KYBERNETIKA ČÍSLO 5, ROČNÍK 5/1969

# The Theory of Regular Events I

IVAN M. HAVEL

**This paper is a survey of the present state of the theory of regular events and regular expressions and of their application to finite automata. Part I of the paper presents an abstract mathematical approach to regular events: the theory of Kleenean algebras. Part II will be mainly concerned with regular expressions and their application to the synthesis of finite automata. The references to both parts, constituting almost exhaustive bibliography of this field, are annexed to the Part II.**

## 1. INTRODUCTION

### 1.1

Regular events were brought to attention by the famous logician S. C. Kleene 13 years ago [42]. Since that time a great number of scientists have contributed to the theory. (The pioneering papers in this field are beside [42] the following: [3], [19], [50], [51], [52], [59], [90], [103]). The reason for the great interest in the theory of regular events may be found in its purely mathematical aspects on one hand and in its applicability to finite automata description and synthesis on the other hand. We illustrate the latter feature by the following simple construction:

Let $\mathscr{A}$ be a complete finite automaton, processing input words over some alphabet $X$ and producing corresponding output words over an alphabet $Y$. Let us look at this automaton as at a *black-box* provided with some *behaviour*. How the behaviour of $\mathscr{A}$ can be defined mathematically? The most plausible way to do this is to define the behaviour of $\mathscr{A}$ as the sequential function $\varphi$ realized by $\mathscr{A}$, i.e. the mapping $\varphi : X^* \rightarrow Y^*$ of input words to output words. A little more difficult problem arises when we have to *describe* or *record* the behaviour of an automaton. This is e.g. the case when we are concerned with practical synthesis of automata. The sequential function $\varphi$ has an infinite domain as well as a range and thus it is not easy to find a general way of its full description (we omit the special cases which can be described by means of finite tables — even if they are most usual for engineers). One of possible approaches is the following: The two significant properties of sequential functions realized by finite automata, viz. the length preserving property: $l(\varphi(u)) = l(u)$ (or $l(\varphi(u)) = l(u) + 1$ for the case of Moore automaton) and the property of reproduction of initial segments of words: $\varphi(uv) = \varphi(u)\, v$, enable us to replace $\varphi$ by the so called *reaction function* $\phi : X^* \to Y$, which maps each input word to the last symbol of the corresponding output word. This function, having finite range, may be characterized by a finite number of the word sets $\phi^{-1}(y)$, $y \in Y$. The description problem of finite automata behaviour is thus transformed to the problem of describing *sets of words* (the so called *events*) *over some finite alphabet*. As we shall see (Sec. 4), the word sets corresponding to finite automata, i.e. the *regular events*, can be well described by means of special formulas, the so called *regular expressions*. This very fact makes the study of regular events and regular expressions very attractive for automata theoretists.

This paper is a survey of some topics of the theory of regular events and regular expressions. In Sec. 2 the attention is paid to the Kleene's theorem on regular events. In Sec. 3–5 the algebraic properties of regular events are studied from the abstract point of view by defining the abstract Kleenean algebra. The proper applications of the theory to finite automata will be the main subject of the second part of the paper.

### 1.2

Throughout the paper the notion of a *finite automaton* refers to the *complete initial Moore automaton with one binary output*, defined by the quintuplet*

$$\mathscr{A} := \langle X, S, s_0, \delta, F \rangle$$

with finite nonempty *input alphabet* $X$, finite *state set* $S$, *initial state* $s_0$, *transition function* $\delta : S \times X \to S$ and the set of *final states* $F \subseteq S$. Automata of this type are sometimes called *recognizing automata*, because they can classify the input words by means of the two-symbol output alphabet (YES and NO). Our restriction to such automata is very convenient for our purposes, because the behaviour of each such automaton is characterized by a single set of words. (The above restriction is principially without any lack of generality: all our considerations can be applied with slight modifications to uncomplete, noninitial and Mealy automata with general output.)

Finite automata can be usefully represented by directed state-graphs the vertices of which correspond to the states from $S$. The branches, labelled by the symbols of $X$, refllect the transition function $\delta$.

Let $X^*$ be the set of all words (finite sequences) over the alphabet $X := \{x_1, \ldots, x_k\}$,

* The symbol ":=" denotes equality by definition.

including $\lambda$, the *empty word* (or the word of zero length). We shall say that the word $w \in X^*$ *leads from* $s_i$ *to* $s_j$ iff* $\delta(s_i, w) = s_j$, $s_i$, $s_j \in S$ (we have extended $\delta$ to $\delta : S \times X^* \to S$ in the usual way). To any automaton $\mathscr{A}$ we can attribute a set

$$T(\mathscr{A}) := \{w \in X^* \mid \delta(s_0, w) \in F\}$$

(the set of all words leading from the initial state to some of the final states) called the *event recognized by the automaton* $\mathscr{A}$. This is precisely the set defining the behaviour of $\mathscr{A}$. The words of $T(\mathscr{A})$ will be said to be *accepted* by $\mathscr{A}$. Two automata $\mathscr{A}_1$ and $\mathscr{A}_2$ are *behaviouraly equivalent* iff $T(\mathscr{A}_1) = T(\mathscr{A}_2)$.

## 2. REGULAR EVENTS

### 2.1

Now we shall proceed to the investigation of word sets and their properties. Our point of departure is the set $X^*$. Let $u$, $v$ be two words. We consider the juxtaposition $uv$ of the words $u$ and $v$ as a special binary operation in $X^*$ (the *concatenation*). This operation is obviously associative and thus $X^*$ forms a *free semigroup*. Moreover as $X^*$ includes also the empty word $\lambda$, for which

$$\forall u \in X^*, \quad u\lambda = \lambda u = u,$$

the set $X^*$ is a *free monoid* with the unity element $\lambda$.

Any subset of $X^*$ i.e. any element of the power set $\mathscr{P}(X^*)$ is called an *event* (some authors use the term *language*). Especially the following sets are events: $\emptyset$ (the empty event), $\Lambda := \{\lambda\}$ (the *silent event*). $\{x\}$ for any $x \in X$ (the *elementary events*), $X$ (the *alphabet*) and $X^*$ (the *universal event*). The sets of finite cardinality are *finite events*. ($X^*$ is an example of an infinite event.)

