Yu. I. Yanov

1962

SovietRxiv

Abstract

Full Text

Yu. I. Yanov

ON IDENTICAL TRANSFORMATIONS OF REGULAR EXPRESSIONS

(Presented by Academician P. S. Novikov on 6 VI 1962)

Let there be an alphabet $A_s = \{a_1, \dots, a_s\}$, $1 \leq s \leq \aleph_0$. We shall consider sets of (finite) words in this alphabet. The sets $\{\lambda\}$ and $\{a_i\}$, $i = 1, \dots, s$, where λ is the empty word, will be called elementary. Consider the following three **basic** operations on sets of words:

1) **union** $A \vee B$—the set-theoretic sum of A and B;
2) **product** AB—the set of all words of the form ab, where $a \in A$, $b \in B$;
3) **closure**

$$\overline{A} = \{\lambda\} \vee A \vee AA \vee AAA \vee \cdots$$

Sets that can be obtained from elementary ones by a finite number of applications of the basic operations will be called **regular sets**. The class of all regular sets for the alphabet A_s will be denoted by $\mathfrak{M}_s$. It is known ([1]) that regular sets, and only they, are the sets representable by finite automata.

We shall omit braces in the notation of one-element sets; then every regular set will be given by some formula representing a superposition of the basic operations, whose arguments are the letters of the original alphabet A_s and, possibly, the symbol of the empty word λ. Such formulas will be called **regular expressions** in the alphabet A_s.

It is obvious that one and the same regular set can be represented by different regular expressions; thus, for example, the regular expressions

$$\overline{(a_1 a_2 \vee a_1 a_2)(a_3 a_4 a_5 \vee a_3 a_6)} \quad \text{and} \quad \overline{a_1 a_2 \vee a_3 a_6 \vee a_4 \vee a_5}$$

define one and the same set. In this connection there arises the problem of identical transformations of regular expressions in the alphabet A_s. It can be shown that this problem reduces to finding a complete system of identities for the algebra

$$\mathfrak{R}_s = \langle \mathfrak{M}_s,\ x \vee y,\ xy,\ \overline{x} \rangle.^*$$

If the empty word is disregarded, then the above problem reduces to finding a complete system of identities for the algebra

$$\mathfrak{R}'_s = \langle \mathfrak{M}'_s,\ x \vee y,\ xy,\ \overline{x} \rangle,$$

where $\mathfrak{M}'_s$ is the class of all those and only those regular sets (in the alphabet A_s) which contain the empty word.

In the present work finite complete systems of identities are constructed for the algebras $\mathfrak{R}'_s$, $1 \le s \le \aleph_0$. It turns out that for any p and q such that $2 \le p, q \le \aleph_0$, the algebras $\mathfrak{R}'_p$ and $\mathfrak{R}'_q$ are similar, i.e. have identical systems of identities.

Consider the following system of identities Σ':

0. $x = x$.
1. $x \vee x = x$.
2. $x \vee y = y \vee x$.
3. $x \vee (y \vee z) = (x \vee y) \vee z$.
4. $x(yz) = (xy)z$.
5. $x(y \vee z) = xy \vee xz$.
6. $(x \vee y)z = xz \vee yz$.
7. $\overline{\overline{x}} = \overline{x}$.
8. $\overline{x}x = \overline{x}$.
9. $\overline{x \vee y} \vee x = \overline{x \vee y}$.
10. $\overline{x \vee y} = \overline{\overline{x} \vee y}$.
11. $\overline{x \vee y} = \overline{xy}$.
12. $\overline{xyx} = \overline{xy}$.
13. $x\overline{xy} = \overline{xy}$.
14. $xy \vee y = xy$.
15. $xy \vee x = xy$.

* As V. N. Red' ko has shown, the algebras $\mathfrak{R}_s$ do not have finite complete systems of identities.

It is not hard to convince oneself that all identities of this system are true for any algebra $\mathfrak{R}'_s$, $1 \le s \le \aleph_0$. We shall show that for every algebra $\mathfrak{R}'_s$, where $2 \le s \le \aleph_0$, the system Σ' is complete, i.e., every identity true in $\mathfrak{R}'_s$ is derivable from Σ' by means of the following two rules of inference:

$$\alpha^0. \quad \frac{R_1(x_1, \dots, x_i, \dots, x_m) = R_2(x_1, \dots, x_i, \dots, x_m)}{R_1(x_1, \dots, S, \dots, x_m) = R_2(x_1, \dots, S, \dots, x_m)} \qquad \text{(substitution rule)},$$

$$\alpha^1. \quad \frac{R_1(S_1) = R_2,\ S_1 = S_2}{R_1(S_2) = R_2} \qquad \text{(replacement rule)}.$$

