Computer Science 351

Nonregular Languages, Part Two

Instructor: Wayne Eberly

Department of Computer Science
University of Calgary

Lecture #9

Goal for Today

Goals for Today:

- Describe how the ***closure properties for regular languages*** can *also* be used to prove that a language $L \subseteq \Sigma^\star$ is ***not*** regular.

Closure Properties

If $L_1, L_2 \subseteq \Sigma^\star$ for some alphabet Σ, and L_1 and L_2 are both regular languages, then the following languages are regular languages as well.

(a) $L_1 \cup L_2$

(b) $L_1 \circ L_2$

(c) $L_1^\star$

See the notes for Lecture #7 for more information about these claims. As noted there, these are examples of ***closure properties***.

Closure Properties

Recall that if $L \subseteq \Sigma^\star$ then the ***complement*** of L is the language

$$L^C = \{\omega \in \Sigma^\star \mid \omega \notin L\}.$$

Theorem #1. Suppose that $L \subseteq \Sigma^\star$ for an alphabet Σ. If L is regular, then L^C is a regular language as well.

Proof: Suppose that $L \subseteq \Sigma^\star$, for an an alphabet Σ, and that L is regular.

- It follows, by the definition of a "regular language", that there exists a ***deterministic finite automaton***

  $$M = (Q, \Sigma, \delta, q_0, F)$$

  such that $L(M) = L$.

Closure Properties

- Now consider a deterministic finite automaton

$$\widehat{M} = (Q, \Sigma, \delta, q_0, \widehat{F})$$

with the same set of states, alphabet, and start state, and with the set of accepting states

$$\widehat{F} = \{q \in Q \mid q \notin F\}.$$

- Since M is a deterministic finite automaton, $\widehat{M}$ is also a deterministic finite automaton with alphabet Σ — and, for every string $\omega \in \Sigma^{\star}$,

$$\omega \in L(M) \quad \text{if and only if} \quad \omega \notin L(\widehat{M}).$$

Closure Properties

- Thus $L(\widehat{M}) = L^C$.
- Since L^C is the language of a deterministic finite automaton, L^C is a regular language.
- Since L was an arbitrarily chosen regular language, it follows that if L is a regular language then L^C is also a regular language, for *every* language L, as claimed. $\square$

Closure Properties

Since $P \Rightarrow Q$ implies that $\neg Q \Rightarrow \neg P$, the above closure properties imply the following.

(a) For all languages $L_1, L_2 \subseteq \Sigma^\star$, for any alphabet Σ, if $L_1 \cup L_2$ is not a regular language then at least of L_1 or L_2 is not a regular language either.

(b) For all languages $L_1, L_2 \subseteq \Sigma^\star$, for any alphabet Σ, if $L_1 \circ L_2$ is not a regular language, then at least one of L_1 or L_2 is not a regular language either.

(c) For every language $L \subseteq \Sigma^\star$ over any alphabet Σ, if $L^\star$ is not a regular language then L is not a regular language either.

(d) For every language $L \subseteq \Sigma^\star$ over any alphabet Σ, if L^C is not a regular language then L is not a regular language either.

Closure Properties

Key Observation

- Each of the above results can be used to show that if one language is not a regular language then another (given) language cannot be regular, either.
- These results would not be useful if we did not already know about some languages where are regular. The "Pumping Lemma for Regular Languages" is useful because it can be used to establish that some languages are not regular, without already knowing some nonregular languages.
- Now that we know some nonregular languages, the above closure properties can be used to identify more — and proofs that use closure properties will often be simpler than proofs using the "Pumping Lemma for Regular Languages",

First Example

Let $\Sigma = \{\mathtt{a}, \mathtt{b}\}$. Consider the following languages.

- $L_1 = \{\mathrm{a}^n\mathrm{b}^n \mid n \in \mathbb{Z} \text{ and } n \geq 0\}$
- $L_2 = \{\mathrm{a}^n\mathrm{b}^m \mid n, m \in \mathbb{Z},\ n, m \geq 0 \text{ and } n \neq m\}$
- $L_3 = \{\omega \in \Sigma^\star \mid \mathtt{ba} \text{ is a substring of } \omega\}$
- $L_4 = L_2 \cup L_3$

First Example

This first result is needed so that a closure property can be used.

Lemma: $L_1 = L_4^C$.

Proof: It is necessary, and sufficient, to prove that $L_1 \subseteq L_4^C$ and that $L_4^C \subseteq L_1$.

- Let $\omega \in L_1$. Then $\omega = \mathrm{a}^n\mathrm{b}^n$ for some integer n such that $n \geq 0$.

  It follows that $\omega \notin L_2$ because the number of copies of “a” is equal to the number of copies of “b”.

  It also follows that $\omega \notin L_3$ because $\mathtt{ba}$ is not a substring of ω.

  Thus $\omega \notin L_4$, since $L_4 = L_2 \cup L_3$ — and it follows, by the definition of “L_4^C”, that $\omega \in L_4^C$.

