# Two Complete Axiom Systems for the Algebra of Regular Events

ARTO SALOMAA
*University of Turku, Turku, Finland*

*Abstract.* The theory of finite automata is closely linked with the theory of Kleene's regular expressions. In this paper, two formal systems for the algebraic transformation of regular expressions are developed. Both systems are consistent and complete; i.e., the set of equations derivable within the system equals the set of equations between two regular expressions denoting the same event. One of the systems is based upon the uniqueness of the solution of certain regular expression equations, whereas some facts concerning the representation theory of regular events are used in connection with the other.

## 1. *Introduction*

The problem concerning the axiomatization of the algebra of regular events has been proposed by many authors [3, 7, 8]. A negative solution has been obtained by V. N. Redko [10]. He has shown that if substitution is the only rule of inference, then no finite set of axioms is sufficient to yield a complete axiomatization of the algebra of events over the alphabet $\{x\}$. It follows that such an axiomatization is not possible for the algebra of events in general. However, this leaves one with the problem of constructing axiom systems with rules of inference stronger than the substitution rule.

In this paper, two such axiom systems are proposed. The feature characteristic for the system $F_1$ is that regular expression equations of the form $\alpha = \alpha\beta + \gamma$ are solvable by a rule of inference. In the system $F_2$, one is allowed to introduce iterations to regular expressions satisfying certain conditions. Both of the systems $F_1$ and $F_2$ are shown to be consistent and complete. Because the rules of inference are in $F_2$ stronger than in $F_1$, the completeness proof of $F_1$ is more complicated than that of $F_2$. Original versions of the systems $F_1$ and $F_2$ were given in [11] without the completeness proof of $F_1$. This exposition is self-contained in Sections 2–4, whereas some results from [7] and [11] are needed in Section 5.

## 2. *Definitions*

Let $X = \{x_1, \cdots, x_r\}$ be a finite nonempty set, called the (input) *alphabet*. The elements of $X$ are referred to as *letters*. Let $X^*$ be the free semi-group with identity generated by $X$, where one may write the operation multiplicatively and denote it by juxtaposition. The elements of $X^*$ are called *words* (over the alphabet $X$) and the subsets of $X^*$, *events* (over $X$). The symbol $\phi$ denotes the empty event. The identity of $X^*$, called the *empty word*, is denoted by $\phi^*$.

Consider finite strings consisting of the elements of $X$, the symbol $\phi$, the so-called *regular operators*, namely, *sum* $(+)$, *product* $(\cdot)$ and *star* $(^*)$, and parentheses. The notion of a *regular expression* (over the alphabet $X$) is defined recursively as follows:

(i) A string consisting of a single letter of the alphabet or a single $\phi$ is a regular expression.



Journal of the Association for Computing Machinery, Vol. 13, No. 1 (January, 1966), pp. 158–169

(ii) If $\alpha$ and $\beta$ are regular expressions, then so are $(\alpha + \beta)$, $(\alpha \cdot \beta)$ and $\alpha^*$.

(iii) Nothing else is a regular expression unless its being so follows from a finite number of applications of (i) and (ii).

In practice, the dot $(\cdot)$ is omitted. For convenience, parentheses are sometimes omitted, and the order of strength of the various operator signs is specified in the usual fashion: product is performed before disjunction and star before both product and disjunction. Thus, $\alpha + \beta\gamma^*$ has to be read $\alpha + (\beta(\gamma^*))$.

Every regular expression $\alpha$ denotes a subset $|\alpha|$ of $X^*$ according to the following conventions:

(i) The regular expression $x_i$, $(i = 1, \cdots, r)$, denotes the unit set of $x_i$. The regular expression $\phi$ denotes the empty set of words.

(ii) Let the sets denoted by the regular expressions $\alpha$ and $\beta$ be $|\alpha|$ and $|\beta|$. Then the regular expression $(\alpha + \beta)$ denotes the union $|\alpha| \cup |\beta|$. The regular expression $(\alpha\beta)$ denotes the set $\{ab \mid a \in |\alpha|, b \in |\beta|\}$. The regular expression $\alpha^*$ denotes the smallest set $|\alpha^*|$ of words, which contains the empty word and contains, for any $a \in |\alpha|$ and $b \in |\alpha^*|$, the word $ab$. $\alpha^*$ is called the *iteration* of $\alpha$.

The notation $\alpha \equiv \beta$ used here means that the regular expressions $\alpha$ and $\beta$ are identical, i.e., contain the same symbols in the same order. Moreover, the equation $\alpha = \beta$ between two regular expressions $\alpha$ and $\beta$ is *valid* if the sets $|\alpha|$ and $|\beta|$ are identical. Thus, although $(x_1 + x_2^*)^* = (x_1 + x_2)^*$ is a valid equation, one does not have $(x_1 + x_2^*)^* \equiv (x_1 + x_2)^*$.

It may be stated that a regular expression $\alpha$ possesses the *empty word property* (e.w.p.) if and only if one of the following conditions is satisfied:

(i) $\alpha \equiv \beta^*$, for some regular expression $\beta$.

(ii) $\alpha$ is a sum of regular expressions, one of which possesses e.w.p.