A complete system of identities for the algebra $\mathfrak{R}'_1$ can be obtained by adding to Σ' the following two identities:

$$xy = yx, \qquad \overline{xy} = \bar{x}\bar{y}.$$

We shall denote by the symbol $\vdash$ derivability by means of the rules α^0, α^1. We note that from Σ' the following identities and rules are derivable:

$$9'.\quad \overline{xy} \vee x = \overline{xy}. \qquad 9''.\quad \overline{xy} \vee \bar{x} = \overline{xy}.$$

$$\gamma.\quad \frac{R_1 \vee S_1 = S_1,\ R_2 \vee S_2 = S_2}{R_1 R_2 \vee S_1 S_2 = S_1 S_2}. \qquad \varepsilon.\quad \frac{R \vee S = S}{\bar{R} \vee S \vee P = S \vee P}.$$

$$\delta.\quad \frac{\bar{R} \vee S = S}{R \vee S = S}. \qquad \zeta.\quad \frac{R_1 \vee S = S,\ R_2 \vee S = S}{R_1 \vee R_2 \vee S = S}.$$

All subsequent arguments are valid for any algebra $\mathfrak{R}'_s$, where $2 \leq s \leq \aleph_0$.

Let R and S be formulas of the algebra $\mathfrak{R}'_s$. We shall write: $R \subseteq S$, if $R \vee S = S$ is an identity true for $\mathfrak{R}'_s$.

Obviously, in order to prove the completeness of the system Σ', it is sufficient for us to prove the following assertion:

1. For any formulas R and S, if $R \subseteq S$, then $\Sigma' \vdash R \vee S = S$.

Formulas of the form

$$\overline{x_{i_1} x_{i_2} \dots x_{i_m}},$$

where $i_1 < i_2 < \dots < i_m$, will be called **cycles**. Variables and cycles will be called **elementary chains**. Products of elementary chains will be called **chains**. The empty chain, as well as chains containing no cycles, will be called **simple chains**. Simple chains will also be regarded by us as words in the alphabet of variables.

For each formula R we define the set $\mathfrak{S}(R)$ of simple chains, which we shall call R-**words**.

1) $\mathfrak{S}(x_i) = \{\lambda, x_i\}$.

2) Suppose $\mathfrak{S}(R_1)$ and $\mathfrak{S}(R_2)$ have been defined; then:

2,1) $\mathfrak{S}(R_1 \vee R_2) = \mathfrak{S}(R_1) \vee \mathfrak{S}(R_2)$;

2,2) $\mathfrak{S}(R_1 R_2) = \mathfrak{S}(R_1)\mathfrak{S}(R_2)$;

2,3) $\mathfrak{S}(\bar{R}_1) = \overline{\mathfrak{S}(R_1)}$.

We shall call a **contraction** of a word S any word S' obtained from S by deleting some occurrences of letters. From the definition of the set $\mathfrak{S}(R)$ the following assertions follow immediately:

2. If $S \in \mathfrak{S}(R)$ and S' is a contraction of the word S, then $S' \in \mathfrak{S}(R)$.

3. If R is a simple chain, then $\mathfrak{S}(R)$ is the set of all contractions of the word R.

It is not hard to prove also the following assertion:

4. Suppose there is a formula $R \simeq R(x_1, \dots, x_m)^*$, and suppose $\mathfrak{S}(R) = \{R_i\}_i$. Denote by $\hat{R}$ (respectively $\hat{R}_i$) the value of the function R (respectively R_i) on the tuple $X_1, \dots, X_m$, where $X_j \in \mathfrak{R}'_s$, $j = 1, \dots, m$. Then

$$\hat{R} = \bigvee_i \hat{R}_i.$$

$*$ By the sign $\simeq$ we denote graphical identity.

From U4 the following assertion easily follows:

U5. The identity $R = S$ is true in $\mathfrak{R}'_s$ if and only if $\mathfrak{S}(R) = \mathfrak{S}(S)$.

Hence it follows immediately:

U6. $R \subseteq S$ if and only if $\mathfrak{S}(R) \subseteq \mathfrak{S}(S)$.

With the aid of identities 0–4, 8, 10, 11 it is easy to prove the following assertion:

U7. For every formula $R(x_1, \dots, x_m)$

$$\Sigma' \vdash \overline{R(x_1, \dots, x_m)} = \overline{x_1 \dots x_m}.$$

From U7 and identities 5, 6 it follows:

U8. For any formula R,

$$\Sigma' \vdash R = R_1 \vee \dots \vee R_m,$$

where $R_1, \dots, R_m$ are chains.

