Lecture 7

Pattern Matching

What happens when one types `rm *` in UNIX? (If you don't know, don't try it to find out!) What if the current directory contains the files

```
a.tex  bc.tex  a.dvi  bc.dvi
```

and one types `rm *.dvi`? What would happen if there were a file named `.dvi`?

What is going on here is *pattern matching*. The `*` in UNIX is a pattern that matches any string of symbols, including the null string.

Pattern matching is an important application of finite automata. The UNIX commands `grep`, `fgrep`, and `egrep` are basic pattern-matching utilities that use finite automata in their implementation.

Let Σ be a finite alphabet. A *pattern* is a string of symbols of a certain form representing a (possibly infinite) set of strings in Σ^*. The set of patterns is defined formally by induction below. They are either *atomic patterns* or *compound patterns* built up inductively from atomic patterns using certain *operators*. We'll denote patterns by Greek letters α, β, γ, ...

As we define patterns, we will tell which strings $x \in \Sigma^*$ *match* them. The set of strings in Σ^* matching a given pattern α will be denoted $L(\alpha)$. Thus

$$L(\alpha) \quad = \quad \{x \in \Sigma^* \mid x \text{ matches } \alpha\}.$$

In the following, forget the UNIX definition of *. We will use the symbol * for something else.

The *atomic patterns* are

- a for each $a \in \Sigma$, matched by the symbol a only; in symbols, $L(a) = \{a\}$;
- $\boldsymbol{\epsilon}$, matched only by ϵ, the null string; in symbols, $L(\boldsymbol{\epsilon}) = \{\epsilon\}$;
- $\boldsymbol{\emptyset}$, matched by nothing; in symbols, $L(\boldsymbol{\emptyset}) = \emptyset$, the empty set;
- $\#$, matched by any symbol in Σ; in symbols, $L(\#) = \Sigma$;
- @, matched by any string in Σ^*; in symbols, $L(@) = \Sigma^*$.

Compound patterns are formed inductively using binary operators $+$, $\cap$, and $\cdot$ (usually not written) and unary operators $^+$, *, and $\sim$. If α and β are patterns, then so are $\alpha + \beta$, $\alpha \cap \beta$, α^*, α^+, $\sim\alpha$, and $\alpha\beta$. The last of these is short for $\alpha \cdot \beta$.

We also define inductively which strings match each pattern. We have already said which strings match the atomic patterns. This is the basis of the inductive definition. Now suppose we have already defined the sets of strings $L(\alpha)$ and $L(\beta)$ matching α and β, respectively. Then we'll say that

- x matches $\alpha + \beta$ if x matches either α or β:

$$L(\alpha + \beta) = L(\alpha) \cup L(\beta);$$

- x matches $\alpha \cap \beta$ if x matches both α and β:

$$L(\alpha \cap \beta) = L(\alpha) \cap L(\beta);$$

- x matches $\alpha\beta$ if x can be broken down as $x = yz$ such that y matches α and z matches β:

$$\begin{aligned} L(\alpha\beta) &= L(\alpha)L(\beta) \\ &= \{yz \mid y \in L(\alpha) \text{ and } z \in L(\beta)\}; \end{aligned}$$

- x matches $\sim\alpha$ if x does not match α:

$$\begin{aligned} L(\sim\alpha) &= \sim L(\alpha) \\ &= \Sigma^* - L(\alpha); \end{aligned}$$

- x matches α^* if x can be expressed as a concatenation of zero or more strings, all of which match α:

$$\begin{aligned} L(\alpha^*) &= \{x_1 x_2 \cdots x_n \mid n \geq 0 \text{ and } x_i \in L(\alpha),\ 1 \leq i \leq n\} \\ &= L(\alpha)^0 \cup L(\alpha)^1 \cup L(\alpha)^2 \cup \cdots \\ &= L(\alpha)^*. \end{aligned}$$

The null string ϵ always matches α^*, since ϵ is a concatenation of zero strings, all of which (vacuously) match α.