(iii) $\alpha$ is a product of regular expressions, each of which possesses e.w.p.

It is obvious that a regular expression $\alpha$ possesses e.w.p. if and only if the set $|\alpha|$ contains the empty word.

Ordered pairs $(\alpha, \beta)$ of regular expressions are also considered. Then by the sum $(\alpha, \beta) + (\gamma, \delta)$ is meant the ordered pair $(\alpha + \gamma, \beta + \delta)$, and by the product $(\alpha, \beta)\gamma$ is meant the ordered pair $(\alpha\gamma, \beta\gamma)$.

## 3. *The axiom system* $F_1$

A purely formal characterization of valid equations between regular expressions will now be given; i.e., no reference will be made to the sets denoted by regular expressions. For this purpose, the *axiom system* $F_1$ is introduced.

There are 11 axioms in the system $F_1$ :

$$A_1 \quad \alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma, \qquad A_7 \quad \phi^*\alpha = \alpha,$$

$$A_2 \quad \alpha(\beta\gamma) = (\alpha\beta)\gamma, \qquad A_8 \quad \phi\alpha = \phi,$$

$$A_3 \quad \alpha + \beta = \beta + \alpha, \qquad A_9 \quad \alpha + \phi = \alpha,$$

$$A_4 \quad \alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma, \qquad A_{10} \quad \alpha^* = \phi^* + \alpha^*\alpha,$$

$$A_5 \quad (\alpha + \beta)\gamma = \alpha\gamma + \beta\gamma, \qquad A_{11} \quad \alpha^* = (\phi^* + \alpha)^*.$$

$$A_6 \quad \alpha + \alpha = \alpha,$$

In $A_1$–$A_{11}$, $\alpha$, $\beta$ and $\gamma$ are arbitrary regular expressions. (In fact, the axioms are infinite axiom schemata.) There are two rules of inference.

R1 (Substitution). Assume that $\gamma'$ is the result of replacing an occurrence of $\alpha$ by $\beta$ in $\gamma$. Then from the equations $\alpha = \beta$ and $\gamma = \delta$ one may infer the equation $\gamma' = \delta$ and the equation $\gamma' = \gamma$.

R2 (Solution of equations). Assume that $\beta$ does not possess e.w.p. Then from the equation $\alpha = \alpha\beta + \gamma$ one may infer the equation $\alpha = \gamma\beta^*$.

A *proof* in the axiom system $F_1$ is a finite sequence of equations where each equation either is an axiom or follows by a rule of inference from the earlier equations in the sequence. An equation $\alpha = \beta$ is *derivable* within the system $F_1$, in symbols $\vdash_1 \alpha = \beta$, if there is a proof ending with the equation $\alpha = \beta$. The axiom system $F_1$ is said to be *consistent* if all derivable equations are valid. It is said to be *complete* if all valid equations are derivable. First the following Theorem 1 must be established.

THEOREM 1. *The axiom system $F_1$ is consistent.*

PROOF. It is obvious that all of the axioms $A_1$–$A_{11}$ are valid and that the rule R1 preserves validity. It is also easy to see (cf. [2] or [11]) that if a regular expression $\beta$ does not possess e.w.p. then the equation $\alpha = \alpha\beta + \gamma$ has only one solution, namely, $\alpha = \gamma\beta^*$. Therefore, the rule R2 preserves validity and thus Theorem 1 follows.

In the sequel, the substitution rule R1 and the results mentioned in the following lemma will be used without being explicitly referred to.

LEMMA 1. *Let $\alpha$, $\beta$, $\gamma$ and $\delta$ be arbitrary regular expressions. Then $\vdash_1 \alpha = \alpha$. If $\vdash_1 \alpha = \beta$, then $\vdash_1 \beta = \alpha$. If $\vdash_1 \alpha = \beta$ and $\vdash_1 \beta = \gamma$, then $\vdash_1 \alpha = \gamma$. If $\vdash_1 \alpha = \beta$ and $\vdash_1 \gamma = \delta$, then $\vdash_1 \alpha + \gamma = \beta + \delta$, $\vdash_1 \alpha\gamma = \beta\delta$ and $\vdash_1 \alpha^* = \beta^*$.*

The proof of Lemma 1 is straightforward, by R1 and $A_6$. (In fact, the rule R1 can be omitted if the sign of equality is regarded as belonging to the syntax language.)

In view of $A_1$ and $A_2$, we shall write sums and products of more than two regular expressions associatively, and use the customary product and $\sum$-notations. The notation

$$\vdash_1 (\alpha, \beta) = (\gamma, \delta)$$

is used here to mean that both $\vdash_1 \alpha = \gamma$ and $\vdash_1 \beta = \delta$.

## 4. *Completeness proof of $F_1$*

The subsequent proof of the completeness of the axiom system $F_1$ is based on two facts. Firstly, because of the rule R2, certain systems of equations possess a solution which is unique up to derivability within $F_1$. This is shown in Lemma 2. Secondly, if the equation $\alpha = \beta$ is valid, then such a system of equations can be constructed for the pair $(\alpha, \beta)$. This is a consequence of Lemma 4.

