_k) =$$
$$= h(\bar{\beta}(t_2(b_1)(b_2)\dots(b_k))) = \bar{\alpha}(t_2(a_1)(a_2)\dots(a_k)) \quad \text{(see 2.3)}.$$

Moreover, the mapping f is a homomorphism from ind $\mathscr{B}$ into ind $\mathscr{G}(\mathbf{A})$, and since ind $\mathscr{B}$ is connected then from Lemmas 2.1 and 1.2 it follows that f is an onto homomorphism.

The same mapping f is a homomorphism from $\mathscr{B}$ onto $\mathscr{G}(\mathbf{A})$.

(c) The proof is a routine computation.

From the assertion 2° (c) of the previous theorem, it follows that each connected designed groupoid is the groupoid of a pseudoalgebra.

3. Finite pseudoalgebras

Let us suppose that the pseudoalgebra $\mathbf{A}=\alpha\langle V, A\rangle$ is finite, i.e., A is a finite set. For arbitrary $v\in V$ and $a\in A$, let $\mathscr{L}(v, a)$ denote the set $\{p\,|\,p\in A^*, \alpha v(p)=a\}$. It is evident, that $\mathscr{L}(v, a)$ is a language over A. We shall say that the pseudoalgebra $\mathbf{A}$ is *regular* iff $\mathscr{L}(v, a)$ is a regular language over A [9] for each $v\in V$ and $a\in A$.

The next theorem is of great importance for our approach.

Theorem 3.1. The groupoid of a finite pseudoalgebra $\mathbf{A}$ is finite iff $\mathbf{A}$ is regular.

Proof. Let $\mathbf{A}=\alpha\langle V, A\rangle$. It is known from the previous chapter, that $\mathscr{G}(\mathbf{A})$ is isomorphic to the designed groupoid $\langle\widehat{[V, A]''}/\varrho'', V, \alpha''\rangle$. Therefore, $\mathscr{G}(\mathbf{A})$ is finite iff there are finitely many equivalence classes modulo ϱ''.

For arbitrary $v\in V$ let $\varrho''_v=\varrho''\cap([\{v\}, A]'')^2$. The equivalence relations ϱ''_v, when v is running through V, have the property that each equivalence class modulo ϱ'' is the union of some equivalence classes modulo ϱ''_v such that for each $v\in V$ at most one equivalence class modulo ϱ''_v occurs in this union. Consequently, for the finiteness of V, there are finitely many equivalence classes modulo ϱ'' iff there are finitely many equivalence classes modulo ϱ''_v, i.e. $[\{v\}, A]''/\varrho''_v$ is finite, for each $v\in V$.

We give now a necessary and sufficient condition for the finiteness of $[\{v\}, A]''/\varrho''_v$. From 2.3 we have that for arbitrary $t_1=v\vec{p}$ and $t_2=v\vec{q}$ from $[\{v\}, A]''$, where p and q are in A^*, $t_1\varrho''_v t_2$ holds iff $\bar{\alpha}(t_1\vec{r})=\bar{\alpha}(t_2\vec{r})$ is valid for every $r\in A^*$, which is the same as

$$3.1. \qquad \alpha v(pr)=\alpha v(qr) \quad \text{for every} \quad r\in A^*.$$

Now, we can induce an equivalence relation σ_v on the set A^* in the next way: $p\sigma_v q$ holds iff 3.1 is valid. It is obvious that A^*/σ_v is finite iff $[\{v\}, A]''/\varrho''_v$ is finite.

An other equivalence relation $\bar{\sigma}_v$ on A^* can be defined by: $p\bar{\sigma}_v q$ iff $\alpha v(p)=\alpha v(q)$. In the members of the partition $A^*/\bar{\sigma}_v$ we can recognise the sets $\mathscr{L}(v, a)$ for those $a\in A$ for which $\mathscr{L}(v, a)$ is nonempty. Moreover, the partition A^*/σ_v is a refinement of $A^*/\bar{\sigma}_v$. From the definitions of σ_v and $\bar{\sigma}_v$ it can be seen that the partition A^*/σ_v is the maximal right automaton-partition of the set A^* written into the partition $A^*/\bar{\sigma}_v$ (see [5]). It is finite iff each member of $A^*/\bar{\sigma}_v$ is a regular language over the alphabet A, in other terms, iff each $\mathscr{L}(v, a)$ is regular.

