Section 3.13: The Pumping Lemma for Regular Languages

In this section we consider techniques for showing that particular languages are not regular.

(3.13) Introduction (Cont.)

Question: Is

$$L = \{\, 0^n 1^n \mid n \in \mathbb{N} \,\} = \{\%, 01, 0011, 000111, \ldots\}$$

a regular language?

Answer: No.

Intuitively, an automaton would have to have infinitely many states to accept this language. A finite automaton won't be able to keep track of how many 0's it has seen so far, and thus won't be able to insist that the correct number of 1's follow.

(3.13) Introduction (Cont.)

We could turn the preceding ideas into a direct proof that L is not regular.

Instead, we will first state a general result, called the Pumping Lemma for regular languages, for proving that languages are non-regular.

Next, we will show how the Pumping Lemma can be used to prove that L is non-regular.

Finally, we will prove the Pumping Lemma.

(3.13) Statement of Pumping Lemma

Lemma 3.13.1 (Pumping Lemma for Regular Languages)
For all regular languages L, there is a $n \in \mathbb{N}$ such that, for all $z \in \mathbf{Str}$, if $z \in L$ and $|z| \geq n$, then there are $u, v, w \in \mathbf{Str}$ such that $z = uvw$ and

(1) $|uv| \leq n$;

(2) $v \neq \%$; and

(3) $uv^iw \in L$, for all $i \in \mathbb{N}$.

(3.13) Interacting with the Pumping Lemma

When we use the Pumping Lemma, we can imagine that we are interacting with it.

We can give the Pumping Lemma a regular language L, and the lemma will give us back a natural number n such that the property of the lemma holds. We have no control over the value of n. We can then give the lemma a string z that is in L and has at least n symbols. The lemma will then break z up into parts u, v and w in such way that (1)–(3) hold. We have no control over how z is broken up into these parts. (1) says that uv has no more than n symbols. (2) says that v is nonempty. And (3) says that, if we "pump" (duplicate) v as many times as we like, the resulting string will still be in L.

(3.13) Application of Pumping Lemma

Before proving the Pumping Lemma, let's see how it can be used to prove that $L = \{\, 0^n 1^n \mid n \in \mathbb{N} \,\}$ is non-regular.

Suppose, toward a contradiction, that L is regular. Thus there is an $n \in \mathbb{N}$ with the property of the Pumping Lemma. Suppose $z = 0^n 1^n$. Since $z \in L$ and $|z| = 2n \geq n$, it follows that there are $u, v, w \in \mathbf{Str}$ such that $z = uvw$ and properties (1)–(3) of the lemma hold. Since $0^n 1^n = z = uvw$, (1) tells us that there are $i, j, k \in \mathbb{N}$ such that

$$u = 0^i, \quad v = 0^j, \quad w = 0^k 1^n, \quad i + j + k = n.$$

By (2), we have that $j \geq 1$, and thus that $i + k = n - j < n$. By (3), we have that

$$0^{i+k} 1^n = 0^i 0^k 1^n = uw = u\%w = uv^0 w \in L.$$

Thus $i + k = n$—contradiction. Thus L is not regular.

(3.13) Proof of Pumping Lemma

Now, let's prove the Pumping Lemma. Suppose L is a regular language. Thus there is a NFA M such that $L(M) = L$. Let $n = |Q_M|$. Suppose $z \in \mathbf{Str}$, $z \in L$ and $|z| \geq n$. Let $m = |z|$. Thus $1 \leq n \leq |z| = m$. Since $z \in L = L(M)$, there is a valid labeled path for M

$$q_1 \overset{a_1}{\Rightarrow} q_2 \overset{a_2}{\Rightarrow} \cdots q_m \overset{a_m}{\Rightarrow} q_{m+1},$$

that is labeled by z and where $q_1 = s_M$, $q_{m+1} \in A_M$ and $a_i \in \mathbf{Sym}$ for all $1 \leq i \leq m$. Since $|Q_M| = n$, not all of the states $q_1, \ldots, q_{n+1}$ are distinct. Thus, there are $1 \leq i < j \leq n+1$ such that $q_i = q_j$.

(3.13) Proof of Pumping Lemma (Cont.)

Hence, our path looks like:

$$q_1 \overset{a_1}{\Rightarrow} \cdots q_{i-1} \overset{a_{i-1}}{\Rightarrow} q_i \overset{a_i}{\Rightarrow} \cdots q_{j-1} \overset{a_{j-1}}{\Rightarrow} q_j \overset{a_j}{\Rightarrow} \cdots q_m \overset{a_m}{\Rightarrow} q_{m+1}.$$

Let

$$u = a_1 \cdots a_{i-1}, \quad v = a_i \cdots a_{j-1}, \quad w = a_j \cdots a_m.$$

Then $z = uvw$. Since $|uv| = j - 1$ and $j \leq n + 1$, we have that $|uv| \leq n$. Since $i < j$, we have that $i \leq j - 1$, and thus that $v \neq \%$.

(3.13) Proof of Pumping Lemma (Cont.)

Finally, since

$$q_i \in \Delta(\{q_1\}, u), \quad q_j \in \Delta(\{q_i\}, v), \quad q_{m+1} \in \Delta(\{q_j\}, w)$$

and $q_i = q_j$, we have that

$$q_j \in \Delta(\{q_1\}, u), \quad q_j \in \Delta(\{q_j\}, v), \quad q_{m+1} \in \Delta(\{q_j\}, w).$$

Thus, we have that $q_{m+1} \in \Delta(\{q_1\}, uv^i w)$ for all $i \in \mathbb{N}$. But $q_1 = s_M$ and $q_{m+1} \in A_M$, and thus $uv^i w \in L(M) = L$ for all $i \in \mathbb{N}$.

(3.13) Another Approach to Showing Non-regularity

Suppose $L' = \{\, w \in \{0, 1\}^* \mid w$ has an equal number of 0's and 1's $\}$.

We could show that L' is non-regular using the Pumping Lemma.

But we can also prove this result by using some of the closure properties of Section 3.11 plus the fact that $L = \{\, 0^n 1^n \mid n \in \mathbb{N} \,\}$ is non-regular.

Suppose, toward a contradiction, that L' is regular. It is easy to see that $\{0\}$ and $\{1\}$ are regular (e.g., they are generated by the regular expressions 0 and 1). Thus, by Theorem 3.11.17, we have that $\{0\}^*\{1\}^*$ is regular. Hence, by Theorem 3.11.17 again, it follows that $L = L' \cap \{0\}^*\{1\}^*$ is regular—contradiction. Thus L' is non-regular.

(3.13) Forlan Implementation of Pumping Lemma

The textbook also describes the implementation in Forlan of the idea behind the Pumping Lemma.