First let us derive two equations, obtained from $A_7$ and $A_8$ by changing the order of the factors on the left sides. By $A_8$,

$$\vdash_1 \phi\phi = \phi$$

and hence, by $A_9$,

$$\vdash_1 \alpha\phi\phi = \alpha\phi, \qquad \vdash_1 \alpha\phi = (\alpha\phi)\phi + \phi.$$

This implies, by R2, that

$$\vdash_1 \alpha\phi = \phi\phi^*.$$

Therefore, by $A_8$, we have

$$\vdash_1 \alpha\phi = \phi. \tag{1}$$

Consequently,

$$\vdash_1 \alpha = \alpha + \phi = \alpha + \alpha\phi = \alpha\phi + \alpha.$$

Hence, by R2,

$$\vdash_1 \alpha\phi^* = \alpha. \tag{2}$$

LEMMA 2. *Assume that $n$ is a natural number and*

$$\vdash_1 (\alpha_i , \beta_i) = \sum_{j=1}^{n} (\alpha_j , \beta_j)\gamma_{ij} + (\gamma_i , \gamma_i) \quad (i = 1, \cdots , n) \tag{3}$$

*where none of the regular expressions $\gamma_{ij}$ possesses e.w.p. Then $\vdash_1 \alpha_i = \beta_i$, for $i = 1, \cdots , n$.*

PROOF. The proof is by induction on the number $n$. If $n = 1$ then (3) has the form:

$$\vdash_1 \alpha_1 = \alpha_1\gamma_{11} + \gamma_1 , \qquad \vdash_1 \beta_1 = \beta_1\gamma_{11} + \gamma_1 .$$

Hence, by R2,

$$\vdash_1 \alpha_1 = \gamma_1(\gamma_{11})^* = \beta_1 .$$

Assuming that $n \geq 2$ and having established the lemma for the numbers $1, \cdots , n-1$, the equation resulting from (3) for $i = n$ can now be separated into two equations:

$$\vdash_1 \alpha_n = \sum_{j=1}^{n-1} \alpha_j\gamma_{nj} + \alpha_n\gamma_{nn} + \gamma_n$$

and

$$\vdash_1 \beta_n = \sum_{j=1}^{n-1} \beta_j\gamma_{nj} + \beta_n\gamma_{nn} + \gamma_n .$$

By applying R2 (and $A_3$), the following results are obtained:

$$\vdash_1 \alpha_n = \left(\sum_{j=1}^{n-1} \alpha_j \gamma_{nj} + \gamma_n\right) (\gamma_{nn})^* \tag{4}$$

and

$$\vdash_1 \beta_n = \left(\sum_{j=1}^{n-1} \beta_j \gamma_{nj} + \gamma_n\right) (\gamma_{nn})^*. \tag{5}$$

Now eliminate the pair $(\alpha_n , \beta_n)$ from (3) according to (4) and (5), and apply the commutative and distributive laws $A_3$–$A_5$ :

$$\vdash_1 (\alpha_i , \beta_i) = \sum_{j=1}^{n-1} (\alpha_j , \beta_j)(\gamma_{ij} + \gamma_{nj}(\gamma_{nn})^*\gamma_{in}) + (\epsilon_i , \epsilon_i) \quad (i = 1, \cdots , n-1),$$

where $\epsilon_i \equiv \gamma_i + \gamma_n(\gamma_{nn})^*\gamma_{in}$. Obviously, none of the coefficients $\gamma_{ij} + \gamma_{nj}(\gamma_{nn})^*\gamma_{in}$ possesses e.w.p. Therefore, by our inductive assumption,

$$\vdash_1 \alpha_i = \beta_i \qquad (i = 1, \cdots, n-1).$$

Hence, by (4) and (5), $\vdash_1 \alpha_n = \beta_n$. This completes the induction and also the proof of Lemma 2.

It may be stated that a regular expression $\alpha$ is *equationally characterized* if there is a finite number of regular expressions $\alpha_1, \cdots, \alpha_n$ such that $\alpha \equiv \alpha_1$ and

$$\vdash_1 \alpha_i = \sum_{j=1}^{r} \alpha_{ij}x_j + \delta(\alpha_i) \qquad (i = 1, \cdots, n), \quad (6)$$

where $\delta(\alpha_i) \equiv \phi$ or $\delta(\alpha_i) \equiv \phi^*$ and, for each $i$ and $j$, there is a $k$, $1 \leq k \leq n$, such that $\alpha_{ij} \equiv \alpha_k$.

It is obvious that in (6), $\delta(\alpha_i) \equiv \phi^*$ if and only if $\alpha_i$ possesses e.w.p. Furthermore, for each $i$, the sets $|\alpha_{ij}x_j|$, $(j = 1, \cdots, r)$, are disjoint. This implies, by Theorem 1, the following:

LEMMA 3. *Assume that the equation $\alpha = \beta$ is valid and*

$$\vdash_1 (\alpha, \beta) = \sum_{j=1}^{r} (\alpha_j, \beta_j)x_j + (\delta(\alpha), \delta(\beta)),$$

*where $\delta(\alpha) \equiv \phi$ or $\delta(\alpha) \equiv \phi^*$, and $\delta(\beta) \equiv \phi$ or $\delta(\beta) \equiv \phi^*$. Then $\delta(\alpha) \equiv \delta(\beta)$ and the equations $\alpha_j = \beta_j$, $(j = 1, \cdots, r)$, are valid.*

The next lemma is the most important tool in the completeness proof.

