Properties of Regular Languages

Reading: Chapter 4

Pumping Lemma for Regular Languages

- **Lemma:** (the pumping lemma)

  Let M be a DFA with |Q| = n states, and let x be a string in L(M), where |x| >= n. Then x = uvw, where u,v, and w are all in Σ* and:

  - 1<= |uv| <= n
  - |v| >= 1
  - uv^iw is in L(M), for all i >= 0

- Note this statement of the pumping lemma is slightly different than the books'.

Pumping Lemma for Regular Languages

- **Proof:**

Let M be a DFA where |Q| = n, and let x be a string in L(M) where |x|>=n. Furthermore, let $x = a_1a_2 \dots a_m$ where $\delta(q_0, a_1a_2 \dots a_p) = q_{jp}$.

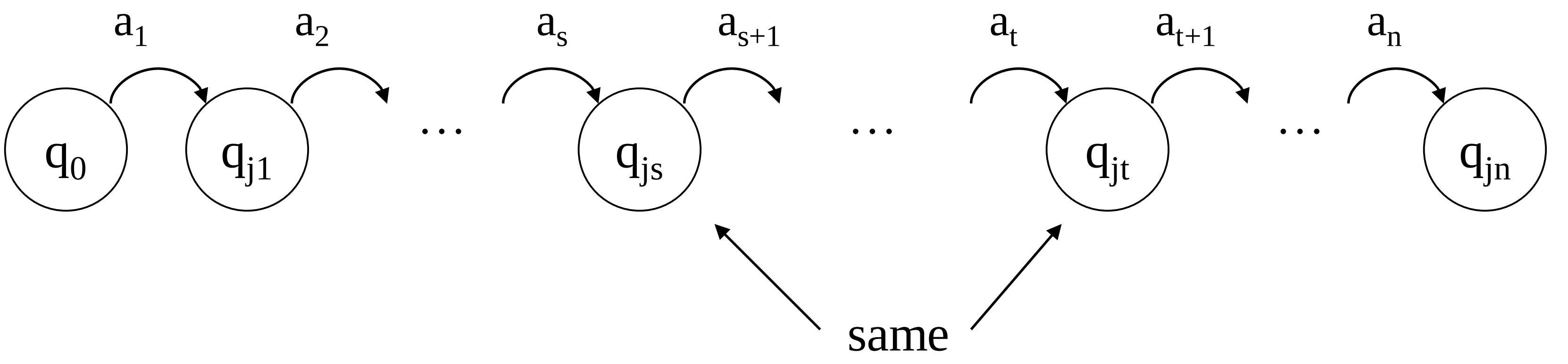
$$\begin{array}{ccccccccc} & a_1 & & a_2 & & a_3 \dots & & a_m & \\ q_{j0} & & q_{j1} & & q_{j2} & & q_{j3} \dots & & q_{jm} \end{array}$$