From U3 and U6 it follows:

U9. $x_{i_1} \dots x_{i_m} \subseteq x_{j_1} \dots x_{j_n}$ if and only if the word $x_{i_1} \dots x_{i_m}$ is a contraction of the word $x_{j_1} \dots x_{j_n}$.

From U9, δ, 1, 2, 11 it follows:

U10. $x_{i_1} \dots x_{i_m} \subseteq \overline{x_{j_1} \dots x_{j_n}}$ if and only if

$$\{x_{i_1}, \dots, x_{i_m}\} \subseteq \{x_{j_1}, \dots, x_{j_n}\}.$$

Similarly one obtains:

U11. $\overline{x_{i_1} \dots x_{i_m}} \subseteq \overline{x_{j_1} \dots x_{j_n}}$ if and only if

$$\{x_{i_1}, \dots, x_{i_m}\} \subseteq \{x_{j_1}, \dots, x_{j_n}\}.$$

From U9–U11, 9′, 9″ it follows:

U12. If R and S are elementary chains and $R \subseteq S$, then

$$\Sigma' \vdash R \vee S = S.$$

We shall say that a chain $R \simeq R_1 \dots R_m$ is a **subchain** of the chain $S \simeq S_1 \dots S_n$, where $R_1, \dots, R_m, S_1, \dots, S_n$ are elementary chains, if there exist $S_{i_1}, \dots, S_{i_m}$ such that

$$1 \leq i_1 < i_2 < \dots < i_m \leq n$$

and

$$R_k \subseteq S_{i_k}, \qquad k = 1, \dots, m.$$

From U12, γ and 14, 15 it follows:

U13. If R is a nonempty subchain of the chain S, then

$$\Sigma' \vdash R \vee S = S.$$

A chain R' obtained from a chain R by repeating certain cycles will be called an **extension** of the chain R.

In view of 8, we have:

U14. If R' is an extension of the chain R, then

$$\Sigma' \vdash R' = R.$$

By induction it is not difficult to prove the following assertion.

U15. If R is a chain, then $\mathfrak{S}(R)$ consists of all simple subchains of extensions of the chain R.

From U3 and U6 it follows:

U16. If R is a simple chain, then $R \subseteq S$ if and only if

$$R \in \mathfrak{S}(S).$$

We shall say that a chain R is **contained** in a chain S, if R is a subchain of some extension of the chain S.

From U13 and U14 it follows:

U17. If a nonempty chain R is contained in a chain S, then

$$\Sigma' \vdash R \vee S = S.$$

U18. If R and S are chains, then $R \subseteq S$ if and only if R is contained in S.

Proof. If R is contained in S, then $R \subseteq S$ by U17. Let $R \subseteq S$; we shall show that then R is contained in S.

1) Let R be an elementary chain.

1.1) If $R \simeq x_i$, the assertion is trivial.

1.2) Let $R \simeq x_{i_1} \dots x_{i_m}$. Then: 1,2,1) if in S there is a cycle

$$\overline{x_{j_1} \dots x_{j_n}}$$

such that

$$\{x_{i_1}, \dots, x_{i_m}\} \subseteq \{x_{j_1}, \dots, x_{j_n}\},$$

then, obviously, R is contained in S. 1,2,2) Suppose that for every cycle

$$\overline{x_{j_1} \dots x_{j_n}}$$

from S,

$$\{x_{i_1}, \dots, x_{i_m}\} \nsubseteq \{x_{j_1}, \dots, x_{j_n}\}.$$

Consider the R-word $R' \simeq (x_{i_1} \dots$

$\cdots x_{i_m})^\alpha$, where $\alpha > C$, C is the "length" of the chain S, i.e., the number of occurrences of variables in S. Let $S \simeq S_1 \dots S_l$, where $S_1, \dots, S_l$ are elementary chains. According to U15 and U9, U10, every S-word has the form

$$S' \simeq (x_{j_1}^{\beta_1} \dots x_{j_n}^{\beta_n})^{\alpha_1} \dots$$

$$\dots (x_{q_1}^{\gamma_1} \dots x_{q_k}^{\gamma_k})^{\alpha_p},$$

where $p \leq l$ and in the parentheses there stand variables belonging to one cycle from S, or else all exponents of the variables and of the parentheses are equal to 1. Suppose that R' is a subchain of the S-word S'. Since no set of variables in parentheses with exponent $\alpha_s > 1$ contains the set $\{x_{i_1}, \dots, x_{i_m}\}$, it follows that R' must be a subchain of an S-word S'' in which all exponents $\beta_1, \dots, \beta_n, \dots, \gamma_1, \dots, \gamma_k$ and $\alpha_1, \dots, \alpha_p$ are equal to 1. But the length of such a word does not exceed C, which contradicts $\alpha > C$. Thus, in view of U15, the R-word R' is not an S-word and therefore, by U6, $R \nsubseteq S$, which contradicts the assumption.