4. Pseudoautomata

Let $\mathbf{A}=\alpha\langle V, A\rangle$ be a pseudoalgebra of type V. Selecting a subset A_F of A, we get from $\mathbf{A}$ a *pseudoautomaton* $\overline{\mathbf{A}}=(\mathbf{A}, A_F)=(\alpha\langle V, A\rangle, A_F)$. Sets V, A and A_F are the sets of *inputs, states* and *final states* of $\overline{\mathbf{A}}$, respectively, while α can be called the *transition function* of $\overline{\mathbf{A}}$ (see [10]). The pseudoautomaton $\overline{\mathbf{A}}$ is *finite, regular* or *connected* iff the pseudoalgebra $\mathbf{A}$ is finite, regular or connected, respectively. The *set of trees represented* by $\overline{\mathbf{A}}$, in symbols $\mathcal{T}(\overline{\mathbf{A}})$, is a subset of $[V]$ defined by

4.1. $$\mathcal{T}(\overline{\mathbf{A}}) = \{t \mid t\in[V],\ \bar{\alpha}(t)\in A_F\}.$$

For an alphabet V, a subset T of trees from $[V]$ will be called *recognizable* iff it is represented by a finite pseudoautomaton. In the case when the pseudoautomaton is regular, T is called *pseudoregular.*

The *language represented* by a pseudoautomaton $\overline{\mathbf{A}}$, in symbols $\mathcal{L}(\overline{\mathbf{A}})$, is defined by

4.2. $$\mathcal{L}(\overline{\mathbf{A}}) = W(\mathcal{T}(\overline{\mathbf{A}})) = \{W(t) \mid t\in\mathcal{T}(\overline{\mathbf{A}})\}.$$

The set $\mathcal{L}(\overline{\mathbf{A}})$ is, obviously, a language over the set of inputs of $\overline{\mathbf{A}}$.

Let $\overline{\mathbf{A}}=(\mathbf{A}, A_F)$ and $\overline{\mathbf{B}}=(\mathbf{B}, B_F)$ be two pseudoautomata, where $\mathbf{A}=\alpha\langle V, A\rangle$ and $\mathbf{B}=\beta\langle V, B\rangle$ are pseudoalgebras of the same type V. If there is a mapping $h\colon B\to A$ which is a homomorphism from $\mathbf{B}$ into $\mathbf{A}$, and in addition for every b from B, $b\in B_F$ iff $h(b)\in A_F$ then h is a *homomorphism from* $\overline{\mathbf{B}}$ *into* $\overline{\mathbf{A}}$. If h is also a homomorphism (isomorphism) of $\mathbf{B}$ onto $\mathbf{A}$, then we say that h is a *homomorphism (isomorphism) of* $\overline{\mathbf{B}}$ *onto* $\overline{\mathbf{A}}$. In this case $\overline{\mathbf{A}}$ *is a homomorphic image of* $\overline{\mathbf{B}}$ (h is an *isomorphism between* $\overline{\mathbf{A}}$ *and* $\overline{\mathbf{B}}$). We shall say that the pseudoautomaton $\overline{\mathbf{B}}$ is a *subpseudoautomaton* of $\overline{\mathbf{A}}$ iff $\mathbf{B}$ is a subpseudoalgebra of $\mathbf{A}$ and $B_F=A_F\cap B$. The pseudoautomaton $\overline{\mathbf{B}}$ is the *trunk* of the pseudoautomaton $\overline{\mathbf{A}}$ iff $\overline{\mathbf{B}}$ is the connected subpseudoautomaton of $\overline{\mathbf{A}}$. (Note that the trunk is completely determined and $B=\{\bar{\alpha}(t) \mid t\in[V]\}$.)

Two pseudoautomata $\overline{\mathbf{A}}$ and $\overline{\mathbf{B}}$ will be called *equivalent* iff $\mathcal{T}(\overline{\mathbf{A}})=\mathcal{T}(\overline{\mathbf{B}})$. It is evident, that for equivalent pseudoautomata $\overline{\mathbf{A}}$ and $\overline{\mathbf{B}}$, $\mathcal{L}(\overline{\mathbf{A}})=\mathcal{L}(\overline{\mathbf{B}})$ also holds (the opposite is not true).

The next result is a direct consequence of our definitions and Lemma 1.1.