- x matches α^+ if x can be expressed as a concatenation of one or more strings, all of which match α:

$$\begin{aligned} L(\alpha^+) &= \{x_1 x_2 \cdots x_n \mid n \geq 1 \text{ and } x_i \in L(\alpha),\ 1 \leq i \leq n\} \\ &= L(\alpha)^1 \cup L(\alpha)^2 \cup L(\alpha)^3 \cup \cdots \\ &= L(\alpha)^+. \end{aligned}$$

Note that patterns are just certain strings of symbols over the alphabet

$$\Sigma \cup \{\, \boldsymbol{\epsilon}, \boldsymbol{\emptyset}, \#, @, +, \cap, \sim, {}^*, {}^+, (,) \,\}.$$

Note also that the meanings of $\#$, $@$, and $\sim$ depend on Σ. For example, if $\Sigma = \{a, b, c\}$ then $L(\#) = \{a, b, c\}$, but if $\Sigma = \{a\}$ then $L(\#) = \{a\}$.

Example 7.1

- $\Sigma^* = L(@) = L(\#^*)$.
- Singleton sets: if $x \in \Sigma^*$, then x itself is a pattern and is matched only by the string x; i.e., $\{x\} = L(x)$.
- Finite sets: if $x_1, \ldots, x_m \in \Sigma^*$, then

$$\{x_1, x_2, \ldots, x_m\} \quad = \quad L(x_1 + x_2 + \cdots + x_m). \qquad \square$$

Note that we can write the last pattern $x_1 + x_2 + \cdots + x_m$ without parentheses, since the two patterns $(\alpha + \beta) + \gamma$ and $\alpha + (\beta + \gamma)$ are matched by the same set of strings; i.e.,

$$L((\alpha + \beta) + \gamma) \quad = \quad L(\alpha + (\beta + \gamma)).$$

Mathematically speaking, the operator $+$ is *associative*. The concatenation operator $\cdot$ is associative, too. Hence we can also unambiguously write $\alpha\beta\gamma$ without parentheses.

Example 7.2

- strings containing at least three occurrences of a:

$$@a@a@a@;$$

- strings containing an a followed later by a b; that is, strings of the form $xaybz$ for some x, y, z:

$$@a@b@;$$

- all single letters except a:

$$\# \cap \sim a;$$

- strings with no occurrence of the letter a:

  $$(\# \cap \sim a)^*;$$

- strings in which every occurrence of a is followed sometime later by an occurrence of b; in other words, strings in which there are either no occurrences of a, or there is an occurrence of b followed by no occurrence of a; for example, aab matches but bba doesn't:

  $$(\# \cap \sim a)^* + @b(\# \cap \sim a)^*.$$

  If the alphabet is $\{a, b\}$, then this takes a much simpler form:

  $$\boldsymbol{\epsilon} + @b.$$ □

Before we go too much further, there is a subtlety that needs to be mentioned. Note the slight difference in appearance between $\boldsymbol{\epsilon}$ and ϵ and between $\boldsymbol{\emptyset}$ and $\emptyset$. The objects $\boldsymbol{\epsilon}$ and $\boldsymbol{\emptyset}$ are *symbols* in the language of patterns, whereas ϵ and $\emptyset$ are *metasymbols* that we are using to name the null string and the empty set, respectively. These are different sorts of things: $\boldsymbol{\epsilon}$ and $\boldsymbol{\emptyset}$ are symbols, that is, strings of length one, whereas ϵ is a string of length zero and $\emptyset$ isn't even a string.

We'll maintain the distinction for a few lectures until we get used to the idea, but at some point in the near future we'll drop the boldface and use ϵ and $\emptyset$ exclusively. We'll always be able to infer from context whether we mean the symbols or the metasymbols. This is a little more convenient and conforms to standard usage, but bear in mind that they are still different things.

While we're on the subject of abuse of notation, we should also mention that very often you will see things like $x \in a^*b^*$ in texts and articles. Strictly speaking, one should write $x \in L(a^*b^*)$, since a^*b^* is a pattern, not a set of strings. But as long as you know what you really mean and can stand the guilt, it is okay to write $x \in a^*b^*$.