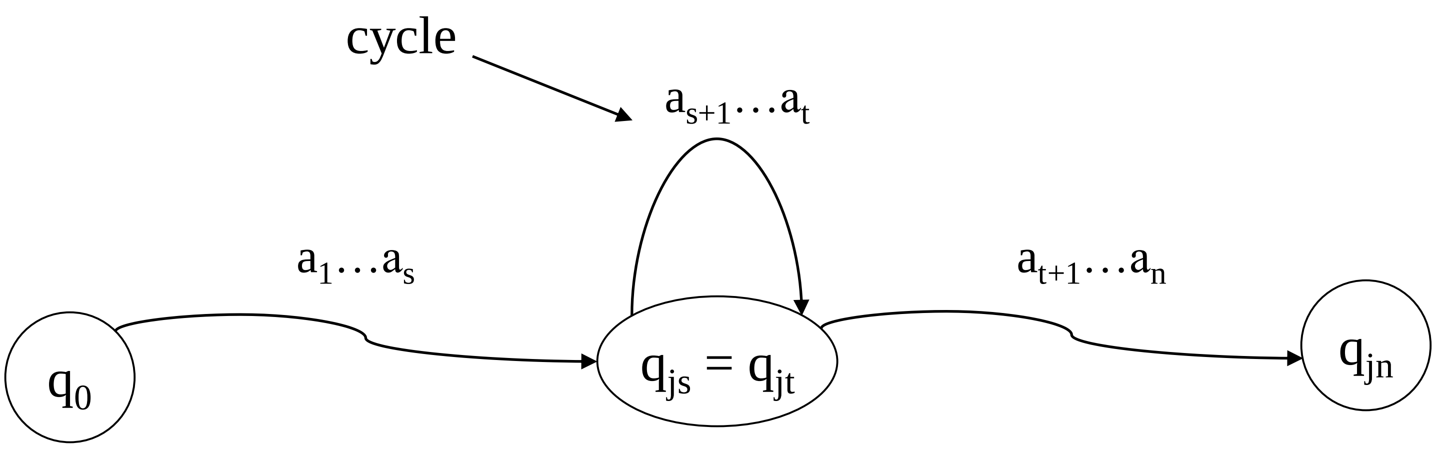
m>=n and q_{j0} is q_0

Consider the first n symbols, and first n+1 states on the above path:

$$\begin{array}{ccccccccc} & a_1 & & a_2 & & a_3 \dots & & a_n & \\ q_{j0} & & q_{j1} & & q_{j2} & & q_{j3} \dots & & q_{jn} \end{array}$$

Since |Q| = n, it follows from the pigeon-hole principle that $j_s = j_t$ for some 0<=s<t<=n, i.e., some state appears on this path twice (perhaps many states appear more than once, but at least one does).

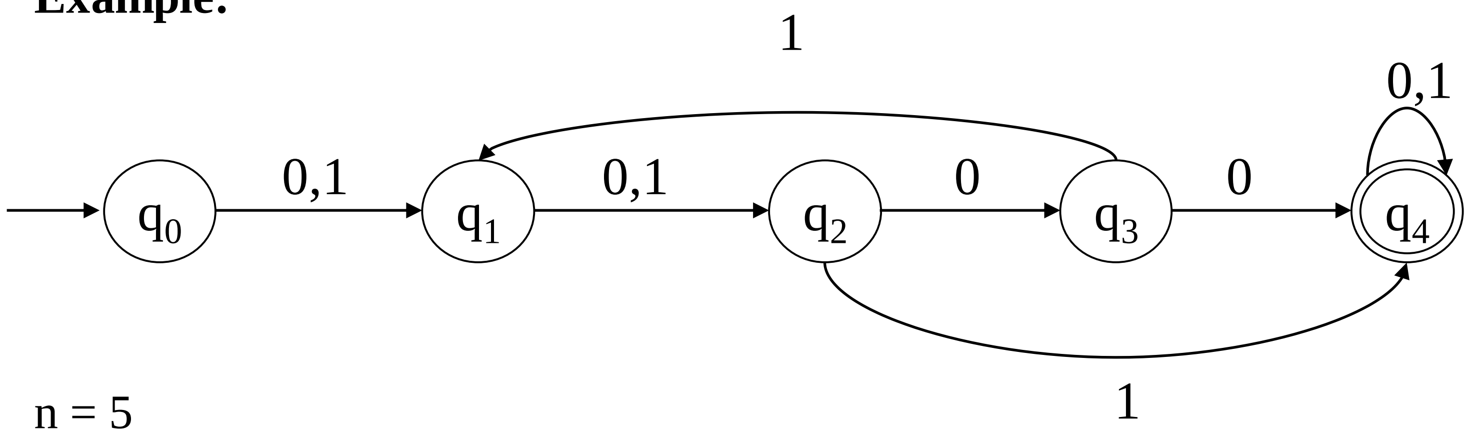

a_1
a_2
a_s
a_{s+1}
a_t
a_{t+1}
a_n
q_0
q_{j1}
…
q_{js}
…
q_{jt}
…
q_{jn}
same


cycle
$a_{s+1} \ldots a_t$
$a_1 \ldots a_s$
$a_{t+1} \ldots a_n$
q_0
$q_{js} = q_{jt}$
q_{jn}

- Let:
  - $u = a_1 \ldots a_s$
  - $v = a_{s+1} \ldots a_t$
- Since $0 <= s < t <= n$ and $uv = a_1 \ldots a_t$ it follows that:
  - $1 <= |v|$ and therefore $1 <= |uv|$
  - $|uv| <= n$ and therefore $1 <= |uv| <= n$
- In addition, let:
  - $w = a_{t+1} \ldots a_m$
- It follows that $uv^iw = a_1 \ldots a_s(a_{s+1} \ldots a_t)^i a_{t+1} \ldots a_m$ is in L(M), for all $i >= 0$.•

In other words, when processing the accepted string x, the cycle was traversed once, but could have been traversed as many times as desired, and the resulting string would still be accepted.