Theorem 4.1. If there exists a homomorphism of the pseudoautomaton $\overline{\mathbf{B}}$ into the pseudoautomaton $\overline{\mathbf{A}}$, then $\overline{\mathbf{A}}$ and $\overline{\mathbf{B}}$ are equivalent. Consequently, $\mathcal{L}(\overline{\mathbf{A}})=\mathcal{L}(\overline{\mathbf{B}})$.

Moreover, we obviously have

Theorem 4.2. Any pseudoautomaton is equivalent to each of its subpseudoautomata. Especially, a pseudoautomaton is equivalent to its trunk.

The second part of the previous theorem shows that connected pseudoautomata are of special interest.

From Lemma 1.2 we can get the next theorem.

Theorem 4.3. Let $\overline{\mathbf{A}}$ and $\overline{\mathbf{B}}$ be connected pseudoautomata with a common set of inputs. Assume that h is a homomorphism of $\overline{\mathbf{B}}$ into $\overline{\mathbf{A}}$. Then h is uniquelly determined and $\overline{\mathbf{A}}$ is a homomorphic image of $\overline{\mathbf{B}}$.

By *the groupoid of a pseudoautomaton* $\overline{\mathbf{A}}=(\mathbf{A}, A_F)$, in symbols $\mathscr{G}(\overline{\mathbf{A}})$ we mean the designed groupoid $\mathscr{G}(\mathbf{A})$.

If for a pseudoautomaton $\overline{\mathbf{B}}=(\mathbf{B}, B_F)$ the pseudoalgebra $\mathbf{B}$ is of the form ind $\mathscr{B}$, where $\mathscr{B}$ is a designed groupoid, we shall call $\overline{\mathbf{B}}$ *groupoid pseudoautomaton.* The groupoid pseudoautomaton $\overline{\mathbf{B}}$ is *the groupoid pseudoautomaton belonging to* $\overline{\mathbf{A}}=(\alpha\langle V, A\rangle, A_F)$, iff $\mathscr{B}=\mathscr{G}(\overline{\mathbf{A}})=\langle\widetilde{[V, A]'/\varrho'}, V, \alpha'\rangle$ and $B_F=\{\varrho'(t)|t\in[V, A]', \bar{\alpha}(t)\in A_F\}$. In this case $\overline{\mathbf{B}}$ is denoted by $G(\overline{\mathbf{A}})$.

On account of our definitions and previous results we can state the following three theorems. The first of them is based on Lemma 2.1.

Theorem 4.4. If $\overline{\mathbf{A}}$ is a connected pseudoautomaton then $G(\overline{\mathbf{A}})$ is connected too.

The next theorem follows from Theorem 2.1, properties of the mapping h from the proof of the assertion 1° in this theorem and from our definitions.

Theorem 4.5. Let $\overline{\mathbf{A}}=(\mathbf{A}, A_F)$ be a pseudoautomaton and $G(\overline{\mathbf{A}})$ the groupoid pseudoautomaton belonging to $\overline{\mathbf{A}}$. The following assertions are valid:

1° There exists a homomorphism h from $G(\overline{\mathbf{A}})$ into $\overline{\mathbf{A}}$.

2° If $\overline{\mathbf{A}}$ is connected then

(a) $G(\overline{\mathbf{A}})$ is connected and h is completely determined *onto* homomorphism;

(b) if $\overline{\mathbf{B}}$ is a connected groupoid pseudoautomaton and there exists a homomorphism of $\overline{\mathbf{B}}$ into $\overline{\mathbf{A}}$ then $\overline{\mathbf{A}}$ is a homomorphic image of $\overline{\mathbf{B}}$ and $G(\overline{\mathbf{A}})$ is a homomorphic image of $\overline{\mathbf{B}}$;

(c) if $\overline{\mathbf{A}}$ is a groupoid pseudoautomaton then $G(\overline{\mathbf{A}})$ is isomorphic to $\overline{\mathbf{A}}$.

From the previous theorem it is clear, that the relation between a pseudoautomaton and its groupoid is similar to the relation between an automaton and its monoid (see [5]).

The following theorem is important in investigating finite pseudoautomata. It follows from Theorem 3.1.

Theorem 4.6. If $\overline{\mathbf{A}}$ is a finite pseudoautomaton then $G(\overline{\mathbf{A}})$ is finite iff $\overline{\mathbf{A}}$ is regular.