### 2.2

We can define some special operations on events:

(a) The *set operations*, esp.

| | |
|---|---|
| *set union* | $A \cup B$; |
| *set intersection* | $A \cap B$; |
| *set complement* | $\bar{A} = X^* - A$; |
| *relative set complement* | $A - B$; |
| *symmetric difference* | $A \triangle B$ etc. |

* If and only if.

All the set operations are related to the *Boolean algebra of subsets* of $X^*$. Accordingly, the inclusion $\subseteq$ is a lattice ordering in $\mathscr{P}(X^*)$ with $\emptyset$ and $X^*$ as the minimal and maximal elements respectively. 

(b) The *dot operation* (*concatenation*)

$$A \,.\, B \quad \text{or} \quad AB := \{w \in X^* \mid w = uv,\ u \in A,\ v \in B\}$$

(especially $A\emptyset = \emptyset A = \emptyset$, $A\Lambda = \Lambda A = A$). $\mathscr{P}(X^*)$ with the dot operation (and $\Lambda$) forms a *monoid*.

The dot operation yields also the powers of sets:

$$A^0 := \Lambda\,, \quad A^n := A \,.\, A^{n-1}\,.$$

(c) The *dagger* and *star operation*:

$$A^\dagger := \bigcup_{n=1}^{\infty} A^n\,, \quad A^* = \bigcup_{n=0}^{\infty} A^n = \Lambda \cup A^\dagger\,.$$

(especially $\emptyset^\dagger = \emptyset$, $\emptyset^* = \Lambda$). This definition is justified by the completeness of $\mathscr{P}(X^*)$ as a lattice. We see that our notation of $X^*$ is in accordance with the definition of star.

(d) The *quotients* [29]:

*right quotient* of $A$ by $B$:

$$A/B := \{w \in X^* \mid \exists u \in B,\ wu \in A\}\ ;$$

*left quotient* of $A$ by $B$:

$$B \setminus A := \{w \in X^* \mid \exists u \in B,\ uw \in A\}\ ;$$

*two-sided quotient* of $A$ by $B$, $C$:

$$B \setminus A/C := \{w \in X^* \mid \exists u \in B,\ \exists v \in C,\ uwv \in A\}\,.$$

As we shall see (cf. 6.2) the quotients (especially with respect to single words) are important in the synthesis of automata.

(e) The *reverse*:

$$\overleftarrow{A} := \{\overleftarrow{w} \mid w \in A\}\,,$$

where $\overleftarrow{w}$ denotes the word $w$ written backward.

The operations union, dot and star are called the *Kleenean operations*.

2.3

Our principal concept will be that of regular events:

**Definition.** An event $A$ is regular iff it is either finite or made up from finite events by means of a finite application of the Kleenean operations union, dot and star.

It is obvious that the definition of regularity may be based only on elementary events $\{x\}$, $x \in X$, and $\emptyset$ instead of on all finite events.

By the definition the family of regular events is closed with respect to Kleenean operations, as is the family of all events. The former, is, however a countable subfamily of the latter (which is uncountable).

The importance of the regular events lies in their connection to events recognized by finite automata:

2.4

**Kleene's Theorem.** *An event may be recognized by a finite automaton iff it is regular.*

An outline of the proof: The regularity of an event $T(\mathscr{A})$ recognized by an automaton $\mathscr{A}$ follows from the properties of the state-graph for $\mathscr{A}$. It may be proved (by induction over the number of states in $S$) that the set of words leading from whatever state to any other state is always regular. $T(\mathscr{A})$ is a finite union of such sets. Conversely, the existence of some automaton $\mathscr{A}$ to any regular event $A \subseteq X^*$ such that $T(\mathscr{A}) = A$ may be proved constructively by induction over the number of regular operators in $A$: The elementary events are trivially recognizable. If $A = T(\mathscr{A}_1)$, and $B = T(A_2)$ then three automata $\mathscr{A}_3$, $\mathscr{A}_4$, $\mathscr{A}_5$ may be constructed so that $T(\mathscr{A}_3) = A \cup B$, $T(\mathscr{A}_4) = AB$ and $T(\mathscr{A}_5) = A^*$. Their construction uses direct products and powers of automata (in the direct product and power the states of the resulting automaton are pairs and sets of original states resp.).

The analysis and synthesis procedures (see Sec. 9) may serve as an alternative proof of the Kleene's theorem.

2.5

A construction, quite similar to that of automata recognizing $A \cup B$, $AB$ and $A^*$, may be used for proving recognizability and hence regularity of some other operations:

**Theorem.** *If $A$, $B$ are regular and $C$, $D$ arbitrary events then the events $\bar{A}$, $A \cap B$ (and also the results of other set operations), $A/C$, $C \setminus A$, $C \setminus A(D)$ and $\overleftarrow{A}$ are also regular.*

It should be pointed out that in the case of quotients for non-regular $C$ or $D$, the proof is not generally effective (the procedure yields an automaton with undetermined initial state; cf. [73]). Stearns and Hartmanis [73] show further special regularity preserving operations as e.g. removing arbitrary halves of words and making some specific changes (errors) in words.

The following operations also preserve regularity:

$(A)^{\exists} := \{w \in X^* \mid$ there exists an initial segment of $w$ belonging to $A\}$,
$(A)^{\forall} := \{w \in X^* \mid$ all initial segments of $w$ belong to $A\}$ (clearly $(A)^{\exists} = AX^*$, $(A)^{\forall} = \overline{\bar{A}X^*}$). These operations represent the quantification in time (cf. [103]).

Last (but not least) the *superposition* [86], [121], also preserving regularity of events, should be mentioned. It is a $k + 1$-ary operation $(A_1, \ldots, A_k) \downarrow B$ – the replacement of symbols $x_1, \ldots, x_k$ in each word $w \in B$ by arbitrary words from $A_1, \ldots, A_k$ resp. (the symbol $x_i$ in different places of the same word $w$ may be replaced by different words from $A_i$). Special cases of superposition are *substitution* [103], *projection* [23] and *permutation.*

**2.6**

The well-known examples of non-regular events are the sets $\{0^n1^n \mid n \geqq 0\}$ and $\{1^{n^2} \mid n \geqq 0\}$ over the alphabets $\{0, 1\}$ and $\{1\}$ respectively.