LEMMA 4. *Every regular expression is equationally characterized.*

PROOF. We follow the recursive definition of regular expressions presented in Section 2. Using axioms $A_6$–$A_9$, the following relations are obtained:

$$\vdash_1 \phi = \sum_{j=1}^{r} \phi x_j + \phi,$$

$$\vdash_1 x_i = \phi x_1 + \cdots + \phi^* x_i + \cdots + \phi x_r + \phi \qquad (i = 1, \cdots, r),$$

$$\vdash_1 \phi^* = \sum_{j=1}^{r} \phi x_j + \phi^*.$$

Thus, it may be concluded that the regular expressions $\phi$ and $x_i$, $(i = 1, \cdots, r)$, are equationally characterized. (For $\phi$ the corresponding set consists of $\phi$ alone, and for $x_i$ it consists of $x_i$, $\phi$ and $\phi^*$.)

Assume that the regular expressions $\alpha$ and $\beta$ are equationally characterized. This implies that there are regular expressions $\alpha_1, \cdots, \alpha_n$ where $\alpha_1 \equiv \alpha$ such that (6) holds. Furthermore, there are regular expressions $\beta_1, \cdots, \beta_m$ where $\beta \equiv \beta_1$ such that

$$\vdash_1 \beta_i = \sum_{j=1}^{r} \beta_{ij}x_j + \delta(\beta_i) \qquad (i = 1, \cdots, m), \quad (7)$$

where $\delta(\beta_i) \equiv \phi$ or $\delta(\beta_i) \equiv \phi^*$ and, for each $i$ and $j$, there is a $k$, $(1 \leq k \leq m)$, such that $\beta_{ij} \equiv \beta_k$. To complete the proof of Lemma 4, it suffices to show that the regular expressions $\alpha + \beta$, $\alpha\beta$ and $\alpha^*$ are equationally characterized.

We denote

$$\xi(u, v) \equiv \alpha_u + \beta_v \quad (u = 1, \cdots, n; \quad v = 1, \cdots, m).$$

Because we have

$$\vdash_1 \phi + \phi = \phi, \qquad \vdash_1 \phi + \phi^* = \phi^* + \phi = \phi^*, \qquad \vdash_1 \phi^* + \phi^* = \phi^*,$$

we obtain, using (6) and (7) (and the commutative and distributive laws),

$$\vdash_1 \xi(u, v) = \sum_{j=1}^{r} (\alpha_{uj} + \beta_{vj}) x_j + \delta(u, v),$$

where $\delta(u, v) \equiv \phi$ or $\delta(u, v) \equiv \phi^*$ and all of the regular expressions $\alpha_{uj} + \beta_{vj}$ are among the regular expressions $\xi$. Since $\xi(1, 1) \equiv \alpha + \beta$, this implies that $\alpha + \beta$ is equationally characterized.

Next let us consider regular expressions

$$\eta(u, v_1, \cdots, v_h) \equiv \alpha\beta_u + \alpha_{v_1} + \cdots + \alpha_{v_h}, \tag{8}$$

where $1 \leq u \leq m$, $h \geq 0$ and $1 \leq v_1 < v_2 < \cdots < v_h \leq n$. (Obviously, the number of the regular expressions (8) equals $m \cdot 2^n$.) Assume first that $\delta(\beta_u) \equiv \phi$. Then, by $A_3$–$A_6$, $A_9$ and (1), the following result may be obtained:

$$\vdash_1 \eta(u, v_1, \cdots, v_h) = \sum_{j=1}^{r} (\alpha\beta_{uj} + \alpha_{v_1 j} + \cdots + \alpha_{v_h j}) x_j + \delta(\eta), \tag{9}$$

where

$$\delta(\eta) \equiv \phi \quad \text{or} \quad \delta(\eta) \equiv \phi^*. \tag{10}$$

If $\delta(\beta_u) \equiv \phi^*$, then we obtain, by $A_3$–$A_6$, $A_9$ and (2), the result:

$$\vdash_1 \eta(u, v_1, \cdots, v_h) = \sum_{j=1}^{r} (\alpha\beta_{uj} + \alpha_{1j} + \alpha_{v_1 j} + \cdots + \alpha_{v_h j}) x_j + \delta(\eta), \tag{11}$$

where (10) is satisfied. Using axioms $A_3$ and $A_6$, the coefficients of $x_j$ on the right sides of (9) and (11) may be replaced by some regular expressions (8). Because $\eta(1) \equiv \alpha\beta$, it is concluded that $\alpha\beta$ is equationally characterized.