Lecture 8

Pattern Matching and Regular Expressions

Here are some interesting and important questions:

- How hard is it to determine whether a given string x matches a given pattern α? This is an important practical question. There are very efficient algorithms, as we will see.
- Is every set represented by some pattern? Answer: no. For example, the set

  $$\{a^n b^n \mid n \geq 0\}$$

  is not represented by any pattern. We'll prove this later.
- Patterns α and β are *equivalent* if $L(\alpha) = L(\beta)$. How do you tell whether α and β are equivalent? Sometimes it is obvious and sometimes not.
- Which operators are redundant? For example, we can get rid of $\boldsymbol{\epsilon}$ since it is equivalent to $\sim(\#@)$ and also to $\boldsymbol{\emptyset}^*$. We can get rid of @ since it is equivalent to $\#^*$. We can get rid of unary $^+$ since α^+ is equivalent to $\alpha\alpha^*$. We can get rid of #, since if $\Sigma = \{a_1, \ldots, a_n\}$ then # is equivalent to the pattern

  $$a_1 + a_2 + \cdots + a_n.$$

The operator $\cap$ is also redundant, by one of the De Morgan laws:

$$\alpha \cap \beta \quad \text{is equivalent to} \quad \sim(\sim\alpha + \sim\beta).$$

Redundancy is an important question. From a user's point of view, we would like to have a lot of operators since this lets us write more succinct patterns; but from a programmer's point of view, we would like to have as few as possible since there is less code to write. Also, from a theoretical point of view, fewer operators mean fewer cases we have to treat in giving formal semantics and proofs of correctness.

An amazing and difficult-to-prove fact is that the operator $\sim$ is redundant. Thus every pattern is equivalent to one using only atomic patterns $a \in \Sigma$, $\boldsymbol{\epsilon}$, $\boldsymbol{\emptyset}$, and operators $+$, $\cdot$, and *. Patterns using only these symbols are called *regular expressions*. Actually, as we have observed, even $\boldsymbol{\epsilon}$ is redundant, but we include it in the definition of regular expressions because it occurs so often.

Our goal for this lecture and the next will be to show that the family of subsets of Σ^* represented by patterns is exactly the family of regular sets. Thus as a way of describing subsets of Σ^*, finite automata, patterns, and regular expressions are equally expressive.

Some Notational Conveniences

Since the binary operators $+$ and $\cdot$ are associative, that is,

$$\begin{aligned} L(\alpha + (\beta + \gamma)) &= L((\alpha + \beta) + \gamma), \\ L(\alpha(\beta\gamma)) &= L((\alpha\beta)\gamma), \end{aligned}$$

we can write

$$\alpha + \beta + \gamma \qquad \text{and} \qquad \alpha\beta\gamma$$

without ambiguity. To resolve ambiguity in other situations, we assign precedence to operators. For example,

$$\alpha + \beta\gamma$$

could be interpreted as either

$$\alpha + (\beta\gamma) \qquad \text{or} \qquad (\alpha + \beta)\gamma,$$

which are not equivalent. We adopt the convention that the concatenation operator $\cdot$ has higher precedence than $+$, so that we would prefer the former interpretation. Similarly, we assign * higher precedence than $+$ or $\cdot$, so that

$$\alpha + \beta^*$$

is interpreted as

$$\alpha + (\beta^*)$$

and not as

$$(\alpha + \beta)^*.$$

All else failing, use parentheses.

Equivalence of Patterns, Regular Expressions, and Finite Automata

Patterns, regular expressions (patterns built from atomic patterns $a \in \Sigma$, $\boldsymbol{\epsilon}$, $\boldsymbol{\emptyset}$, and operators $+$, *, and $\cdot$ only), and finite automata are all equivalent in expressive power: they all represent the regular sets.