## 3. ABSTRACT KLEENEAN ALGEBRA

**3.1**

We have seen that the family of all events as well as its subfamily of regular events are closed with respect to some important operations. It seems to be plausible to introduce the concept of related abstract algebra.

We shall present one possible approach, based on three Kleenean operations (2.2). The abstract algebra will be defined by means of a set of axioms. These axioms are due to Salomaa [63], [64], who used them in a somewhat less general sense. Similar axiom system was proposed by Aanderaa [1] (Cf. also [32], 4.11).

**Definition.** Let a set $\mathscr{K}$ be given with two binary operations $\cup$ and $\cdot$ (*union* and *dot*), one unary operation $*$ (*star*) and one nullary operation or constant $\emptyset$ (*empty element* or *zero*).

Then the algebraic system

$$\mathscr{K} := \langle \mathscr{K}, \cup, \cdot, *, \emptyset \rangle$$

is called the *Kleenean algebra* whenever the following axioms are satisfied:

A1 $\quad a \cup a = a \quad$ (independent law for union),

A2 $\quad a \cup b = b \cup a \quad$ (commutative law for union)

A3 $\quad a \cup (b \cup c) = (a \cup b) \cup c$

A4 $\quad a(bc) = (ab)\, c \quad$ } (associative laws),

A5 $a(b \cup c) = ab \cup ac$
A6 $(a \cup b)\,c = ac \cup bc$ (distributive laws),

A7 $a\Lambda = a$ where $\Lambda := \emptyset^*$,

A8 $a\emptyset = \emptyset$ (properties of the unity and zero),

A9 $a \cup \emptyset = a$,

A10 $a^* = \Lambda \cup aa^*$
A11 $a^* = (\Lambda \cup a)\Lambda^*$ (special laws for star),

R If $b \cup \Lambda \neq b$ then $a = ba \cup c$ implies $a = b^*c$ (solution of equations).

We see that all the above axioms but one are equalities of terms over an alphabet consisting of variables and given operational symbols (we shall call them *regular terms*). The equalities are supposed to hold for arbitrary elements of $\mathscr{K}$ substituted for variables. Thus these equalities are the identities of the algebra (cf. Kuroš: Lekcii po obščej algebre, III.6). Other identities in $\mathscr{K}$ may be derived from A1 – A11 by means of a substitution, replacement and of a special rule of inference R.

In fact the definition of Kleenean algebra defines the whole primitive class of algebras differing not only in extent but also in possible additional identities not derivable from the above axioms.

**3.2**

Some examples of Kleenean algebra:

**Example 1.** $\mathscr{K}_{\text{ev}}$ – *Kleenean algebra of all events* over some alphabet $X$. $\mathscr{K}_{\text{ev}} := \mathscr{P}(X^*)$; the operations of the algebra are the Kleenean operations on events.

The verification of validity of A1 – A11 in $\mathscr{K}_{\text{ev}}$ is quite easy. The proof of validity of R is as follows (cf. [3], [11]): Let $A, B, C \subseteq X^*$, $A = BA \cup C$ and $\Lambda \notin B$. Clearly $C \subseteq A$ and $D \subseteq A \Rightarrow BD \subseteq A$ for every $D \subseteq X^*$. Hence $B^*C \subseteq A$. Now let $B^*C \cup E \subseteq A$ for some $E$, $E \cap B^*C = \emptyset$. Then $B^*C \cup E = B(B^*C \cup E) \cup C = BE \cup (\Lambda \cup B^*B)\,C \Rightarrow E \subseteq BE$. But the shortest word in $E$ yields a contradiction, because $\Lambda \notin B$. Thus $E = \emptyset$ and $A = B^*C$.

For given $X$ the algebra $\mathscr{K}_{\text{ev}}$ is a maximal Kleenean algebra of events: any other algebra of events over $X$ is its subalgebra. Thus for the proof that some $\mathscr{K} \subseteq \mathscr{P}(X^*)$ is a Kleenean algebra it is sufficient to show that $\mathscr{K}$ is closed with respect to $\cup$, $\cdot$, $*$, $\emptyset$. In the following examples this closure properties are provable without difficulties.

**Example 2.** $\mathscr{K}_{\text{rc}}$ – *Kleenean algebra of all recursive events* over some alphabet $X$.

**Example 3.** $\mathscr{K}_{\text{cf}}$ – *Kleenean algebra of all context-free languages* over $X$. (Cf. Ginsburg: The mathematical Theory of Context-free languages).

**Example 4.** $\mathscr{K}_{rg}$ – *Kleenean algebra of all regular events* over $X$. This algebra is closed with respect to $\cup$, $\cdot$, $*$ by the definition. 

**Example 5.** $\mathscr{K}_{(2)} := \{\emptyset, \Lambda\}$ – *trivial two-element algebra.*

**Example 6.** $\mathscr{K}_{(1)} := \{\emptyset\}$ – *trivial one-element algebra.*

For the fixed alphabet $X$ clearly

$$\mathscr{K}_{(1)} \subset \mathscr{K}_{(2)} \subset \mathscr{K}_{rg} \subset \mathscr{K}_{cf} \subset \mathscr{K}_{rc} \subset \mathscr{K}_{ev},$$

where "$\subset$" means the relation of being proper subalgebra. Beside above special event algebras there is a great number of Kleenean event algebras $\mathscr{P}(A^*)$ for each $A \subseteq X^*$.

### 3.3

Generally, every algebra with elements which are subsets of a finitely generated free monoid and the operations interpreted as the set operations described previously, will be called the *event algebra*. (We might generalize the concept of event for subsets of an arbitrary, not necessarily free, semigroup or monoid [11], [33], [34] and investigate corresponding generalized event algebras with appropriate modifications of the axiomatic system.)