- **Example:**



n = 5

x = 0001000 is in L(M)

0 0 0 1 0 0 0

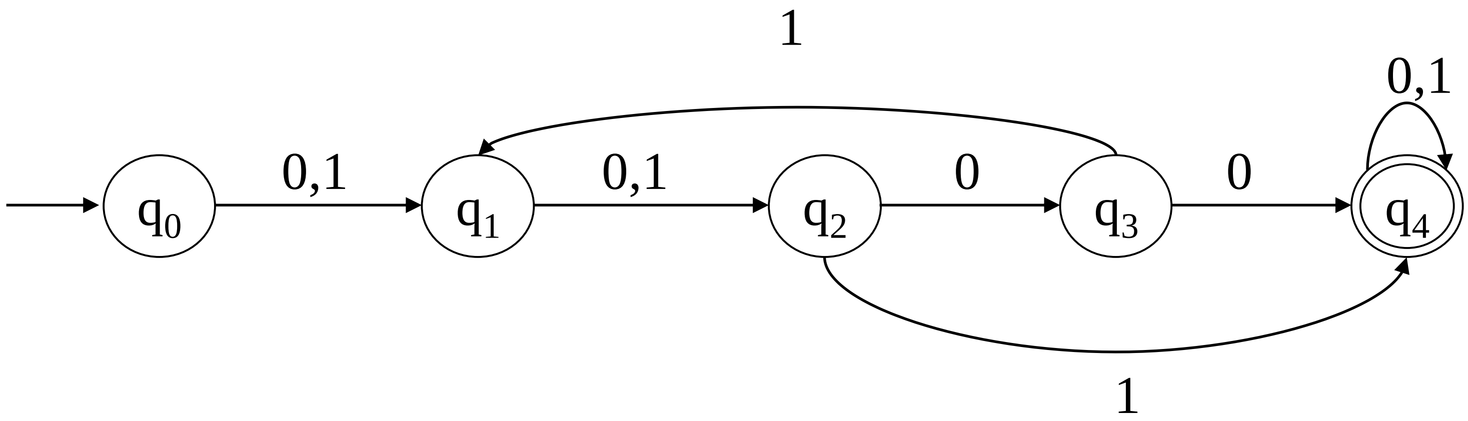
q_0 $\mathbf{q_1}$ q_2 q_3 $\mathbf{q_1}$ q_2 q_3 q_4