Theorem 8.1 *Let $A \subseteq \Sigma^*$. The following three statements are equivalent:*

(i) *A is regular; that is, $A = L(M)$ for some finite automaton M;*

(ii) *$A = L(\alpha)$ for some pattern α;*

(iii) *$A = L(\alpha)$ for some regular expression α.*

Proof. The implication (iii) $\Rightarrow$ (ii) is trivial, since every regular expression is a pattern. We prove (ii) $\Rightarrow$ (i) here and (i) $\Rightarrow$ (iii) in Lecture 9.

The heart of the proof (ii) $\Rightarrow$ (i) involves showing that certain basic sets (corresponding to atomic patterns) are regular, and the regular sets are closed under certain closure operations corresponding to the operators used to build patterns. Note that

- the singleton set $\{a\}$ is regular, $a \in \Sigma$,
- the singleton set $\{\epsilon\}$ is regular, and
- the empty set $\varnothing$ is regular,

since each of these sets is the set accepted by some automaton. Here are nondeterministic automata for these three sets, respectively:

Also, we have previously shown that the regular sets are closed under the set operations $\cup$, $\cap$, $\sim$, $\cdot$, *, and $^+$; that is, if A and B are regular sets, then so are $A \cup B$, $A \cap B$, $\sim A = \Sigma^* - A$, AB, A^*, and A^+.

These facts can be used to prove inductively that (ii) $\Rightarrow$ (i). Let α be a given pattern. We wish to show that $L(\alpha)$ is a regular set. We proceed by

induction on the structure of α. The pattern α is of one of the following forms:

(i) a, where $a \in \Sigma$;
(ii) $\boldsymbol{\epsilon}$;
(iii) $\boldsymbol{\emptyset}$;
(iv) $\#$;
(v) @;
(vi) $\beta + \gamma$;
(vii) $\beta \cap \gamma$;
(viii) $\beta\gamma$;
(ix) $\sim\beta$;
(x) β^*;
(xi) β^+.

There are five base cases (i) through (v) corresponding to the atomic patterns and six induction cases (vi) through (xi) corresponding to compound patterns. Each of these cases uses a closure property of the regular sets previously observed.

For (i), (ii), and (iii), we have $L(a) = \{a\}$ for $a \in \Sigma$, $L(\boldsymbol{\epsilon}) = \{\epsilon\}$, and $L(\boldsymbol{\emptyset}) = \varnothing$, and these are regular sets.

For (iv), (v), and (xi), we observed earlier that the operators $\#$, @, and $^+$ were redundant, so we may disregard these cases since they are already covered by the other cases.

For (vi), recall that $L(\beta+\gamma) = L(\beta) \cup L(\gamma)$ by definition of the $+$ operator. By the induction hypothesis, $L(\beta)$ and $L(\gamma)$ are regular. Since the regular sets are closed under union, $L(\beta + \gamma) = L(\beta) \cup L(\gamma)$ is also regular.

The arguments for the remaining cases (vii) through (x) are similar to the argument for (vi). Each of these cases uses a closure property of the regular sets that we have observed previously in Lectures 4 and 6. □

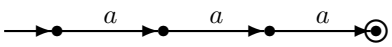
Example 8.2 Let's convert the regular expression

$$(aaa)^* + (aaaaa)^*$$

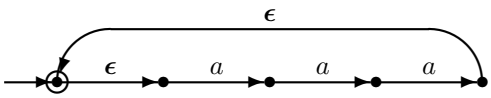
for the set

$$\{x \in \{a\}^* \mid |x| \text{ is divisible by either 3 or 5}\}$$

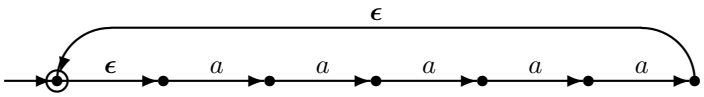
to an equivalent NFA. First we show how to construct an automaton for $(aaa)^*$. We take an automaton accepting only the string aaa, say