### 3.4

The following two examples are non-event Kleenean algebras. (Proofs of the validity of the axioms are quite straightforward.)

**Example 7.** Any finite set of integers including 0 and 1 under operations

$$a \cup b := \max(a, b),$$

$$a \,.\, b := \begin{cases} b & \text{if } a = 1 \text{ or } b = 0, \\ a & \text{otherwise}, \end{cases}$$

$$a^* := \begin{cases} 1 & \text{if } a = 0, \\ a & \text{otherwise}, \end{cases}$$

$$\emptyset := 0.$$

This algebra is idempotent ($a \,.\, a = a$). It is also an example of a nontrivial finite Kleenean algebra.

 **Example 8.** The set of all real numbers $\geqq 0$ together with $+\infty$, under operations

$$a \cup b := \min(a, b),$$
$$a \,.\, b := a + b,$$
$$a^* := 0,$$
$$\emptyset := +\infty,$$

This algebra is commutative $(ab = ba)$.

**3.5**

Some useful identities derivable within the Kleenean algebra are:

$$\left.\begin{array}{lr} (1) & \Lambda a = a \\ (2) & \emptyset a = \emptyset \end{array}\right\} \quad \text{(reverse to A7, A8)},$$

$$\left.\begin{array}{lr} (3) & a^{**} = a^* \\ (4) & aa^* = a^*a \\ (5) & a^*a^* = a^* \\ (6) & a^* \cup a = a^* \end{array}\right\} \quad \text{(properties of } *),$$

$$\left.\begin{array}{lr} (7) & (a \cup b)^* = (a^* \cup b^*)^* \\ (8) & (a \cup b)^* = (a^*b^*)^* \\ (9) & (a \cup b)^* = (a^*b)^*\, a^* \\ (10) & (a \cup b)^* = (a^* \cup b)^* \end{array}\right\} \quad \text{(reduction of the so called star-height -- see 11.2)},$$

$$(11) \qquad (a \cup b)^* \cup a = (a \cup b)^*,$$

$$(12) \qquad (ab)^*\, a = a(ba)^*.$$

Let us define a very useful relation in $\mathscr{K}$ corresponding to the set inclusion in $\mathscr{K}_{ev}$:

$$a \subseteq b \quad \text{iff}_{df} \quad a \cup b = b.$$

The following rules are derivable:

$$(13) \qquad a \subseteq b \Rightarrow a^* \subseteq b^*,$$

$$(14) \qquad a = \Lambda \cup ba \Rightarrow a \supseteq b^*,$$

$$(15) \qquad a = \Lambda \cup ab \Rightarrow a \supseteq b^*.$$

(Due to (3), (6) i.e. $a \subseteq a^*$ and (13) the operation $*$ is a closure operation.)

**3.6**

When using the rule R of Kleenean algebra we always have to check the condition $b \cup \Lambda \neq b$ i.e. $\Lambda \nsubseteq b$. The following construction may be useful.

Let us define $\mathscr{K}^{\Lambda} := \{a \in \mathscr{K} \mid \Lambda \subseteq a\}$. Then $\mathscr{K}^{\Lambda}$ is the least set with the following properties: 

(i) $$\forall a \in \mathscr{K}, \quad a^* \in \mathscr{K}^{\Lambda};$$

(ii) $$\forall a \in \mathscr{K}, \quad \forall b \in \mathscr{K}^{\Lambda}, \quad a \cup b \in \mathscr{K}^{\Lambda};$$

(iii) $$\forall a, b \in \mathscr{K}^{\Lambda}, \quad ab \in \mathscr{K}^{\Lambda}.$$

$\mathscr{K}^{\Lambda}$ is thus some sort of an ideal in $\mathscr{K}$. $\mathscr{K}^{\Lambda} \cup \{\emptyset\}$ is a subalgebra of $\mathscr{K}$.

**3.7**

Let us return to the definition of Kleenean algebra. We shall mention some slight modifications of Kleenean algebra, which may have some justification when applied to automata.

(a) ($\mathscr{K}'$) First of all we may leave out the axioms A8 and A9 and define the more general Kleenean algebra $\mathscr{K}' := \langle \mathscr{K}', \cup, \cdot, *, \Lambda \rangle$ with $\Lambda$ as a primitive nullary operation. Evidently $\mathscr{K}$ is a special case of $\mathscr{K}'$.

(b) ($\mathscr{K}^{\Lambda}$) If we now add the new axiom

A 12 $$a \cup \Lambda = a,$$

we obtain the special algebra $\mathscr{K}^{\Lambda} := \langle \mathscr{K}^{\Lambda}, \cup, \cdot, *, \Lambda \rangle$, which was just mentioned above (3.6). The example of $\mathscr{K}^{\Lambda}$ is e.g. the set of all (all regular) $\Lambda$-events (events containing the empty word), described by Janov [12]. Note that as a consequence of A12 the axiom R may be deleted. The following identities hold in $\mathscr{K}^1$:

(16) $$(a \cup b)^* = (ab)^*,$$

(17) $$(ab)^* a = (ab)^*,$$

(18) $$a(ab)^* = (ab)^*,$$

(19) $$a \subseteq ab,$$

(20) $$b \subseteq ab.$$

(c) ($\mathscr{K}^{(\dagger)}$) On the other hand we may define the algebra without $\Lambda$ [113]. For this we must, however, replace the star by the dagger operation $\dagger$. The algebra will be $\mathscr{K}^{(\dagger)} := \langle \mathscr{K}^{(\dagger)}, \cup, \cdot, \dagger \rangle$ and the only sufficient axioms A1 – A6, together with

A10′ $$a^{\dagger} = a \cup aa^{\dagger}$$

and

R′ $$a = ba \cup c \Rightarrow a = b^{\dagger}c \cup c.$$

(The axiom system A1 – A6, A10′, R′ is independent.)