In the next theorem languages represented by regular pseudoautomata are characterised. (For language-theoretic terminology used here, see [9]. It should be emphasized that here in the definition of context-free grammar we take a set of initial letters instead of a single letter. Obviously, this modification does not alter the generative capacity of context-free grammars.)

Theorem 4.7. A language over an alphabet V is a λ-free context-free language iff it is represented by a regular pseudoautomaton with set of inputs V.

Proof. Since a finite groupoid pseudoautomaton is regular, by Theorems 4.1, 4.5 (assertion 1°) and 4.6, it will be sufficient to prove that a language is λ-free and context-free iff it is represented by a finite groupoid pseudoautomaton.

First we prove the sufficiency of the condition.

Let $\overline{\mathbf{A}}=(\mathbf{A}, A_F)$ be a finite groupoid pseudoautomaton, i.e. $\mathbf{A}=\text{ind}\,\mathscr{A}$, where $\mathscr{A}=\langle\tilde{A}, V, \xi\rangle$ is a finite designed groupoid. Using $\overline{\mathbf{A}}$ one can construct a λ-free context-free grammar $\Gamma(\overline{\mathbf{A}})$ of Chomsky normal form in the next way: $\Gamma(\overline{\mathbf{A}})=$ $=(A, V, A_F, \Pi)$, where A and V are the nonterminal and terminal alphabets, respectively. Moreover, A_F is the set of initial letters and Π is the set of productions for which $\Pi=\Pi_1\cup\Pi_2$ where

$$\Pi_1 = \{a \to v | v\in V,\ a\in A,\ \xi(v) = a\},$$

and

$$\Pi_2 = \{a \to a_1 a_2 | a, a_1, a_2\in A,\ a = a_1 \cdot a_2 \text{ in } \tilde{A}\}.$$

It can be shown that the language generated by $\Gamma(\bar{\mathbf{A}})$, in symbols $\mathcal{L}(\Gamma(\bar{\mathbf{A}}))$, equals to $\mathcal{L}(\bar{\mathbf{A}})$. For this purpose it is enough to show that for arbitrary $a\in A$ and $p\in V^+$, $a\Rightarrow^* p$ is valid iff there exists a t from $[V]$ for which $W(t)=p$ and $\bar{\alpha}(t)=a$ where α is the transition function of $\bar{\mathbf{A}}$. The proof is by induction. First, it can be seen easily that if $p=v$ and $t=v$ for an arbitrary v from V, then $a\Rightarrow^* p$ holds iff $\bar{\alpha}(t)=a$. Furthermore, let us suppose that for $p_i\in V^+$, $a_i\in A$, $1\leqq i\leqq k$, $k\geqq 1$, it have been shown that $a_i\Rightarrow^* p_i$ is valid iff there exists a $t_i\in[V]$ for which $W(t_i)=p_i$ and $\bar{\alpha}(t_i)=a_i$ hold. But then, there is a sequence of productions from Π_2: $a\to b_k a_k$, $b_k\to b_{k-1}a_{k-1}$, ..., $b_2\to b_1 a_1$, and a production from Π_1: $b_1\to v$ such that by a succesive application of them we get the derivation $a\Rightarrow^* v a_1 a_2 \dots a_k$, and by $a_i\Rightarrow^* p_i$, $a\Rightarrow^* v p_1 p_2 \dots p_k = p$, $p\in V^+$ iff there is a sequence of identities $a=b_k\cdot a_k$, $b_k=b_{k-1}\cdot a_{k-1}$, ..., $b_2=b_1\cdot a_1$, $b_1=\xi(v)$ in $\tilde{A}$, from which we get $a=(\dots((\xi(v)\cdot a_1)\cdot \cdot a_2)\cdot\dots)\cdot a_k=\alpha v(a_1 a_2\dots a_k)$, and for $a_i=\bar{\alpha}(t_i)$, $a=\bar{\alpha}(t)$, where $t=v(t_1)(t_2)\dots(t_k)$, and moreover $W(t)=v p_1 p_2\dots p_k=p$.