We denote, finally,

$$\xi(0) \equiv \alpha^*, \qquad \xi(u_1, \cdots, u_h) \equiv \alpha^*(\alpha_{u_1} + \cdots + \alpha_{u_h}), \tag{12}$$

where $h \geq 1$ and $1 \leq u_1 < u_2 < \cdots < u_h \leq n$. (Obviously, the number of the regular expressions (12) equals $2^n$.) By (6),

$$\vdash_1 \alpha = \sum_{j=1}^{r} \alpha_{1j} x_j + \delta(\alpha).$$

Hence, by $A_9$ or $A_{11}$,

$$\vdash_1 \alpha^* = \left( \sum_{j=1}^{r} \alpha_{1j} x_j \right)^*.$$

From this relation we obtain, by $A_{10}$,

$$\vdash_1 \xi(0) = \sum_{j=1}^{r} \alpha^* \alpha_{1j} x_j + \phi^*. \tag{13}$$

Assume that the regular expression $\alpha_{u_1} + \cdots + \alpha_{u_h}$ does not possess e.w.p. Then we obtain the relation:

$$\vdash_1 \xi(u_1, \cdots, u_h) = \sum_{j=1}^{r} \alpha^*(\alpha_{u_1 j} + \cdots + \alpha_{u_h j}) x_j + \phi. \tag{14}$$

If the regular expression $\alpha_{u_1} + \cdots + \alpha_{u_h}$ possesses e.w.p., then the following relation is derived:

$$\vdash_1 \xi(u_1, \cdots, u_h) = \sum_{j=1}^{r} \alpha^*(\alpha_{1j} + \alpha_{u_1 j} + \cdots + \alpha_{u_h j})x_j + \phi^*. \tag{15}$$

Again, using axioms $A_3$ and $A_6$, the coefficients of $x_j$ on the right sides of (13), (14) and (15) may be replaced by some regular expressions (12). Hence, by (12–15), it is concluded that $\alpha^*$ is equationally characterized. This proves Lemma 4.

*Remark 1.* Lemma 4 comprises two facts; namely, that the number of dissimilar derivatives of regular expressions is finite (cf. Theorem 5.2 in [4]) and that the corresponding characteristic equations are derivable within $F_1$. The definition of similarity presented in [4] has to be modified to include operations with $\phi$ and $\phi^*$. Otherwise, [4, Theorem 5.2] does not hold true.

A suitable position has now been reached to begin the proof of the completeness of $F_1$.

THEOREM 2. *The axiom system $F_1$ is complete.*

PROOF. Let $\alpha = \beta$ be an arbitrary valid equation. By Lemma 4, both $\alpha$ and $\beta$ are equationally characterized. Let the corresponding regular expressions be $\alpha_1, \cdots, \alpha_n$ and $\beta_1, \cdots, \beta_m$, where $\alpha \equiv \alpha_1$, $\beta \equiv \beta_1$ and the conditions (6) and (7) are satisfied. By Lemma 3, we obtain

$$\vdash_1 (\alpha, \beta) = (\alpha_1, \beta_1) = \sum_{j=1}^{r} (\alpha_{1j}, \beta_{1j})x_j + (\delta(\alpha), \delta(\alpha)),$$

where the pairs $(\alpha_{1j}, \beta_{1j})$ are among the pairs $(\alpha_h, \beta_k)$. By Lemma 3, we obtain similarly

$$\vdash_1 (\alpha_{1j}, \beta_{1j}) = \sum_{j=1}^{r} (\alpha'_{1j}, \beta'_{1j})x_j + (\delta(\alpha_{1j}), \delta(\alpha_{1j})) \quad (j = 1, \cdots, r),$$

where the pairs $(\alpha'_{1j}, \beta'_{1j})$ are among the pairs $(\alpha_h, \beta_k)$. The procedure is carried on until no new pairs appear. Thus, we can construct a set of pairs

$$(\alpha^{(1)}, \beta^{(1)}), (\alpha^{(2)}, \beta^{(2)}), \cdots, (\alpha^{(u)}, \beta^{(u)}) \tag{16}$$

with $u \leq mn$ and

$$\vdash_1 (\alpha^{(i)}, \beta^{(i)}) = \sum_{j=1}^{r} (\alpha_j^{(i)}, \beta_j^{(i)})x_j + (\gamma_i, \gamma_i) \quad (i = 1, \cdots, u), \tag{17}$$

where all of the pairs $(\alpha_j^{(i)}, \beta_j^{(i)})$ are among the pairs (16). By (1), we may write (17) in the form

$$\vdash_1 (\alpha^{(i)}, \beta^{(i)}) = \sum_{j=1}^{u} (\alpha^{(j)}, \beta^{(j)})\gamma_{ij} + (\gamma_i, \gamma_i) \quad (i = 1, \cdots, u),$$

where for each $i$ and $j$ either $\gamma_{ij} \equiv \phi$ or $\gamma_{ij} \equiv x_{j_1} + \cdots + x_{j_v}$ for some $v \geq 1$ and $1 \leq j_1 < \cdots < j_v \leq r$. Thus, none of the regular expressions $\gamma_{ij}$ possesses e.w.p. This implies, by Lemma 2, that

$$\vdash_1 \alpha^{(i)} = \beta^{(i)} \qquad (i = 1, \cdots, u).$$

In particular, we have $\vdash_1 \alpha = \beta$. Thus, Theorem 2 follows.

