Non-Regular Languages

David Robinson

Pigeonhole Principle

If n items are put into m containers with $n > m$, at least one container contains more than one item.

Pumping Lemma

If A is a regular language, then there is a number p, the pumping length, so that if s is a string in A with length of at least p, then $s = xyz$, so that

1. xy^iz is a string in A for all $i \geq 0$, where y^i is y concatenated to itself i times
2. $|y| > 0$
3. $|xy| \leq p$

Proof

Let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA recognizing language A, and let $p = |Q|$.

- Consider $s \in A$ so that $|s| = n$, with $n \geq p$.
- Show that $s = xyz$ so that xy^iz is a string in A for all $i \geq 0$, with $|y| > 0$ and $|xy| \leq p$.

Let $s = s_1 s_2 \cdots s_n$ be a string accepted by M, with $n \geq p$.

Let $r_1 r_2 \cdots r_{n+1}$ be the sequence of states that M enters while computing S. Observe that:

- The state sequence has length $n + 1$, which is at least $p + 1$
- Within the first $p + 1$ states in the sequence, two different points in the sequence have to be the same state, by the pigeonhole principle
- Call the first one r_j and the second one r_k

Now let $x = s_1 \cdots s_{j-1}$, $y = s_j \cdots s_{k-1}$, and $z = s_k \cdots s_n$.

Observe that:

- x takes M from r_1 to r_j
- y takes M from r_j to r_k
- z takes M from r_k to r_{n+1}
- Therefore, M must accept xy^iz for all $i \geq 0$

Observe that:

- Since $j \neq k, |y| > 0$

Finally, observe that:

- Since $k \leq p + 1, |xy| \leq p$

We have shown that all three conditions of the pumping lemma hold.