To prove the necessity of the condition, let us suppose that L is a λ-free context-free language over V. Then there exists a grammar in Chomsky normal form generating L. From this grammar, by the method applied to the proof of Theorem 3.1 of Part Three in [9] (with the difference that here the set S_0' contains the empty subset of S_0, as well) we can get an equivalent grammar Γ of the form (A, V, A_F, Π). For the set of productions Π we have $\Pi=\Pi_1\cup\Pi_2$, where Π_1 contains productions of the form $a\to v$ only ($a\in A$, $v\in V$) such that for each v from V there is exactly one production of this form, while Π_2 contains productions of the form $a\to a_1 a_2$ only ($a, a_1, a_2\in A$) such that for each (a_1, a_2) from A^2 there is exactly one production of this form. From these properties of Γ it follows that there is a finite groupoid pseudoautomaton $\bar{\mathbf{A}}$ with $\Gamma=\Gamma(\bar{\mathbf{A}})$.

NOTE. If for a λ-free context-free grammar $\Gamma=(N, V, N', \Pi)$ we introduce the grammar $\Gamma^T=(N, V\cup\pi, N', \Pi^T)$ where Π^T is obtained from Π by substituting every production $a\to a_1 a_2\dots a_k$ ($a\in N, k\geqq 1, a_i\in N\cup V$, $1\leqq i\leqq k$) in Π by the production $a\to a_1(a_2(\dots(a_k)))$ then the language $\mathcal{L}(\Gamma^T)$ generated by Γ^T is a subset of $[V]$. Let us call $\mathcal{L}(\Gamma^T)$ *the set of trees generated by* Γ. It was shown in [3] that the following assertion is valid: *a set of trees is pseudoregular iff it is the set of trees generated by a λ-free context-free grammar.* Since $W(\mathcal{L}(\Gamma^T))=\mathcal{L}(\Gamma)$ is valid, our Theorem 4.7 is now a consequence of this assertion. (Moreover, for $\Gamma(\bar{\mathbf{A}})$ from the proof of Theorem 4.7, $\mathcal{L}(\Gamma(\bar{\mathbf{A}})^T)=\mathcal{T}(\bar{\mathbf{A}})$ is valid.)

5. Relations between various types of the sets of trees

We shall say that a set of trees is *regular* iff it is represented by a such modification of our finite pseudoautomaton that if the transition function, the set of inputs and the set of states are denoted by α, V, and A, respectively, than for any $v\in V$, αv: $\{p | p\in A^*,\ \lg(p)\in K_v\}\to A$, where K_v is a finite nonempty subset of ω. It can be checked that this definition of regular sets of trees is equivalent to the definition of recognizable sets in [11].

Each regular set of trees is pseudoregular. (It can be seen by adding to our modified finite pseudoautomaton a new state b and mapping by αv all remaining words from $(A\cup\{b\})^*$ into it. The pseudoautomaton obtained by this procedure is regular.) The opposite is not true, i.e. there exist such sets of trees which are pseudoregular but not regular (for example $[V]$ for an alphabet V).

If a set of trees is pseudoregular it is recognizable by definition. However, there are recognizable sets of trees which are not pseudoregular. To show it, let us take an alphabet by a single letter v. Then each subset of the set

$$\{t\,|\,t = v\underbrace{(v)(v)\dots(v)}_{k \text{ times}} = v(v)^k,\ k\in\omega\},$$

is represented by a pseudoautomaton which has at most three states. Therefore, selecting a subset T of this set, for which $W(T)$ is not context-free, we get a recognizable set of trees and it is not pseudoregular by Theorem 4.7.

To finish these discussions, we demonstrate that for any alphabet V there are subsets of $[V]$ which are not recognizable. Let us suppose that v is an element of V and define a subset U of $[V]$ in the next way:

(1) $v\in U$;

(2) if $t\in U$, then $v(t)\in U$;

(3) the elements of U are those and only those which we get from (1) and (2) in a finite number of steps.

Every recognizable subset of U is regular, therefore, it is pseudoregular. Selecting from U a subset S for which $W(S)$ is not context-free, we get a set which is not recognizable.

6. Conclusion

From our point of view, we shall now answer to the question: what is the connection between pseudoautomata and automata?

The importance of connected pseudoautomata follows from Theorem 4.2. By the assertion 2° (a) of Theorem 4.5 every pseudoautomaton of this kind is a homomorphic image of a connected groupoid pseudoautomaton, and therefore (for Theorem 4.1) equivalent to it.