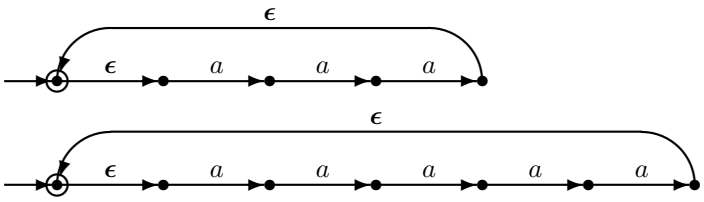
Applying the construction of Lecture 6, we add a new start state and ϵ-transitions from the new start state to all the old start states and from all the old accept states to the new start state. We let the new start state be the only accept state of the new automaton. This gives



The construction for $(aaaaa)^*$ is similar, giving



To get an NFA for $(aaa)^* + (aaaaa)^*$, we can simply take the disjoint union of these two automata:



□

Lecture 9

Regular Expressions and Finite Automata

Simplification of Expressions

For small regular expressions, one can often see how to construct an equivalent automaton directly without going through the mechanical procedure of the previous lecture. It is therefore useful to try to simplify the expression first.

For regular expressions α, β, if $L(\alpha) = L(\beta)$, we write $\alpha \equiv \beta$ and say that α and β are *equivalent*. The relation $\equiv$ on regular expressions is an equivalence relation; that is, it is

- reflexive: $\alpha \equiv \alpha$ for all α;
- symmetric: if $\alpha \equiv \beta$, then $\beta \equiv \alpha$; and
- transitive: if $\alpha \equiv \beta$ and $\beta \equiv \gamma$, then $\alpha \equiv \gamma$.

If $\alpha \equiv \beta$, one can substitute α for β (or vice versa) in any regular expression, and the resulting expression will be equivalent to the original.

Here are a few laws that can be used to simplify regular expressions.

$$\alpha + (\beta + \gamma) \equiv (\alpha + \beta) + \gamma \tag{9.1}$$
$$\alpha + \beta \equiv \beta + \alpha \tag{9.2}$$
$$\alpha + \boldsymbol{\emptyset} \equiv \alpha \tag{9.3}$$
$$\alpha + \alpha \equiv \alpha \tag{9.4}$$
$$\alpha(\beta\gamma) \equiv (\alpha\beta)\gamma \tag{9.5}$$
$$\boldsymbol{\epsilon}\alpha \equiv \alpha\boldsymbol{\epsilon} \equiv \alpha \tag{9.6}$$
$$\alpha(\beta + \gamma) \equiv \alpha\beta + \alpha\gamma \tag{9.7}$$
$$(\alpha + \beta)\gamma \equiv \alpha\gamma + \beta\gamma \tag{9.8}$$
$$\boldsymbol{\emptyset}\alpha \equiv \alpha\boldsymbol{\emptyset} \equiv \boldsymbol{\emptyset} \tag{9.9}$$
$$\boldsymbol{\epsilon} + \alpha\alpha^* \equiv \alpha^* \tag{9.10}$$
$$\boldsymbol{\epsilon} + \alpha^*\alpha \equiv \alpha^* \tag{9.11}$$
$$\beta + \alpha\gamma \leq \gamma \Rightarrow \alpha^*\beta \leq \gamma \tag{9.12}$$
$$\beta + \gamma\alpha \leq \gamma \Rightarrow \beta\alpha^* \leq \gamma \tag{9.13}$$

In (9.12) and (9.13), $\leq$ refers to the subset order:

$$\begin{aligned} \alpha \leq \beta &\stackrel{\text{def}}{\iff} L(\alpha) \subseteq L(\beta) \\ &\iff L(\alpha + \beta) = L(\beta) \\ &\iff \alpha + \beta \equiv \beta. \end{aligned}$$

Laws (9.12) and (9.13) are not equations but rules from which one can derive equations from other equations. Laws (9.1) through (9.13) can be justified by replacing each expression by its definition and reasoning set theoretically.