(d) ($\mathscr{K}^{\Omega}$) In the Kleenean algebra $\mathscr{K}$ the existence of some maximal element is not generally guaranteed. But we may define $\mathscr{K}^{\Omega} := \langle \mathscr{K}^{\Omega}, \cup, \cdot, *, \emptyset, \Omega \rangle$ adding to A1 – A11, R two new axioms:

A13 $$a \cup \Omega = \Omega ,$$

A14 $$a\Omega = \Omega \quad \text{for} \quad a \neq \emptyset .$$

In $\mathscr{K}_{\mathrm{ev}}$, $\mathscr{K}_{\mathrm{rc}}$, $\mathscr{K}_{\mathrm{cf}}$, $\mathscr{K}_{\mathrm{rg}}$ the maximal element is the universal event: $\Omega = X^*$.

(e) ($\mathscr{K}^{0}$) Sometimes we may omit even the star as a primitive operation. The algebra $\mathscr{K}^{0} := \langle \mathscr{K}^{0}, \cup, \cdot, \Lambda \rangle$ with axioms A1 – A7 and (1) is the so called semi-lattice ordered semigroup (cf. [85]; the ordering is given by $\subseteq$). If the semi-lattice is complete (i.e. if also infinite unions exist) and completely distributive, the star operation may be defined:

$$a^* := \bigcup_{n=0}^{\infty} a^n$$

as well as the maximal element:

$$\Omega := \bigcup_{A \in \mathscr{K}} A .$$

This is e.g. the case of $\mathscr{K}_{\mathrm{ev}}$.

## 4. FINITARY KLEENEAN ALGEBRAS. REGULAR EXPRESSIONS

### 4.1

Now we shall proceed to a special class of Kleenean algebras which proves to have a principal importance in automata theory.

Let $\mathscr{K}$ be a Kleenean algebra. We shall call a *generating set* of $\mathscr{K}$ any set $G$ of its elements so that $\mathscr{K}$ is an algebraic closure of $G$ with respect to the operations union, dot, star and zero:*

$$\mathscr{K} := [G; \cup, \cdot, *, \emptyset] ,$$

An element $x$ in a Kleenean algebra $\mathscr{K}$ is an *atom* iff $x \neq \Lambda$ and for every $a, b \in \mathscr{K}$.

(1) $$x = a \cup b \Rightarrow a = \emptyset \quad \text{or} \quad b = \emptyset \quad \text{or} \quad a = b ;$$

(2) $$x = ab \Rightarrow a = \Lambda \quad \text{or} \quad b = \Lambda .$$

The atoms are not (nontrivially) decomposable and thus every generating set of $\mathscr{K}$ must contain all atoms.

* In the following, square brackets denote the algebraic closure of a given set with respect to given operations.

A generating set consisting only of atoms is called a *basis* in $\mathscr{K}$. If a Kleenean algebra has a basis, then this basis is unique.

**Definition.** A Kleenean algebra is *finitary* (*normal*) iff it has a finite generating set (a finite basis).

The algebra $\mathscr{K}_{\mathrm{rg}}$ is an example of a normal Kleenean algebra (its basis is $X$). No other Kleenean event algebra (over $X$) containing $X$ possesses a basis. On the other hand $\mathscr{K}_{\mathrm{cf}}$ is an example of a relatively "small" but non-finitary Kleenean algebra.*

Let us note that every finitary Kleenean algebra possesses a maximal element (satisfying A13, A14). It is clearly the element $\Omega = (\bigcup_{g \in G} g)^*$, where $G$ is a finite set of generators.

By the definition, every normal algebra is finitary. The converse, however, is not, in general, true:

**Example 9.** The set of all non-negative integers together with $+\infty$; the operations are defined similarly as in Example 8 (3.4).

This algebra has a single generator 1, which is, however, not an atom: $1 \cup 2 = 1$.

### 4.2

The next very important theorem, essentially due to Salomaa [64], asserts the property of normal Kleenean algebras which we shall call a *completeness with resp. to* $\mathscr{K}_{\mathrm{rg}}$.

**Theorem.** *Any equality between regular terms valid in* $\mathscr{K}_{\mathrm{rg}}$ *(independently on* $X$*) is valid in every normal Kleenean algebra.*

In other words any equality valid in $\mathscr{K}_{\mathrm{rg}}$ is derivable within A1 – A11 and R with possible application of the fact that there exists a finite basis in $\mathscr{K}_{\mathrm{rg}}$.

Proof of the theorem would be only a slight adaptation of the Salomaa's proof of the completeness of this axiomatic system $F_1$ [64].

The parallel theorem may be stated for the already mentioned modified algebra $\mathscr{K}^{(\dagger)}$ (with respect to $\mathscr{K}_{\mathrm{rg}}^{(\dagger)} := \{A \in \mathscr{K}_{\mathrm{rg}} \mid A \neq \emptyset,\ \lambda \notin A\}$. The completeness for this case is proved in another Salomaa's paper [113].

### 4.3

As a matter of fact, the axiomatic system (the set of identities A1 – A11 with the special rule of inference R) for Kleenean algebra was developed as a useful tool for

* It follows from one general result of Dr Gruska.

 deriving and proving equalities in the algebra of regular events $\mathscr{K}_{rg}$. In this connection one may ask, whether the axiom R, which is not an identity, is necessary. In other words, whether there exists some finite system of identities, equivalent to A1 – A11, R. The answer is negative. As Redko [111] has shown, there is no finite complete set of identities for a special case of one-letter event algebra $\mathscr{K}'$ and thus neither for the algebra of regular events in general ($\mathscr{K}$ or $\mathscr{K}'$) – cf. also [32], 4.10. The role of the rule R is to replace an infinite set of identities by the generative power of R. As it is known to the author, no infinite set of identities has been described till now (Redko [110] has presented such infinite set for the commutative Kleenean algebra.)

The reader may remember that in the algebra $\mathscr{K}^\Lambda$ (3.7(b)) with the unit element $\Lambda$ satisfying $a \cup \Lambda = a$, the rule R has vanished. Janov [120] has proved the completeness of the finite set of identities A1 – A6, (3), (5), (10), (11), (16) – (20) with respect to $\mathscr{K}^\Lambda_{rg} := \{A \in \mathscr{K}_{rg} \mid \lambda \in A\}$. On the other hand, adding to $\mathscr{K}^\Lambda_{rg}$ at least one event consisting of a single non-empty word $w \in X^*$ (and then taking the closure with resp. to $\cup$, $\cdot$, $*$) yields an algebra similar to $K'$ without a finite complete set of identities (Janov [123]).

