Tracing the Decision Procedure of Regular Expressions Equality

Benoît Razet & Bodhayan Roy

TIFR

Automata, Concurrency and Timed Systems (ACTS) 2011
CMI, Chennai, India
2011-01-29

Regular expressions

Let $\Sigma = \{a, b, c \cdots, z\}$ is an alphabet.

$$\begin{array}{rcll} E, F & := & 0 & \\ & \mid & 1 & \\ & \mid & a & a \in \Sigma \\ & \mid & E + F & \\ & \mid & E \cdot F & \\ & \mid & E^* & \end{array}$$

Language model

Standard interpretation

We interpret inductiveley a regular expression into a language, that is a set of words.

$$
\begin{aligned}
L(0) &= \emptyset \\
L(1) &= \{\epsilon\} \\
L(a) &= \{\ a\ \} \\
L(E+F) &= L(E) \cup L(F) \\
L(E \cdot F) &= L(E) \cdot L(F) \\
&= \{ww' \mid w \in L(E) \wedge w' \in L(F)\} \\
L(E^*) &= \bigcup_{n \in \mathbb{N}} L(E)^n
\end{aligned}
$$

Some axioms

For any expression A, B and C, we have some basic identities:

$$\begin{aligned} (A+B)+C &= A+(B+C) \\ (AB)C &= A(BC) \\ A+0 &= A \\ &\vdots \\ A^{**} &= A^* \\ (A+B)^* &= A^*(BA^*)^* \\ (AB)^* &= 1+A(BA)^*B \\ &\vdots \end{aligned}$$

Open Problem (Kleene 1951)

Is there a ***complete axiomatisation*** *for the equality ?*

A negative result

Theorem (Redko 1964)

*Any **complete** axiomatisation for the equality of regular expressions must involve **infinitely many axioms**.*

Axiomatisation: a long history

Axiomatisations of regular languages has a long history because it is at the crossroads of algebra, computer science with theoretical and practical impact.

- Kleene (1956)
- Redko (1964)
- Salomaa (1966)
- Conway (1971)
- Kozen (1991)
- Pratt (1991)
- Bloom-Esik (1993)

Salomaa's axiomatisation

Idempotent Semiring

- $(+, 0)$ is a commutative semigroup

$$(A + B) + C = A + (B + C) \tag{1}$$
$$A + 0 = 0 + A = A \tag{2}$$
$$A + B = B + A \tag{3}$$

- $(\cdot, 1)$ is a semigroup

$$(A \cdot B) \cdot C = A \cdot (B \cdot C) \tag{4}$$
$$1 \cdot A = A \cdot 1 = A \tag{5}$$

- distributivity of $\cdot$ over $+$

$$A \cdot (B + C) = A \cdot B + A \cdot C \tag{6}$$
$$(A + B) \cdot C = A \cdot C + B \cdot C \tag{7}$$

- 0 is an annihilator for $\cdot$

$$0 \cdot A = A \cdot 0 = 0 \tag{8}$$

- Idempotence of $+$

$$A + A = A \tag{9}$$

Salomaa's Axiomatisation (1966)

Salomaa's axiomatisation consists in the axioms of idempotent semiring, plus the following axioms for Kleene star operator.

$$(A+1)^* = A^* \tag{S1}$$

$$1 + A \cdot A^* = A^* \tag{S2}$$

$$X = AX + B \text{ and } \epsilon \notin A \Rightarrow X = A^*B \tag{SI3}$$

The axiomatisation contains only **one** non-equational axiom, all the others are equational.

A complete axiomatisation

Theorem (Salomaa 1966)

Salomaa's Axiomatisation is complete.
For all expressions A and B, if $L(A) = L(B)$ then the equality $A = B$ can be proved using Salomaa's axiomatisation.

Proof sketch:

- Construct a system of equations conresponding to a deterministic automaton.

$$\begin{array}{lcl} (a+b)^*L & = & a \cdot X_1 + b \cdot X_2 + 1 \\ X_1 & = & a \cdot X_{a,1} + \cdots + z \cdot X_{z,1} + \delta_1 \\ \vdots & \vdots & \vdots \\ X_n & = & a \cdot X_{a,n} + \cdots + z \cdot X_{z,n} + \delta_n \end{array}$$

- Unify the 2 systems: duplicate equations.
- Use axiom (SI3) to eliminate variable and obtain a common expression C such that $A = C$ and $B = C$.