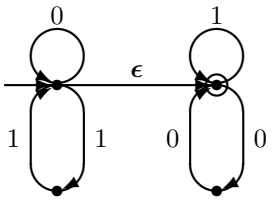
Here are some useful equations that follow from (9.1) through (9.13) that you can use to simplify expressions.

$$(\alpha\beta)^*\alpha \equiv \alpha(\beta\alpha)^* \tag{9.14}$$
$$(\alpha^*\beta)^*\alpha^* \equiv (\alpha + \beta)^* \tag{9.15}$$
$$\alpha^*(\beta\alpha^*)^* \equiv (\alpha + \beta)^* \tag{9.16}$$
$$(\boldsymbol{\epsilon} + \alpha)^* \equiv \alpha^* \tag{9.17}$$
$$\alpha\alpha^* \equiv \alpha^*\alpha \tag{9.18}$$

An interesting fact that is beyond the scope of this course is that all true equations between regular expressions can be proved purely algebraically from the axioms and rules (9.1) through (9.13) plus the laws of equational logic [64].

To illustrate, let's convert some regular expressions to finite automata.

Example 9.1 $(11+0)^*(00+1)^*$



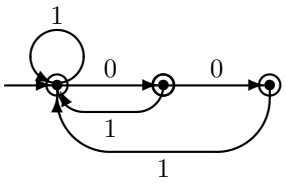
This expression is simple enough that the easiest thing to do is eyeball it. The mechanical method described in Lecture 8 would give more states and ϵ-transitions than shown here. The two states connected by an ϵ-transition cannot be collapsed into one state, since then 10 would be accepted, which does not match the regular expression. □

Example 9.2 $(1+01+001)^*(\epsilon+0+00)$

Using the algebraic laws above, we can rewrite the expression:

$$\begin{aligned}(1+01+001)^*(\epsilon+0+00) &\equiv ((\epsilon+0+00)1)^*(\epsilon+0+00)\\ &\equiv ((\epsilon+0)(\epsilon+0)1)^*(\epsilon+0)(\epsilon+0).\end{aligned}$$

It is now easier to see that the set represented is the set of all strings over $\{0,1\}$ with no substring of more than two adjacent 0's.



□

Just because all states of an NFA are accept states doesn't mean that all strings are accepted! Note that in Example 9.2, 000 is not accepted.

Converting Automata to Regular Expressions

To finish the proof of Theorem 8.1, it remains to show how to convert a given finite automaton M to an equivalent regular expression.

Given an NFA

$$M = (Q, \Sigma, \Delta, S, F),$$

a subset $X \subseteq Q$, and states $u, v \in Q$, we show how to construct a regular expression

$$\alpha_{uv}^X$$

representing the set of all strings x such that there is a path from u to v in M labeled x (i.e., such that $v \in \widehat{\Delta}(\{u\}, x)$) and all states along that path, with the possible exception of u and v, lie in X.

The expressions are constructed inductively on the size of X. For the basis $X = \varnothing$, let $a_1, \ldots, a_k$ be all the symbols in Σ such that $v \in \Delta(u, a_i)$. For $u \neq v$, take

$$\alpha_{uv}^{\varnothing} \stackrel{\text{def}}{=} \begin{cases} a_1 + \cdots + a_k & \text{if } k \geq 1, \\ \boldsymbol{\emptyset} & \text{if } k = 0; \end{cases}$$

and for $u = v$, take

$$\alpha_{uv}^{\varnothing} \stackrel{\text{def}}{=} \begin{cases} a_1 + \cdots + a_k + \boldsymbol{\epsilon} & \text{if } k \geq 1, \\ \boldsymbol{\epsilon} & \text{if } k = 0. \end{cases}$$

For nonempty X, we can choose any element $q \in X$ and take

$$\alpha_{uv}^{X} \stackrel{\text{def}}{=} \alpha_{uv}^{X-\{q\}} + \alpha_{uq}^{X-\{q\}} (\alpha_{qq}^{X-\{q\}})^* \alpha_{qv}^{X-\{q\}}. \tag{9.19}$$