2) Suppose that the assertion is true for R' and R'', and $R \simeq R'R''$. Since $R' \subseteq R'R''$ and $R'' \subseteq R'R''$, we have $R' \subseteq S$ and $R'' \subseteq S$. Let $S \simeq S_1 \dots S_l$, where $S_1, \dots, S_l$ are elementary chains, and let p and q be such that

$$R' \subseteq S' \simeq S_1 \dots S_p, \quad R' \nsubseteq S_1 \dots S_{p-1}, \quad R'' \subseteq S'' \simeq S_{qS_{q+1}} \dots S_l, \quad R'' \nsubseteq S_{q+1} \dots S_l.$$

The following cases are possible.

2.1) $p < q$. Then, since by the induction hypothesis R' fits into S', and R'' fits into S'', R fits into the subchain $S'S''$ of the chain S, and consequently R fits into S.

2.2) $p = q$ and S_p is a cycle. Then, obviously, R fits into the expansion

$$S_1 \dots S_{pS} p \dots S_l$$

of the chain S.

2.3) $p = q$ and S_p is a variable. Let $P \simeq S_1 \dots S_{p-1}$, $Q \simeq S_{p+1} \dots S_l$, i.e., $S \simeq PS_{pQ} \simeq Px_{iQ}$. Since $R' \nsubseteq P$ and $R'' \nsubseteq Q$, by U6, U15, and U16 there exist an R'-word T_1 and an R''-word T_2 such that T_1 is not a subchain of any expansion of the chain P, and T_2 is not a subchain of any expansion of the chain Q. And since every expansion of the chain S has the form $P' x'_{iQ}$, where P' and Q' are expansions of the chains P and Q, the simple chain $T_1 T_2$ is not a subchain of any expansion of the chain S, i.e., by U15, the R-word $T_1 T_2$ is not an S-word. In view of U6 this means that $R \nsubseteq S$, which contradicts the assumption.

2.4) $p > q$. Then, obviously, there exist chains P and Q such that $S \simeq PQ$ and $R' \nsubseteq P$, $R'' \nsubseteq Q$. Analogously to what was done in item 2.3), one can show that then $R \nsubseteq S$. The assertion is proved.

From U17 and U18 it follows:

U19. *If R and S are chains and $R \subseteq S$, then $\Sigma' \vdash R \vee S = S$.*

U20. *Let $R \subseteq S$, where $S \simeq S_1 \vee \dots \vee S_n$ and $R, S_1, \dots, S_n$ are chains. Then $\Sigma' \vdash R \vee S = S$.*

Proof. In view of U19 and ε, it is enough for us to prove that if $R \subseteq S \simeq S_1 \vee \dots \vee S_n$, then there is an S_k, $1 \leq k \leq n$, such that $R \subseteq S_k$. Suppose the contrary, i.e., that $R \nsubseteq S_k$ for every $k = 1, \dots, n$. Then, according to U6, for every $k = 1, \dots, n$ there is an R-word R_k that is not an S_k-word. But in view of U15, for the set of R-words $\{R_k\}_{k=1}^{n}$ there is an R-word R' such that every R-word from this set is its subchain. Since for every $k = 1, \dots, n$, $R_k \subseteq R'$ and $R_k \nsubseteq S_k$, it follows that for every $k = 1, \dots, n$, $R' \nsubseteq S_k$. But, obviously, every S-word is an S_k-word for some k, $1 \leq k \leq n$, and therefore $R' \nsubseteq S$, i.e., in view of U16 $R \nsubseteq S$. The contradiction obtained proves the assertion.

Obviously:

U21. *If $R_1 \vee \dots \vee R_m \subseteq S$, then for every $i = 1, \dots, m$, $R_i \subseteq S$.*

From U8, U20, and U21, in view of ζ, U1 follows.

Received
30 V 1962

CITED LITERATURE

M. O. Rabin, D. Scott. I. B. M. J. Res. Develop., **3**, No. 2, 114 (1959).

Note: Figure translations are in progress. See original paper for figures.