The mentioned completeness of normal Kleenean algebra with respect to $\mathscr{K}_{rg}$ is in close relation to the following statement: Any Kleenean algebra $\mathscr{K}_{rg}$ of regular events is isomorphic to some free normal Kleenean algebra $\mathscr{K}$: it is sufficient to take the alphabet $X$ of $\mathscr{K}_{rg}$ (or better to say the set $\{x_1\}, \ldots, \{x_k\}$ of elementary events in $\mathscr{K}_{rg}$) as a free basis of the algebra $\mathscr{K}$.

## 4.4

Let us consider the most significant property of finitary Kleenean algebra – the possibility of being described by means of regular expressions. Finitary algebra has a finite set of generators so that it is sufficient to take a finite set of symbols denoting the generators, a few operational symbols $\cup$, $*$, $\emptyset$ (and $\Lambda$ which otherwise is not necessary) and parentheses to describe uniquely any element of the algebra:

**Definition.** Let $X := \{x_1, \ldots, x_k\}$ be a finite nonempty set of symbols and let $\cup$, $*$, $\emptyset$, $\Lambda$, ( , ) be another symbols not in $X$.

*Regular expressions* over $X$ are defined recursively as follows:

1. Each of the symbols $\emptyset$, $\Lambda$, $x_1, \ldots, x_k$ is a regular expression;
2. If $\alpha$ and $\beta$ are regular expressions, so are $(\alpha \cup \beta)$, $(\alpha\beta)$ and $\alpha^*$.
3. No string is a regular expression unless it is formed by 1 and 2 in a finite number of steps (we admit, however, some abbreviations in expressions).

Regular expressions may be also considered as elements of a free algebra under $\cup$, $*$, $\emptyset$ and juxtaposition. We shall use usual conventions omitting superfluous parantheses (thus we shall e.g. write $\alpha\beta \cup \gamma$ instead of $((\alpha\beta) \cup \gamma))$. Note that we use three different types of letters in connection with Kleenean algebra: $a, b, c, \ldots$ for elements of abstract algebra, $A, B, \ldots$ for events i.e. subsets of $\mathscr{P}(X^*)$ and $\alpha, \beta, \ldots$ for regular expressions. 

Let $\mathscr{K}$ be a finitary Kleenean algebra with $G := \{g_1, \ldots, g_k\}$ as its generating set. Let $X := \{x_1, \ldots, x_k\}$ be a finite set of symbols. We define the *interpretation* of regular expressions over $X$ as a mapping $| \; |$ to $\mathscr{K}$ recursively as follows:

$$1. \quad |\emptyset| = \emptyset \; ; \quad |\Lambda| = \Lambda \; ; \quad |x_i| = g_i \quad (i = 1, \ldots, k) \; ;$$
$$2. \quad |\alpha \cup \beta| = |\alpha| \cup |\beta| \; ; \quad |\alpha\beta| = |\alpha| . |\beta| \; ; \quad |\alpha^*| = |\alpha|^* \, .$$

Thus for an arbitrary expression $\alpha$ we may find the unique element $|\alpha| \in \mathscr{K}$. On the other hand for an arbitrary element $a \in \mathscr{K}$ there exists at least one expression $\alpha$ so that $|\alpha| = a$ (the mapping $| \; |$ is onto). In fact, there is an infinite number of such expressions to each $a \in \mathscr{K}$. This is a situation quite analogous to that of Boolean expressions and Boolean functions. But here the problem of equivalence of regular expressions is much more intricate (see Sec. 8). For the present we shall use only the associative law for $\cup$ and for juxtaposition. (Thus we shall write $\alpha\beta\gamma$ instead of $((\alpha\beta)\,\gamma)$ and $\alpha \cup \beta \cup \gamma$ instead of $((\alpha \cup \beta) \cup \gamma)$.)

Due to the Kleene's theorem, regular events play the principal role in the behavioural theory of finite automata. The algebra of regular events is finitary and thus we may describe the behaviour of any finite automaton by means of a regular expression.

On Table 1 we present some illustrative examples of events described by regular expressions over the alphabet $\{0, 1\}$.

When dealing with regular expressions it is useful to accept some abbreviations. For example:

| | | |
|---|---|---|
| $X$ | for | $x_1 \cup x_2 \cup \ldots \cup x_k$, |
| $\tilde{x}$ | for | $\bigcup_{x_i \neq x} x_i$, |
| $\alpha^n$ | for | $\alpha\alpha \ldots \alpha$ ($n$times), |
| $\alpha^\exists$ | for | $\alpha X^*$ (existential quantifier, cf. 2.5), |
| $\mathsf{C}_n(x_i)$ | for | $(\tilde{x}_i^* x_i)^{n*} \, \tilde{x}_i^*$ (counter mod $n$), |
| $\alpha \subseteq \beta$ | for | $\alpha \cup \beta = \beta$. |

## 5. KLEENE–BOOLEAN ALGEBRA. EXTENDED REGULAR EXPRESSIONS

### 5.1

We have seen that the class of all regular events is closed with respect to the complement and intersection (see 2.5). It implies that $\mathscr{K}_{rg}$ has a structure of Boolean