  Since ω was arbitrarily chosen from L_1, this implies that $L_1 \subseteq L_4^C$.

First Example

- Let $\omega \in L_4^C$.
  Then $\omega \in (L_2 \cup L_3)^C = L_2^C \cap L_3^C$, so that $\omega \in L_2^C$ and $\omega \in L_3^C$.
  Since $\omega \in L_3^C$, $\mathtt{ba}$ is not a substring of ω. Thus ω cannot include a copy of "$\mathtt{b}$" after a copy of "$\mathtt{a}$",[1] so that $\omega = \mathtt{a}^n\mathtt{b}^m$ for some integers m and n such that $m, n \geq 0$.

  Since $\omega \notin L_3$, $n = m$ — so that $\omega = \mathtt{a}^n\mathtt{b}^n$ and $\omega \in L_1$.

  Since ω was arbitrarily chosen from L_4^C, it follows that $L_4^C \subseteq L_1$.

Since $L_1 \subseteq L_4^C$ and $L_4^C \subseteq L_1$ it follows that $L_1 = L_4^C$, as claimed.
$\square$

[1] To see this, consider the *last* copy of "$\mathtt{b}$" in ω that has a copy of "$\mathtt{a}$" immediately after it. What symbol must appear immediately after this copy of $\mathtt{b}$?

First Example

Claim: L_4 is not a regular language.

Proof: Suppose, to obtain a contradiction, that L_4 is a regular language.

- Then it follows by the ***closure*** of the set of regular languages under ***complementation*** (Theorem #1, above) that L_4^C is a regular language too.
- Since $L_1 = L_4^C$ (by the previous Lemma), L_1 is a regular language.
- However, we proved that L_1 is ***not*** a regular language (using the "Pumping Lemma for Regular Languages") during the previous lecture.
- A ***contradiction*** has been obtained, so our assumption must be false: L_4 is not a regular language. $\square$

Second Example

Claim: L_2 is not a regular language.

Proof: Suppose, to obtain a contradiction, that L_2 is a regular language.

- The language L_3 is certainly a regular language, since this is the language of the regular expression

$$((((\Sigma)^\star \circ \mathtt{b}) \circ \mathtt{a}) \circ (\Sigma)^\star)$$

- It follows by the ***closure*** of the set of regular operations under ***union*** (Property (a) from the beginning of this lecture) that $L_2 \cup L_3$ must be a regular language.
- However, $L_2 \cup L_3 = L_4$, and it follows by the previous claim that L_4 is ***not*** a regular language.
- A ***contradiction*** has been obtained, so our assumption must be false: L_2 is not a regular language. $\square$

Summary of Process

To prove that a language $L \subseteq \Sigma^\star$ (for some alphabet Σ) is not regular . . .

Assume — to obtain a ***contradiction*** — that L *is* a regular language.

- If you have already proved (or know) that L^C is *not* a regular language, then you should note that — by the closure of the set of regular languages under ***complementation*** — L^C *is* a regular language, so that a ***contradiction*** has been obtained.
- If you have already proved (or know) that $L^\star$ is *not* a regular language, then you should note that — by the closure of the set of regular languages under the ***Kleene star*** operation — that $L^\star$ *is* a regular language, so that a contradiction has been obtained.

Summary of Process

If you *do not* know (and cannot prove) that either L^C or $L^\star$ is not regular, then you should (somehow) find another language $\widehat{L} \subseteq \Sigma^\star$ such that one of the languages $L \cup \widehat{L}$, $L \circ \widehat{L}$, or $\widehat{L} \circ L$ is not a regular language.

- If you have already proved (or know) that $L \cup \widehat{L}$ is *not* a regular language, then you should note that — by the closure of the set of regular languages under ***union*** — $L \cup \widehat{L}$ *is* a regular language, so that a ***contradiction*** has been obtained.
- If you have already proved (or know) that either $L \circ \widehat{L}$ or $\widehat{L} \circ L$ is not a regular language, then you should note that — by the closure of the set of regular languages under ***concatenation*** — $L \circ \widehat{L}$ and $\widehat{L} \circ L$ are both regular languages, so that a ***contradiction*** has been obtained here, as well.

Summary of Process

In every case you should note that the ***contradiction*** establishes that the original assumption was false: L is not a regular language.

- Sometimes, you might need to identify other “intermediate” languages and apply this process several times, to prove that the intermediate languages are not regular, before you can use this to show that the *language you are interested in* is not regular.

  For example, you can think of the *pair* of examples, before this, as a *single* example, where the fact that L_1 is not a regular language is used to prove that L_2 is not a regular language — with L_4 used as an “intermediate” language that is proved not to be regular, along the way.

Summary of Process

- Additional ***closure properties*** for the set of regular languages can be established, and then used (to prove that languages are not regular) in, essentially, the same way.
- The lecture presentation will include more about this.