*Remark 2.* The completeness proof is constructive in the sense that, for any

valid equation $\alpha = \beta$, it gives a method to construct a proof of it. One may even compute an upper bound, depending on the number of the regular operators included in $\alpha$ and $\beta$, for the length of the shortest proof of $\alpha = \beta$. However, the general method is in most cases not economical. For simple derivations within $F_1$ of some of the most common regular expression equations, the reader is referred to [11].

As an illustration, let us now establish the derivability (within $F_1$) of the equation $\alpha_1 = \beta_1$, where

$$\alpha_1 \equiv (0 + 01 + 10)^* \quad \text{and} \quad \beta_1 \equiv (10 + 0^*01)^*0^*.$$

We denote

$$\alpha_2 \equiv \alpha_1 + \alpha_1 1 \equiv (0 + 01 + 10)^* + (0 + 01 + 10)^*1$$

and

$$\beta_2 \equiv \beta_1 + (10 + 0^*01)^*1 \equiv (10 + 0^*01)^*0^* + (10 + 0^*01)^*1.$$

By $A_{10}$, the following result is obtained:

$$\vdash_1 \alpha_1 = \phi^* + \alpha_1(0 + 01 + 10) = \phi^* + \alpha_2 0 + \alpha_1 01.$$

Because we have

$$\vdash_1 \alpha_1 = \phi^* + \alpha_1 0 + \alpha_1 01 + \alpha_1 10 = \phi^* + \alpha_1 0 + \alpha_1 0 + \alpha_1 01 + \alpha_1 10 = \alpha_1 0 + \alpha_1,$$

we may write

$$\vdash_1 \alpha_2 = \alpha_1 + \alpha_1 1 = \phi^* + \alpha_2 0 + \alpha_1 01 + \alpha_1 1$$
$$= \phi^* + \alpha_2 0 + (\alpha_1 0 + \alpha_1)1 = \phi^* + \alpha_2 0 + \alpha_1 1.$$

Using (2) and the relation $\vdash_1 0^* = \phi^* + 0^*0$, we obtain

$$\vdash_1 \beta_1 = (10 + 0^*01)^* + \beta_1 0 = \phi^* + (10 + 0^*01)^*(10 + 0^*01) + \beta_1 0$$
$$= \phi^* + \beta_2 0 + \beta_1 01$$

and

$$\vdash_1 \beta_2 = \beta_1 + (10 + 0^*01)^*1 = \phi^* + \beta_2 0 + \beta_1 01 + (10 + 0^*01)^*1$$
$$= \phi^* + \beta_2 0 + (\beta_1 0 + (10 + 0^*01)^*)1 = \phi^* + \beta_2 0 + \beta_1 1.$$

Thus we have established the relations

$$\vdash_1 (\alpha_1, \beta_1) = \phi^* + (\alpha_2, \beta_2)0 + (\alpha_1, \beta_1)01$$

and

$$\vdash_1 (\alpha_2, \beta_2) = \phi^* + (\alpha_2, \beta_2)0 + (\alpha_1, \beta_1)1.$$

Two successive applications of R2 give, finally, the result

$$\vdash_1 \alpha_1 = \beta_1 = 0^*(01 + 10^*0)^*.$$

The method used in this example is, in principle, the same as the one used in the completeness proof of $F_1$. However, the proof has been simplified by allowing the word 01 to appear as a coefficient.

*Remark 3.* Note that intersection and complement are not included in our

language of regular expressions (cf. [4] and [9]). Also note that it follows from the completeness of $F_1$ that all valid regular expression equations can be derived using Ghiron's rules (cf. [6]), provided $A_6$ and R1 are added to his rules. In addition, some of Ghiron's rules can be omitted (cf. [11, footnote on p. 19]).

*Remark 4.* Although it is easy to see that the "essential" axioms $A_{10}$ and $A_{11}$, as well as the rule R2 are independent, the problem of the independence of the remaining axioms is open. A related problem is whether or not R2 can be weakened, for instance, to the following form: If $\beta$ does not possess e.w.p. and $\alpha = \alpha\beta + \phi^*$ then $\alpha = \beta^*$. On the other hand, it is obvious that the axiom system $F_1'$, obtained from $F_1$ by reversing the order of factors in products appearing in $A_7$, $A_8$, $A_{10}$ and R2, is complete.

*Remark 5.* One may omit the symbol $\phi$ from the language of regular expressions. Then the events denoted by regular expressions do not contain the empty word. Therefore, $| \alpha^* |$ is defined to be the union of the sets $| \alpha^i |$, $i \geq 1$. (This definition of iteration is proposed in [5].) We mention without proof that a complete axiomatization for this restricted language consists of the axioms $A_1$–$A_6$ and

$$A_{10}' \qquad \alpha^* = \alpha^*\alpha + \alpha$$

and of the rules R1 and R2′: From the equation $\alpha = \alpha\beta + \gamma$ one may infer the equation $\alpha = \gamma\beta^* + \gamma$.

## 5. *The axiom system $F_2$*

The representation theory of finite automata gives various possibilities to construct complete axiom systems for the algebra of regular events. Such a system is presented in what follows. Its characteristic rule of inference is more difficult to apply than the rule R2.