To justify the definition (9.19), note that any path from u to v with all intermediate states in X either (i) never visits q, hence the expression

$$\alpha_{uv}^{X-\{q\}}$$

on the right-hand side of (9.19); or (ii) visits q for the first time, hence the expression

$$\alpha_{uq}^{X-\{q\}},$$

followed by a finite number (possibly zero) of loops from q back to itself without visiting q in between and staying in X, hence the expression

$$(\alpha_{qq}^{X-\{q\}})^*,$$

followed by a path from q to v after leaving q for the last time, hence the expression

$$\alpha_{qv}^{X-\{q\}}.$$

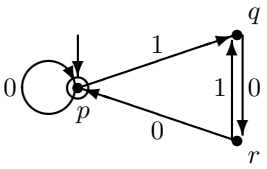
The sum of all expressions of the form

$$\alpha_{sf}^{Q},$$

where s is a start state and f is a final state, represents the set of strings accepted by M.

As a practical rule of thumb when doing homework exercises, when choosing the $q \in X$ to drop out in (9.19), it is best to try to choose one that disconnects the automaton as much as possible.

Example 9.3 Let's convert the automaton



to an equivalent regular expression. The set accepted by this automaton will be represented by the inductively defined regular expression

$$\alpha_{pp}^{\{p,q,r\}},$$

since p is the only start and the only accept state. Removing the state q (we can choose any state we like here), we can take

$$\alpha_{pp}^{\{p,q,r\}} \quad = \quad \alpha_{pp}^{\{p,r\}} + \alpha_{pq}^{\{p,r\}}(\alpha_{qq}^{\{p,r\}})^*\alpha_{qp}^{\{p,r\}}.$$

Looking at the automaton, the only paths going from p to p and staying in the states $\{p, r\}$ are paths going around the single loop labeled 0 from p to p some finite number of times; thus we can take

$$\alpha_{pp}^{\{p,r\}} \quad = \quad 0^*.$$

By similar informal reasoning, we can take

$$\begin{aligned}
\alpha_{pq}^{\{p,r\}} &= 0^*1, \\
\alpha_{qq}^{\{p,r\}} &= \boldsymbol{\epsilon} + 01 + 000^*1 \\
&\equiv \boldsymbol{\epsilon} + 0(\boldsymbol{\epsilon} + 00^*)1 \\
&\equiv \boldsymbol{\epsilon} + 00^*1, \\
\alpha_{qp}^{\{p,r\}} &= 000^*.
\end{aligned}$$

Thus we can take

$$\alpha_{pp}^{\{p,q,r\}} \quad = \quad 0^* + 0^*1(\boldsymbol{\epsilon} + 00^*1)^*000^*.$$

This is matched by the set of all strings accepted by the automaton. We can further simplify the expression using the algebraic laws (9.1) through (9.18):

$$\begin{aligned}
&0^* + 0^*1(\boldsymbol{\epsilon} + 00^*1)^*000^* \\
&\quad \equiv \quad 0^* + 0^*1(00^*1)^*000^* && \text{by (9.17)} \\
&\quad \equiv \quad \boldsymbol{\epsilon} + 00^* + 0^*10(0^*10)^*00^* && \text{by (9.10) and (9.14)} \\
&\quad \equiv \quad \boldsymbol{\epsilon} + (\boldsymbol{\epsilon} + 0^*10(0^*10)^*)00^* && \text{by (9.8)} \\
&\quad \equiv \quad \boldsymbol{\epsilon} + (0^*10)^*00^* && \text{by (9.10)} \\
&\quad \equiv \quad \boldsymbol{\epsilon} + (0^*10)^*0^*0 && \text{by (9.18)} \\
&\quad \equiv \quad \boldsymbol{\epsilon} + (0 + 10)^*0 && \text{by (9.15).}
\end{aligned}$$

□

Historical Notes

Kleene [**?**] proved that deterministic finite automata and regular expressions are equivalent. A shorter proof was given by McNaughton and Yamada [71].

The relationship between right- and left-linear grammars and regular sets (Homework 5, Exercise 1) was observed by Chomsky and Miller [17].