Let $\mathbf{\bar{A}}=(\alpha\langle V, A\rangle, A_F)$ be a connected groupoid pseudoautomaton, i.e. $\alpha\langle V, A\rangle = =\text{ind}\,\mathscr{A}$ for some connected designed groupoid $\mathscr{A}=\langle \tilde{A}, V, \xi\rangle$. Moreover, let $\tilde{A}$ be a monoid (semigroup with identity). Then, from the associativity it follows, that for each $p\in V^+$ the set $T(p)=\{t\,|\,t\in[V], W(t)=p\}$ has the property, that the whole set is represented by a single state of $\mathbf{\bar{A}}$. (This means that from $t_1, t_2\in T(p)$ it follows $\bar{\alpha}(t_1)=\bar{\alpha}(t_2)$. It can be proved by induction on $\lg(p)$.) For this reason it may be chosen a representative from $T(p)$ which is simpler than other members of this set, and only it must be represented by the pseudoautomaton $\mathbf{\bar{A}}$. If $p=v_1v_2\dots v_k$ $(k\geqq 1, v_i\in V, 1\leqq i\leqq k)$, then this representative can be $t=v_1(v_2(\dots(v_k)\dots))$ where arities of symbols $v_1, v_2, \dots, v_{k-1}$ equal to 1 and of v_k to 0. However, the situation becomes yet more simpler, if arities of each v from V equal to 1 and the other arities are ignored because they are unnecessary. But, it needs the introducing of a nullary symbol Λ which is not in V and whose realization is the identity e of the monoid $\tilde{A}$. Then the representative of $T(p)$ is the tree $v_1(v_2(\dots(v_k(\Lambda))\dots))$. By these modifications we got from $\bar{\mathbf{A}}$ a (connected) automaton in the sense of [4], [5], with initial state e. Now, an arbitrary homomorphic image of this automaton is a (connected) automaton too, its initial state is the image of e under the homomorphism, and moreover these automata are equivalent. (The first of them is a "monoid" automaton, but the second is an arbitrary one.)

By this interpretation, we got that the (ordinary) automaton is a simplification of the pseudoautomaton for the case when its groupoid is a monoid, and conversely, the concept of the pseudoautomaton is such a generalization of the concept of the automaton where its monoid is replaced with an arbitrary groupoid.

Abstract

The notion of a pseudoalgebra and that of a pseudoautomaton are introduced in a paper by THATCHER (1967). In this work it is shown that with a pseudoalgebra and with a pseudoautomaton a groupoid can be associated, in the same way as to a unary universal algebra and to an automaon a monoid can be corresponded.

FACULTY OF CIVIL ENGINEERING
24000 SUBOTICA
YUGOSLAVIA

References

[1] BRAINERD, W. S., The minimalization of tree automata, *Inform. and Control*, v. *13*, 1968, pp. 484—491.
[2] FERENCI, F., A new representation of context-free languages by tree automata, *Found. Control Engrg.*, v. 1, 1976, pp. 217—222.
[3] FERENCI, F., *Generalized automata*, Dissertation, University of Novi Sad, 1977 (in Serbocroatian).
[4] GÉCSEG, F. and I. PEÁK, *Algebraic theory of automata*, Akadémiai Kiadó, Budapest, 1972.
[5] Глушков, В. М., Абстрактная теория автоматов, *Успехи матем. наук*, 16:5 (101), 1961, pp. 3—62.
[6] GRÄTZER, G. *Universal algebra*, D. van Nostrand Company, inc. Princeton, 1968.
[7] Мальцев, А. И., Алгебраические системы, Москва, 1970.
[8] Медведев, Ю. Т., О классе событий, допускающих представление в конечном автомате, *Сборник «Автоматы»*, Москва, 1956, pp. 385—401.
[9] SALOMAA, A., *Formal languages*, Academic Press, New York, 1973.
[10] THATCHER, J. W., Characterizing derivation trees of context-free grammars through a generalization of finite automata theory, *J. Comput. System. Sci.*, v. 1, 1967, pp. 317—322.
[11] THATCHER, J. W., Generalized² sequential machine maps, *J. Comput. System. Sci.*, v. 4, 1970, pp. 257—287.
[12] THATCHER, J. W. and J. B. WRIGHT, Generalized finite automata theory with an application to a decision-problem of second-order logic, *Math. Systems Theory*, v. 2, 1968, pp. 57—81.

(Received August 29, 1979)