first n symbols

u = 0

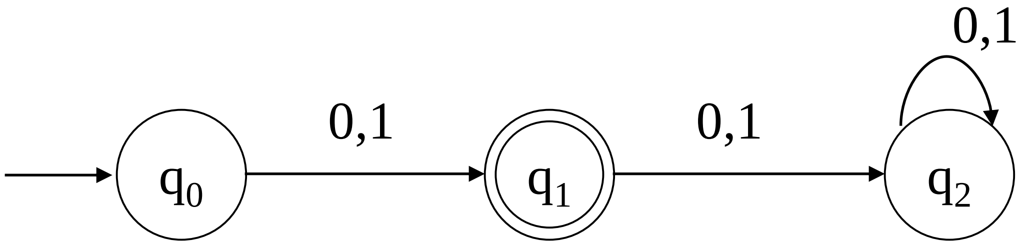
v = 001

w = 000 uv^iw is in L(M), i.e., $0(001)^i000$ is in L(M), for all i>= 0

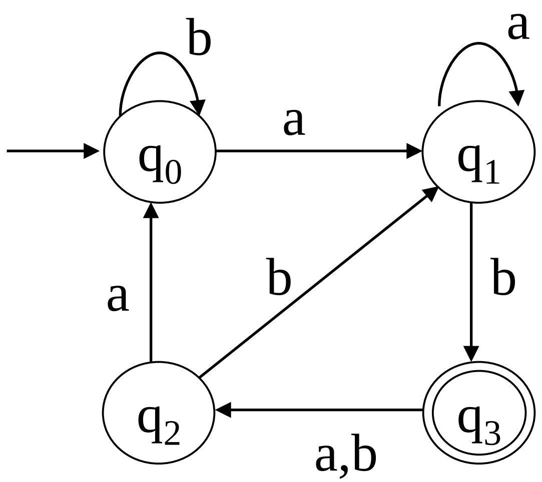


- Note this does not mean that every string accepted by the DFA has this form:
  - 001 is in L(M) but is not of the form $0(001)^i000$
- Similarly, this doesn't even mean that every <u>long</u> string accepted by the DFA has this form:
  - 0011111 is in L(M), is very long, but is not of the form $0(001)^i000$
- Note, however, in this latter case 0011111 could be similarly decomposed.

- **Note:** It may be the case that no x in L(M) has |x| >= n.



- **Example:**



n = 4

x = bbbab is in L(M)

|x| = 5

b b b a b

q0 **q0** q0 q0 q1 q3

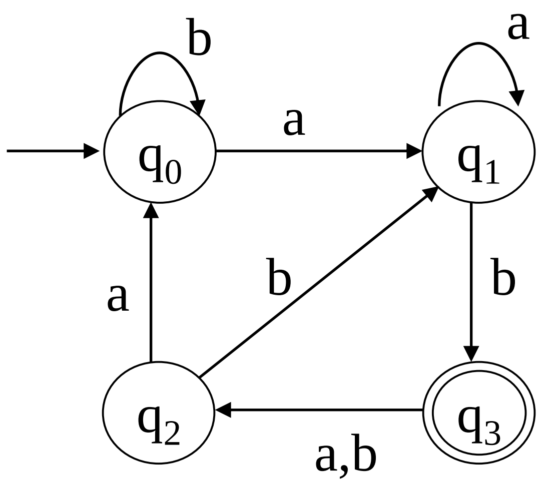
u = ε

v = b

w = bbab

$(b)^i$bbab is in L(M), for all i>=0

- **Example:**



n = 4

x = bbbab is in L(M)

|x| = 5

b b b a b

q0 **q0** **q0** q0 q1 q3

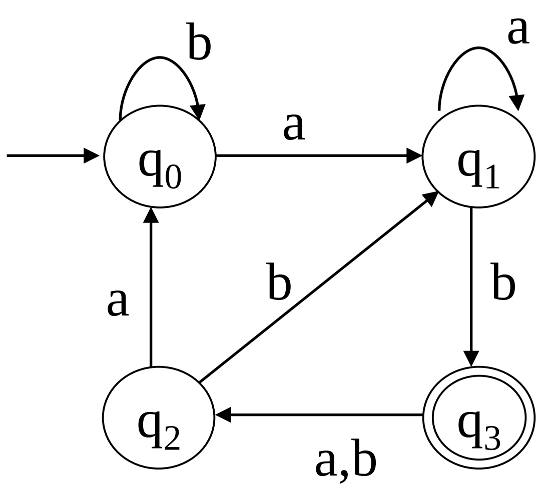
u = b

v = b

w = bab

$b(b)^i bab$ is in L(M), for all i>=0

- **Example:**



n = 4

x = bbbab is in L(M)

|x| = 5

b b b a b

q0 q0 **q0** **q0** q1 q3

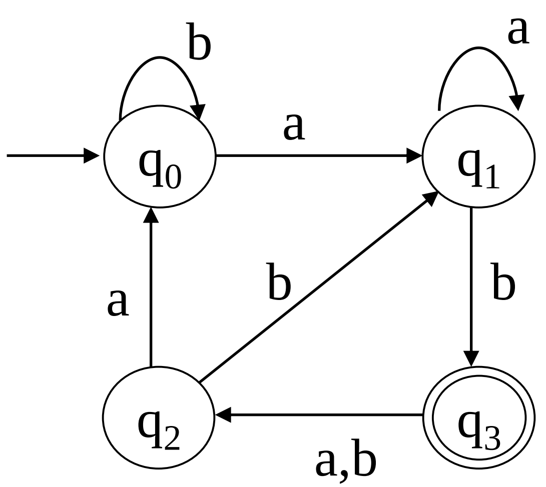
u = bb

v = b

w = ab

bb(b)^{i}ab is in L(M), for all i>=0

- **Example:**



n = 4

x = bbbab is in L(M)

|x| = 5

b b b a b

q0 **q0** q0 **q0** q1 q3

u = b

v = bb

w = ab

b(bb)^{i}ab is in L(M), for all i>=0

NonRegularity Example

- **Theorem:** The language:

  $$L = \{0^k 1^k \mid k >= 0\} \qquad (1)$$

  is not regular.

