COMPSCI 501: Formal Language Theory

Lecture 5: Nonregular Languages. Pumping Lemma

Marius Minea
marius@cs.umass.edu

University of Massachusetts Amherst

1 February, 2019

Regular Languages: Intuitions

- Accepted by finite automata
  bounded amount of memory
  clearly not possible for all languages
- Automata may have cycles
  strings in regular languages may have repeated portions
- Can we find a property
  of nonregular languages?
  of regular languages? the latter

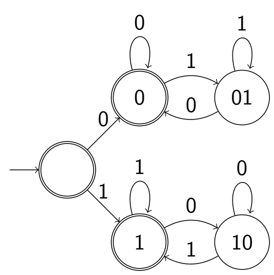
Careful with Intuition

Consider two languages that appear to need unbounded counters

$C = \{w \mid w$ has equal number of 0s and 1s $\}$

$D = \{w \mid w$ has equal number of 01 and 10 substrings $\}$

Are both of them nonregular?
D is not: count *parity* of alternations



Pumping Lemma

Pumping Lemma If A is a regular language, then there is a number p (*pumping length*) so that any string in A of length at least p can be divided into three pieces, $s = xyz$, with the conditions

1. $xy^iz \in A$ for any $i \geq 0$
2. $|y| > 0$
3. $|xy| \leq p$

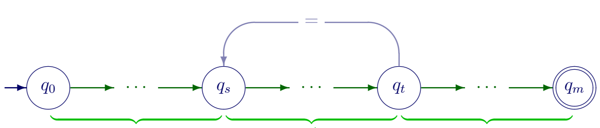
(1) tells us string y can be repeated (*pumped*) any number of times
(2) y is nonempty (else trivially true)
(3) says y found "early enough" (up to length p)

One-way Implication:

- **if** a language is regular, the Pumping Lemma applies
- **if** Pumping Lemma applies, language may or may not be regular

Proof of Pumping Lemma

Consider a DFA $M = (Q, \Sigma, \delta, q_0, F)$ recognizing A.
The pumping length p will be the number of states of M.
Consider a string $w = a_1a_2 \dots a_m$ of length $m \geq p$, and the sequence $q_0q_1 \dots q_m$ of $m + 1$ states traversed when processing w. q_m accepts, since $w \in A$. Since $m + 1 > p$, at least one state repeats. Let q_s be the first such state and q_t its first repeat occurrence.



Take $x = a_1 \dots a_s$, $y = a_{s+1} \dots a_t$ and $z = a_t \dots a_m$.
Since $t > s$, string y is nonempty (2); $|xy| = t \leq p$ (3).
y takes M from q_s back to $q_s(= q_t)$. String xy^iz takes M from q_0 to q_s, then i times back to q_s, then to q_m (accept). (1), q.e.d.

Figure: Jochen Burghardt / Wikipedia

Using the Pumping Lemma

Prove that a language is not regular, by contradiction.

- Assume language is regular, with some pumping length p.
- Construct a convenient accepted string of length $\geq p$
- Identify pumped substring y and its properties
- Show that one of the strings obtained by pumping y is not accepted (contradiction!)
  often by pumping up (repeating): $xyyz$, sometimes more y's
  sometimes by pumping down (deleting y): $xy^0z = xz$

Example: 0^n1^n

We prove $B = 0^n1^n$ is not regular, by contradiction.
What string should we choose?

Assume pumping length p, choose $s = 0^p1^p$.

Pumping lemma assures $|xy| \leq p$, thus y is formed just of 0s.

Then $xy^2z = xyyz$ has |y| more zeroes than ones and should not be accepted, but it is, contradiction!

Condition (3) simplifies the reasoning; otherwise, need to consider:
– y formed of 1s (same reasoning)
– y formed of 0s and 1s $\Rightarrow$ yy has 01...01 (bad)

Example: Equal count of 0 and 1

$C = \{w \mid w$ has equal number of 0s and 1s $\}$ is not regular.

Some strings in C *can* be pumped, e.g., $(01)^k$.

Again, assume C regular, choose $s = 0^p1^p$ (p = pumping length).

Could x and z be empty and $y = s = 0^p1^p$?

No, because (3) tells us $|xy| \leq p$.
Thus, y is all zeroes, by pumping we again get more zeroes.

Can also prove indirectly (language intersection).

If C (equal ones and zeroes) were regular, since 0^*1^* is regular, their intersection 0^n1^n would be regular (but we saw it isn't).

Example: Duplicate word

$F = \{ww \mid w \in \{0,1\}^*\}$ is not regular.

Similar idea: pumped string should ensure result does not have two equal halves.

Choose $w = 0^p10^p1$.

Then, since $|xy| \leq p$, y is formed just of zeroes.
$xyyz$ has more zeroes before first 1 than before second.

Example: Unary languages

$D = \{1^{n^2} \mid n \geq 0\}$ is not regular.

Intuition: length of strings grows further and further apart.

$xyyz$ is at most p longer than xyz.
$\Rightarrow$ choose initial $w = xyz$ so length of the pumped string falls in "gap" between two consecutive strings in language.

Choose $w = 1^{p^2}$. Then $p^2 < |xyyz| \leq p^2 + p < (p+1)^2$ (since $p \geq 1$), thus $xyyz \notin D$, contradiction.

Example: More zeroes followed by fewer ones

$E = \{0^i1^j \mid i > j\}$ is not regular.

Choosing w so that pumped string has just zeroes does not help when pumping up (number of zeroes increases).

But can pump down! $xy^0z = xz$ should be in the language!

Choose $w = 0^{p+1}1^p$.
Then since $|xy| \leq p$, y has just zeroes, and $|y| \geq 1$.
Thus, $xz = 0^k1^p$ with $k \leq p$ (at least one 0 less than w), and is not accepted!

Example: word followed by reverse

$G = \{ww^R \mid w \in \Sigma^*, |\Sigma| > 1\}$ is not regular.

Choose $w = 0^p110^p$.

Again, y will be all zeroes, and thus xz (or $xyyz$) are not symmetric.