For a regular expression $\alpha$, we define

$$C(\alpha) = 2^{s(\alpha)} + 2,$$

where $s(\alpha)$ is the number of letters occurring in $\alpha$. (Each letter is counted as many times as it occurs.)

We say that a regular expression $\alpha$ possesses *$\phi$-property* if and only if one of the following conditions is satisfied: (i) $\alpha \equiv \phi$, (ii) $\alpha$ is a disjunction of regular expressions, each of which possesses $\phi$-property, or (iii) $\alpha$ is a product of regular expressions, one of which possesses $\phi$-property.

Clearly, $\alpha$ possesses $\phi$-property if and only if $| \alpha |$ is the empty set.

The axioms in the system $F_2$ are $A_1$–$A_{10}$ and the following two equations (which correspond to (1) and (2)):

$$A_{12} \qquad \alpha\phi = \phi,$$

$$A_{13} \qquad \alpha\phi^* = \alpha.$$

The rules of inference are R1 and the following:

R3 (Introduction of iterations). For any regular expressions $\alpha$, $\beta$, $\gamma$ and $\delta$ where $\gamma$ does not possess $\phi$-property, the equation $\alpha + \beta\gamma^*\delta = \alpha$ may be inferred from the $C(\alpha) + 1$ equations:

$$\alpha + \beta\gamma^i\delta = \alpha \qquad (i = 0, 1, \cdots, C(\alpha)).$$

Derivability, consistency and completeness are defined as in Section 3. Derivability within $F_2$ is denoted by the symbol $\vdash_2$. The *iterative degree* of a regular expression is defined as follows:

(i) The letters of the alphabet and $\phi$ have iterative degree 0.

(ii) If $\alpha$ and $\beta$ have iterative degrees $i$ and $j$ then $\alpha\beta$ and $\alpha+\beta$ have iterative degree $\max(i, j)$.

(iii) If $\alpha$ has iterative degree $i$ then $\alpha^*$ has iterative degree $i+1$.

Three lemmas are presented first.

LEMMA 5. *Assume that, for each $i$ where $0 \leq i \leq C(\alpha)$, we have*

$$| \beta\gamma^i\delta | \subset | \alpha |, \tag{18}$$

*where $\gamma$ does not possess $\phi$-property. Then also $| \beta\gamma^*\delta | \subset | \alpha |$.*

PROOF. It suffices to show that the inclusion (18) holds for all values of $i \geq 0$. By the assumption, (18) holds for $i \leq C(\alpha)$. We make the following inductive hypothesis: (18) holds for $i \leq C(\alpha) + k$. To complete the induction, it must be proved that

$$| \beta\gamma^{C(\alpha)+k+1}\delta | \subset | \alpha |. \tag{19}$$

Assume the contrary: There is a word $P$ such that

$$P \in | \beta\gamma^{C(\alpha)+k+1}\delta |, \qquad P \notin | \alpha |. \tag{20}$$

Obviously, $P$ is of the form

$$P = P^{(\beta)}P_1^{(\gamma)} \cdots P_{C(\alpha)+k+1}^{(\gamma)}P^{(\delta)}, \qquad P^{(\beta)} \in | \beta |, \qquad P_i^{(\gamma)} \in | \gamma |, \qquad P^{(\delta)} \in | \delta |.$$

According to [7, Theorem 16], $| \alpha |$ can be represented in an automaton with $C(\alpha) - 1$ states. This implies that the states

$$s_0P^{(\beta)}P_1^{(\gamma)}, \cdots, s_0P^{(\beta)}P_1^{(\gamma)} \cdots P_{C(\alpha)+k}^{(\gamma)},$$

where $s_0$ is the initial state, cannot all be distinct. Hence, for some $u$ and $v$,

$$s_0P^{(\beta)}P_1^{(\gamma)} \cdots P_u^{(\gamma)} = s_0P^{(\beta)}P_1^{(\gamma)} \cdots P_u^{(\gamma)} \cdots P_{u+v}^{(\gamma)} \quad (1 \leq u < u + v \leq C(\alpha) + k).$$

Consequently,

$$s_0P' = s_0P, \tag{21}$$

where

$$P' = P^{(\beta)}P_1^{(\gamma)} \cdots P_u^{(\gamma)}P_{u+v+1}^{(\gamma)} \cdots P_{C(\alpha)+k+1}^{(\gamma)}P^{(\delta)} \in | \beta\gamma^i\delta |, \tag{22}$$

for some $i \leq C(\alpha) + k$. By (20) and (21), we obtain

$$P' \notin | \alpha |. \tag{23}$$

Because (22) and (23) contradict our inductive assumption, we may conclude that (19) holds. This proves Lemma 5.