**Table 1.**

| Regular expression | Event |
|---|---|
| 0*1 | $\{0^n1 \mid n \geqq 0\}$ (the automaton indicates the first 1 on the input) |
| (01 ∪ 1)* | Set of words where every 0 is followed by 1 |
| ((0 ∪ 1) (0 ∪ 1))* | Set of words of even length (the automaton is autonomous and measures the time modulo 2) |
| (000* ∪ 1)* | Set of words without any occurence of a single 0 |
| (0 ∪ 1)* 00(0 ∪ 1)* | Set of words with at least one occurence of two consecutive 0's (the automaton indicates and remembers the occurence of 00) |
| (0*10*1)*0* | Set of words with even total number of 1's (the automaton counts modulo 2) |
| (0*1111*)* 0* | Set of words where all 1's occur in blocks of at least 3 members |
| (0*(11)* 10*(11)* 1)* 0* | Set of words where all 1's occur in blocks of odd length, the total number of blocks being even |
| ((1 ∪ 0) 000(1 ∪ 0) 0)* | Set of words where no 1 occurs at a time divisible by 2 or 3 |
| (00)* 0(0 ∪ 1) 1(11)* | $\{0^n1^m \mid n, m \geqq 1,\ n + m \text{ odd}\}$ (i.e. words of odd length with an occurence of both 0 and 1 but no occurence of 10) |

algebra. Because $\mathscr{K}_{rg}$ is a closure of $X$ with resp. to $\cup$, $\cdot$, $*$, and because $\cap$ and $^-$ do not bring anything new, we may surmise that the Boolean structure of $\mathscr{K}_{rg}$ is involved already in its Kleenean structure. Especially we may ask whether the operations $\cap$ and $^-$ are definable within Kleenean algebra. As it is known to the author this question has not yet been answered satisfactorily. We shall discuss the problem in this section.

In order to involve the Boolean structure into Kleenean algebra we have to extend the set of axioms by any complete set of identities of Boolean algebra. In order to reduce the total number of axioms we shall use the compact system due to Huntington:

A2 $$a \cup b = b \cup a\,,$$

A3 $$a \cup (b \cup c) = (a \cup b) \cup c\,,$$

A15 $\overline{\overline{a} \cup \overline{b}} \cup \overline{\overline{a} \cup b} = a$. 

**Definition.** Let $\mathcal{K}$ be a Kleenean algebra, $\mathcal{K} := \langle \mathcal{K}, \cup, \cdot, *, \emptyset \rangle$, with an additional unary operation $^{-}$ (*complement*). The algebraic system

$$\tilde{\mathcal{K}} := \langle \mathcal{K}, \cup, \cdot, *, ^{-}, \emptyset \rangle$$

is called the *Kleene-Boolean algebra* iff the additional axiom A15 is satisfied. (Indeed, the axioms A1 and A9 can be then omitted.)

Further we shall use the operation $\cap$ (*intersection*), definable in $\tilde{\mathcal{K}}$:

$$a \cap b := \overline{\overline{a} \cup \overline{b}}.$$

(In fact, the empty element $\emptyset$ is also definable in $\tilde{\mathcal{K}}$, $\emptyset := a \cap \overline{a}$, as well as the maximal element $\Omega := \overline{\emptyset}$. Thus $\tilde{\mathcal{K}} = \tilde{\mathcal{K}}^{\Omega}$.)

Kleene-Boolean algebra is thus a special case of a lattice-ordered semigroup ([84], cf. also Birkhoff; Lattice Theory, Ch. XIII).

**5.2**

Note that the dot operation does not, in general, distribute over intersection. With respect to this, the following result holds in event algebras [11]: Let us say that the element $a \in \tilde{\mathcal{K}}$ has the *prefix* resp. *suffix property* iff for any $b, c \in \tilde{\mathcal{K}}$

$$bc \cup b \subseteq c \quad \text{resp.} \quad cb \cup b \subseteq c \quad \text{implies} \quad a \subseteq \Lambda \quad \text{or} \quad b \subseteq \Lambda.$$

If $a$ has the prefix resp. suffix property, then

$$\forall b, c, \quad a(b \cap c) = ab \cap ac$$

resp.

$$(b \cap c)\, a = ba \cap ca.$$

**5.3**

Kleenean event algebras $\mathcal{K}_{rg}$, $\mathcal{K}_{rc}$ and $\mathcal{K}_{ev}$ are closed with resp. to the set complement and thus they are also Kleene-Boolean event algebras. On the other hand there is a well-known fact that $\mathcal{K}_{cf}$ is not closed with resp. to the set complement. This means that $\mathcal{K}_{cf}$ is not a Kleene-Boolean event algebra (which, however, tells us nothing about the possibility of $\mathcal{K}_{cf}$ being the Kleene-Boolean algebra under some non-set operation satisfying A15).

In general, Kleenean algebra need not be Kleene-Boolean algebra even in the finitary case. This may be seen in Example 9, where the only two possibilities $a \cup \overline{a} = a$ or $a \cup \overline{a} = \overline{a}$ lead to $a = \emptyset$ or $a = \Lambda$ which does not admit a non-trivial algebra.

However, the most usual case is that of normal Kleenean algebras, i.e. algebras with a finite basis $X = \{x_1, \ldots, x_k\}$ (the basic alphabet). For this case Salomaa and Tixier [65] presented two extensions of the axiom system A1 – A11, R to obtain an axiom system complete with respect to $\mathscr{K}_{rg}$ with the complement and intersection. In the first extension the following two identities are taken as axioms:

A16 $\quad \overline{x_1 a_1 \cup x_2 a_2 \cup \ldots \cup x_k a_k} = x_1 \bar{a}_1 \cup x_2 \bar{a}_2 \cup \ldots \cup x_k \bar{a}_k \cup \Lambda$ ;

A17 $\quad \overline{x_1 a_1 \cup x_2 a_2 \cup \ldots \cup x_k a_k \cup \Lambda} = x_1 \bar{a}_1 \cup x_2 \bar{a}_2 \cup \ldots \cup x_k \bar{a}_k$ .

In the second extension (cf. also [74]) the axioms A1, A2, A3 and A9 are replaced by a complete axiom system for any Boolean algebra (similarly as we did in sec. 5.1) and the following axiom and rule are added:

A18 $\quad \forall x \in X , \quad \Lambda \cap ax = \emptyset$ ;

R2 $\quad \forall x_i, x_j \in X \quad \text{if} \quad x_i \neq x_j \quad \text{then} \quad ax_i \cap ax_j = \emptyset$ .