- **Proof:** (by contradiction) Suppose that L is regular. Then there exists a DFA M such that:

  $$L = L(M) \qquad (2)$$

  We will show that M accepts some strings not in L, contradicting (2).

  Suppose that M has n states, and consider a string $x=0^m1^m$, where m>>n.

  By (1), x is in L.

  By (2), x is also in L(M).

Since x is very long, i.e., $|x| = 2*m >> n$, it follows from the pumping lemma that:

- $x = uvw$
- $1 <= |uv| <= n$
- $1 <= |v|$, and
- uv^iw is in L(M), for all $i >= 0$

Since $1<=|uv|<=n$ and $n<<m$, it follows that $1<=|uv|<=m$.

Also, since $x = 0^m1^m$ it follows that uv is a substring of 0^m.

In other words $v=0^j$, for some $j>=1$.

Since uv^iw is in L(M), for all $i>=0$, it follows that $0^{m+cj}1^m$ is in L(M), for all $c>=1$.

But by (2), $0^{m+cj}1^m$ is in L, for any $c>=1$, a contradiction.•

NonRegularity Example

- **Theorem:** The language:

  $$L = \{0^k 1^k 2^k \mid k>=0\} \qquad (1)$$

  is not regular.

- **Proof:** (by contradiction) Suppose that L is regular. Then there exists a DFA M such that:

  $$L = L(M) \qquad (2)$$

  We will show that M accepts some strings not in L, contradicting (2).

  Suppose that M has n states, and consider a string $x=0^m1^m2^m$, where m>>n.

  By (1), x is in L.

  By (2), x is also in L(M).

Since is very long, i.e., |x| = 3*m >> n, it follows from the pumping lemma that:

- x = uvw
- 1 <= |uv| <= n
- 1 <= |v|, and
- uv^iw is in L(M), for all i >= 0

Since 1<=|uv|<=n and n<<m, it follows that 1<=|uv|<=m.

Also, since x = $0^m1^m2^m$ it follows that uv is a substring of 0^m.

In other words v=0^j, for some j>=1.

Since uv^iw is in L(M), for all i>=0, it follows that $0^{m+cj}1^m2^m$ is in L(M), for all c>=1.

But by (2), $0^{m+cj}1^m2^m$ is in L, for any c>=1, a contradiction.•

NonRegularity Example

- **Theorem:** The language:

  $$L = \{0^m 1^n 2^{m+n} \mid m,n >= 0\} \quad (1)$$

  is not regular.

- **Proof:** (by contradiction) Suppose that L is regular. Then there exists a DFA M such that:

  $$L = L(M) \quad (2)$$

  We will show that M accepts some strings not in L, contradicting (2).

  Suppose that M has n states, and consider a string $x=0^m 1^n 2^{m+n}$, where m>>n.

  By (1), x is in L.

  By (2), x is also in L(M).

Since |x| = 2(m+n) >> n, it follows from the pumping lemma that:

- x = uvw
- 1 <= |uv| <= n
- 1 <= |v|, and
- uv^iw is in L(M), for all i >= 0

Since 1<=|uv|<=n and n<<m, it follows that 1<=|uv|<=m.

Also, since x = $0^m1^n2^{m+n}$ it follows that uv is a substring of 0^m.

In other words v=0^j, for some j>=1.

Since uv^iw is in L(M), for all i>=0, it follows that $0^{m+cj}1^m2^{m+n}$ is in L(M), for all c>=1. In other words v can be "pumped" as many times as we like, and we still get a string in L(M).

But by (2), $0^{m+cj}1^n2^{m+n}$ is in L, for any c>=1, a contradiction.•

- Note that the above proof is almost identical to the previous proof.