LEMMA 6. *Assume that $k \geq 1$ and $\alpha$, $\beta_i$ $(1 \leq i \leq k+1)$ and $\gamma_i$ $(1 \leq i \leq k)$ are regular expressions such that none of the $\gamma_i$'s possesses $\phi$-property and*

$$\vdash_2 \alpha + \beta_1\gamma_1^{n_1}\beta_2 \cdots \beta_k\gamma_k^{n_k}\beta_{k+1} = \alpha$$

*whenever $0 \leq n_i \leq C(\alpha)$, $i = 1, \cdots, k$. Then also*

$$\vdash_2 \alpha + \beta_1\gamma_1^*\beta_2 \cdots \beta_k\gamma_k^*\beta_{k+1} = \alpha.$$

The proof is by induction on $k$. The reader is referred to the proof of [11, Theorem 11] where a similar argument is carried out.

LEMMA 7. *Assume that $\gamma$ is a regular expression with iterative degree 0 and $|\gamma| \subset |\alpha|$. Then $\vdash_2 \alpha + \gamma = \alpha$.*

The proof is performed by induction on the number of the regular operators in $\alpha$. For the details, the reader is referred to the proofs of [11, Theorems 4, 5]. In fact, the equation $\alpha + \gamma = \alpha$ can be derived in $F_2$ without an application of R3.

THEOREM 3. *The axiom system $F_2$ is both consistent and complete.*

PROOF. The consistency is due to the fact that, by Lemma 5, rule R3 preserves validity. To prove the completeness, we show by an induction on the iterative degree of $\beta$ that the inclusion $|\beta| \subset |\alpha|$ implies the relation

$$\vdash_2 \alpha + \beta = \alpha. \tag{24}$$

By Lemma 7, (24) follows if the iterative degree of $\beta$ equals 0. The inductive step is completed using Lemma 6 (and axioms $A_7$ and $A_{13}$ to take care of possible occurrences of $\phi^*$). The details can be found in the proof of [11, Theorem 12].

Let now $\alpha = \beta$ be an arbitrary valid equation. Then we have both $|\alpha| \subset |\beta|$ and $|\beta| \subset |\alpha|$ and, hence, $\vdash_2 \beta + \alpha = \beta$ and $\vdash_2 \alpha + \beta = \alpha$. Therefore, we obtain the relation $\vdash_2 \alpha = \beta$ which proves the completeness of the system $F_2$. Thus, Theorem 3 has been established.

*Remark 6.* Another complete axiom system $F_3$ can be constructed by using the fact that a regular event is completely characterized by a sufficiently large finite part of it. More specifically, define

$$C(\alpha, \beta) = 2^{s(\alpha, \beta)} + 1$$

where $s(\alpha, \beta)$ is the number of letters occurring in the regular expressions $\alpha$ and $\beta$. (Each letter is counted as many times as it occurs.) Furthermore, let $U(k)$ be the set of all words over $X^*$ with length less than or equal to $k$. Then $\alpha = \beta$ is valid if and only if

$$|\alpha| \cap U(C(\alpha, \beta)) = |\beta| \cap U(C(\alpha, \beta)).$$

*Acknowledgments.* Independently of [11], S. Aanderaa [1] has presented an axiom system almost identical to $F_1$ and independently presented a completeness proof for his system. Some parts of our completeness proof of $F_1$ (Lemma 2, Lemma 3 and the final argument of the proof) were essentially simplified by Aanderaa's proof. On the other hand, Aanderaa uses a result of Brzozowski [4], which is not quite correct (cf. remark 1 above). The argument given in the proof of Lemma 4 makes the usage of this result unnecessary. The author wishes to thank Professor P. C. Fischer for information concerning Aanderaa's work and Professor R. McNaughton for pointing out the possibility to obtain a finitary system $F_2$.

RECEIVED JULY, 1965

## REFERENCES

1. AANDERAA, S. On the algebra of regular expressions. Appl. Math., Harvard U., Cambridge., Jan. 1965, pp. 1-18. (Ditto)
2. ARDEN, D. N. Delayed logic and finite state machines. In *Theory of Computing Machine Design*, pp. 1-35. U. of Michigan Press, Ann Arbor. 1960.

3. Brzozowski, J. A. Canonical regular expressions and minimal state graphs for definite events. *Proc. Symp. Mathematical Theory of Automata*, Polytechnic Press, 1963, 529–561.
4. ——. Derivatives of regular expressions. *J. ACM 11* (1964), 481–494.
5. Copi, I. M., Elgot, C. C., and Wright, J. B. Realization of events by logical nets. *J. ACM 5* (1958), 181–196.
6. Ghiron, H. Rules to manipulate regular expressions of finite automata. *IRE Trans. EC-11* (1962), 574–575.
7. Gluschkow, W. M. *Theorie der abstrakten Automaten.* VEB Deutscher Verlag der Wissenschaften, Berlin, 1963.
8. Kleene, S. C. Representation of events in nerve nets and finite automata. In *Automata Studies*, Princeton U. Press, 1956, 3–41.
9. McNaughton, R., and Yamada, H. Regular expressions and state graphs for automata. *IRE Trans. EC-9* (1960), 39–47.
10. Redko, V. N. On defining relations for the algebra of regular events. *Ukrain. Mat. Ž.* 16 (1964), 120–126. (Russian)
11. Salomaa, A. Axiom systems for regular expressions of finite automata. *Ann. Univ. Turku.*, Ser. A I 75 (1964).