Kozen's axiomatisation

Relational model

Let D be a set. We consider **binary relations** on D. We consider the composition of relations $\circ$.

$$\rho \circ \rho' = \{(x, z) \mid \exists y, (x, y) \in \rho \wedge (y, z) \in \rho'\}$$

Suppose you are given some relations R_a associated to each letter $a \in \Sigma$. We interpret regular expressions as relations as the following:

$$\begin{aligned} R(0) &= \emptyset \\ R(1) &= \{\ (x, x) \mid x \in D\} \\ R(a) &= R_a \\ R(E + F) &= R(E) \cup R(F) \\ R(E \cdot F) &= R(E) \circ R(F) \\ R(E^*) &= \bigcup_{n \in \mathbb{N}} R(E)^n \end{aligned}$$

Kozen's Axiomatisation (1991)

Kozen's Axiomatisation consists in the axioms of idempotent semiring, plus the following axioms for the Kleene star operator.

$$A \leq B \stackrel{\text{def}}{=} A + B = B$$

$$1 + AA^* \leq A^* \tag{K1}$$

$$1 + A^*A \leq A^* \tag{K2}$$

$$AB + C \leq B \Rightarrow A^*C \leq B \tag{K3}$$

$$BA + C \leq B \Rightarrow CA^* \leq B \tag{K4}$$

Kozen axiomatisation gets rid of the guarded condition of Salomaa (SI3) in the implication rules, which make them Horn clauses.

Definition (Kozen 1991)

Kozen's axiomatisation is the definition of **Kleene Algebras**.

Another complete axiomatisation

Theorem (Kozen 1991)

Kozen's Axiomatisation is complete.
For all expressions A and B, if $L(A) = L(B)$ then the equality $A = B$ can be proved using Kozen's axiomatisation.

Proof: Elegant proof using the fact that matrices (automata) form also a Kleene algebra. Determinisation and Minimisation are operations provable with the axiomatisation.

Kozen's axiomatisation defines what is now called as
Kleene algebras

Pratt's axiomatisation

Pratt's Axiomatisation (1991) Action Algebra

(first version)

Consider two new operators called **residuations** $\leftarrow$ and $\rightarrow$, interpreted as the following on the language model:

$$A \rightarrow B = \{v \mid \forall u \in A, uv \in B\}$$

$$B \leftarrow A = \{v \mid \forall u \in A, vu \in B\}$$

Using these new operators, Pratt propose the following axiomatisation:

$$AB \leq C \;\Leftrightarrow\; B \leq A \rightarrow C \tag{P1}$$

$$AB \leq C \;\Leftrightarrow\; A \leq C \leftarrow B \tag{P2}$$

$$1 + A^*A^* + A \;\leq\; A^* \tag{P3}$$

$$1 + BB + A \leq B \;\Rightarrow\; A^* \leq B \tag{P4}$$

Pratt's Axiomatisation (1991) Action Algebra

(second version)

This presentation can be restated avoiding any implication rule. The axioms for the residuations are the following:

$$\begin{aligned} A \to B &\leq A \to (B + B') \\ B &\leq A \to AB \\ A(A \to B) &\leq B \end{aligned}$$

$$\begin{aligned} B \leftarrow A &\leq (B + B') \leftarrow A \\ B &\leq BA \leftarrow A \\ (B \leftarrow A)A &\leq B \end{aligned}$$

Axioms for Kleene Star * are the following:

$$\begin{aligned} 1 + A^*A^* + A &\leq A^* \\ A^* &\leq (A + B)^* \\ (A \to A)^* &\leq A \to A \qquad \text{(P5)} \end{aligned}$$

A conservative extension

Theorem (Pratt 1991)

Pratt's Action Algebras are a conservative extension of Kleene Algebras.

For all expressions A and B, if $A = B$ is provable in Kleene Algebra then it is provable in Action Algebra.

Implementation

Thank you !