There are languages that are not regular

There is an uncountably infinite number of possible languages over a given alphabet. (Why?) However, there are only countably many regular languages over the same alphabet. (Why?) Therefore, there must be many languages that are not regular languages.

However, this reasoning does not help us identify particular languages that are not regular. For that, we can use the pumping lemma for regular languages.

Pumping Lemma

Let L be a regular language. There exists an integer n such that for any string \( w \in L \) with \( |w| \geq n \), \( w \) may be written as \( w = xyz \), for some \( x, y, \) and \( z \) satisfying the following:

- \( |xy| \leq n \),
- \( |y| \geq 1 \),
- and \( xy^iz \in L \) for every \( i \geq 1 \)

Proof idea

If a DFA has \( n \) states, then any path of length \( n \) must visit \( n+1 \) states, and contains a cycle. (This is an application of the “pigeonhole principle.”)

This part of the string can be “pumped” to produce other strings in the language.

Proof idea again

If an infinite language is regular, it is accepted by a DFA. The DFA has some finite number of states, say, \( n \). Because the language is infinite, some strings must have length \( > n \).

For a string of length \( > n \) accepted by the DFA, a “walk” through the DFA must contain a cycle. Repeating the cycle an arbitrary number of times must yield another string accepted by the DFA.

How to use the pumping lemma

The Pumping Lemma describes a property that is possessed by every regular language. So if we show that a language does not possess this property, we know that it is not regular.

The strategy is proof by contradiction. We assume a language has the property described by the pumping lemma, and then we show that this leads to a contradiction.
Theorem: The language $L = \{ a^k b^k \mid k \geq 0 \}$ is not regular.

The proof is by contradiction. If $L$ is regular, it must be accepted by some DFA. Let $n$ be the number of states of the DFA and consider some $w \in L$ such that $|w| \geq n$. By the pumping lemma, we can split $w$ into three pieces, $w = xyz$, such that for any $k \geq 0$, the string $xy^kz$ is in $L$. So let $w = a^n b^n$. Because $|xy| \leq n$, $y$ must consist of all a’s. But then $xy^2z$ will contain more a’s than b’s, which is a contradiction.

Example

Use the pumping lemma to show that the language of palindromes $L = \{ w \mid w = w^R, w \in \{ a,b \}^* \}$ is not regular.

Exercise

Use the pumping lemma (plus some closure properties of regular languages) to show that the language $L = \{ w \in \{ a,b \}^* \mid w$ contains equal number of a’s and b’s $\}$ is not regular.

Exercise

Use the pumping lemma to show that the language $L = \{ ww \mid w \in \{ a,b \}^* \}$ is not regular.