Both extensions depend explicitely on the basic alphabet. The problem of their equivalence with our definition of the normal Kleene-Boolean algebra remains open, as well as the question of the independence of axioms. This last question is closely related to the problem of the definability of complement within the normal Kleenean algebra. It can be shown that in the case of free normal Kleenean algebra the complement is always definable e.g. by means of the identities A16, A17. For this purpose we have to prove only that the defined complement is really an operation (i.e. $\forall a\ \exists! \bar{a}$), and that it is unique. But as we shall see later (7.3, 8.1) any element of a free normal Kleenean algebra is uniquely characterized by a set of equations, and the identities (1) and (2) serve only as transformation rules which, applied to a set of equations for some element, yield a set of equations for its complement. Note that the construction of a complement by this definition involves essentially a synthesis and analysis of an automaton (9.1, 9.2).

## 5.4

It is very useful to use complement and intersection also as operators in the language of regular expressions (the definition of appropriate extended language of regular expressions is obvious). It makes possible to use all other operators of Boolean algebra, as well as further definable operators.

Some examples of useful operators follow:

$\alpha - \beta := \alpha \cap \bar{\beta}$ (relative complement),

$\alpha \triangle \beta := (\alpha \cap \bar{\beta}) \cup (\bar{\alpha} \cap \beta)$ (symmetric difference),

$\alpha^{\forall} := \overline{\bar{\alpha} X^*} = \overline{\bar{\alpha}^{\exists}}$ (universal quantification over time – cf. 2.5),

$\neg\alpha := \overline{X^*\alpha X^*}$ (the set of all words having no segment contained in $\alpha$),

$\alpha \to \beta := \overline{X^*\alpha\overline{\beta X^*}}$ (the set of all words having for every segment contained in $\alpha$ a segment immediately following contained in $\beta$)

$\alpha|^n := \alpha \cap \bigcup_{i=0}^{n} X^i$ (the set of all words from $\alpha$ of length less than or equal to $n$).

**5.5**

Generally speaking, the extensions of the language of regular expressions by means of new operators may be very desirable for applications to finite automata. Beside direct definitions of the above type there are also more complex definitions by recursion over other operators. For example the operator corresponding to the reverse (see 2.2(e)) of events may be defined as follows:

1. $\overleftarrow{\phi} = \phi\,;\quad \overleftarrow{x_i} = x_i\ (x_i \in X)\,;$

2. $\overleftarrow{\alpha \cup \beta} = \overleftarrow{\alpha} \cup \overleftarrow{\beta}\,;\quad \overleftarrow{\alpha^*} = (\overleftarrow{\alpha})^*\,;$

$\overleftarrow{\alpha\beta} = \overleftarrow{\beta}\overleftarrow{\alpha}\,;\quad \overleftarrow{\overleftarrow{\alpha}} = \alpha$

The second example of recursive definition may be that of the derivative (see 7.1). Any such recursive definition of a new operator should be provided with a proof of its existence and uniqueness. Usually it is done indirectly by proving that the corresponding operation in $\mathscr{K}_{\mathrm{rg}}$ preserves the regularity of events.

**5.6**

We conclude the theoretical approach to regular events by a short remark on other possibilities of definition of associated algebra. The question is, whether some other family of operations is capable (similarly as union, dot and star) of generating all regular events from some small set of simple events. We shall shortly mention three of possible approaches (the basic alphabet will be always $X := \{x_1, \ldots, x_k\}$).

(1) In his paper [103] Medvedev used four operations (we shall use our own notation): $\cup$, $\overline{\ }$, $^{\vee}$ and a simple substitution $[x_i \to x]\, A := (x_1, x_2, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_k) \downarrow A$ (cf. 2.5) and $2k + 2$ elementary events $X, X^2, X^*\{x_i\}, X \cup X^*\{x_i\}\, X$, $i = 1, \ldots, k$ (note that neither dot nor star is used at all). All $\Lambda$-free events are generated.

Next two approaches are due to Bodnarčuk [86].

(2) The only operation is the $k + 1$-ary superposition $(A_1, \ldots, A_k) \downarrow B$. The

 number of elementary events is $k + 3$:

$$\{x_i\} \quad (i = 1, \dots, k), \quad \{x_1 x_2\}, \quad \{x_1, x_2\}, \quad X^* - \Lambda .$$

Again all $\Lambda$-free events are generated.

(3) The operation is the substitution $(x_{i_1}, \dots, x_{i_k}) \downarrow A$ and elementary events are $2^{2k+1}$ of the so called $\mathscr{R}$-events

$$R_{KL} := \{w \in X^* \mid w = x_{i_1} x_{i_2}, \dots, x_{i_m}, m \geqq 1, x_{i_1} \in K, x_{i_m} \in L, (x_{i_j}, x_{i_{j+1}}) \in \mathscr{R}\},$$
$$K, L \subseteq X \quad \text{and} \quad R_{KL}^{\Lambda} := R_{KL} \cup \Lambda ,$$

where $\mathscr{R} \subseteq X \times X$ is an arbitrary but fixed relation.

Some general relating results are given by Janov [122] (for $\Lambda$-free events and $X$ possibly infinite). An operation is called invariant iff, roughly speaking, it (1) preserves the set of all symbols occuring in words of any event and (2) commutes with the superposition. The examples of invariant operations are union, dot, star and reverse. The intersection, complement and superposition itself are not invariant. The system of operations is called *strong* iff it generates all regular events from elementary events: letters of the alphabet $X$. Janov has proved:

(a) any strong system of invariant operations has at least three elements;
(b) all invariant (regularity preserving) operations can be derived from whatever strong system of invariant operations.

It implies the non-existence of finite set of identities for any invariant strong system.

(Received July 31st, 1968.)

# Teorie regulárních událostí I

IVAN M. HAVEL

Práce je přehledem současného stavu teorie regulárních událostí a regulárních výrazů a jejich použití v teorii konečných automatů. První část obsahuje obecný matematický přístup spočívající v zavedení abstraktní Kleeneho algebry, jejímž speciálním případem je pak algebra událostí. Je věnována pozornost zejména axiomatizaci a rozšíření Kleeneho algebry o operace Booleovy algebry.

Druhá část práce se bude zabývat převážně použitím regulárních výrazů k syntéze konečných automatů. K druhé části bude též připojena obsáhlá bibliografie této oblasti.

*Ing. Ivan M. Havel, Ústředí pro rozvoj automatizace a výpočetní techniky, Loretánské nám. 3, Praha 1.*