Teoría de Lenguajes

Clase Teórica 5

Lema de ”Pumping” y Propiedades de Lenguajes Regulares

Primer cuartimestre 2016

Material compilado por Julio Jacobo a lo largo de distintas ediciones de la materia Teoría de Lenguajes en el Departamento de Computación, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, revisado recientemente por Verónica Becher.

Bibliografía: Capítulo 4, *Introduction to Automata Theory, Languages and Computation*, J. Hopcroft, R. Motwani, J. Ullman, Second Edition, Addison Wesley, 2001.

Lema (”Pumping”, Scott & Rabin 1959, Bar Hillel, Perles & Shamir 1961)**.** *Sea L un lenguaje regular. Existe entonces una longitud n tal que*

todas las cadenas z de L de longitud mayor o igual que n pueden ser escritas de la forma

$$z = uvw,$$

donde

$$|uv| \leq n,$$

$$|v| \geq 1,$$

$$\forall i \ \geq 0,\ uv^i w \in L.$$

Empecemos con un par de ejemplos de aplicación. Luego daremos la demostración.

Ejemplo. $L = \{a^k b^k :\ k \geq 0\}$ *sobre* $\Sigma = \{a, b\}$*no es regular.*

Demostración. Supongamos L es regular. Sea n la constante del Lema de Pumping.

Sea $z = a^n b^n$. Por el Lema de Pumping, hay una descomposicion $z = uvw$ con $|uv| \leq n$ y $|v| \geq 1$ tal que $uv^i w$ en L, para $i \geq 0$.

Usando que $|uv| \leq n$, concluimos que v consiste solamente de $a$es.
Más aún, como $|v| \geq 1$, v contiene al menos una a.

Ahora bombeamos v y obtenemos $uv^2 w$.
Por Lema de Pumping, $uv^2 w$ está en L.
Pero $uv^2 w$ tiene más $a$es que bs, entonces no está en L.

Llegamos a una contradicción que provino de asumir que L es regular. Por lo tanto, L no es regular. □

Ejemplo. $L = \{0^{k^2} :\ k \geq 1\}$ *sobre* $\Sigma = \{0\}$ *no es regular.*

Demostración. Supongamos L es regular. Sea n la constante del Lema de Pumping.

Sea $z = 0^{n \times n}$. Por el Lema de Pumping, hay una descomposición $z = uvw$ con $|uv| \leq n$ y $|v| \geq 1$ tal que $uv^i w$ en L, para $i \geq 0$.

Entonces $v = 0^m$ para algun valor de m entre 1 y n.
Por el Lema de Pumping, $uw = 0^{n \times n - m}$ está en L.

Dado que $1 \leq m < n + 1$, y asumiendo que $n \geq 2$ Entonces

$$n \times n - m = n^2 - m > n^2 - (n + 1) \geq n^2 - (2n - 1) = n^2 - 2n + 1 = (n - 1)^2$$

Entonces $n^2 - m$ no es cuadrado perfecto y por lo tanto uw no está en L. .

La contradicción que provino de asumir que L es regular.
Por lo tanto, L no es regular. □

Ejemplo. $L = \{w : \ w$ *tiene la misma cantidad de* $0s$ *que de* $1s\}$ *no es regular.*

Demostración. Supongamos L es regular. Sea n la constante del Lema de Pumping.

Sea $z = 0^n 1^n$. Por el Lema de Pumping, hay una descomposición $z = uvw$ con $|uv| \leq n$ y $|v| \geq 1$ tal que $uv^i w$ en L, para $i \geq 0$.

Entonces v tiene exclusivamente 0s.

Luego, uw tiene distinta cantidad de 0s que de 1s.

Por lo tanto, uw no está en L.

La contradicción que provino de asumir que L es regular.

□

Definición (Configuración instantánea de un AFD). *Es un par* (q, α) *en* $Q \times \Sigma^*$ *donde* q *es el estado en el que está el autómata y* α *es la cadena de entrada aún no consumida.*

Definición (Transición entre configuraciones instantáneas $\vdash$). *Llamamos transición a la siguiente relación sobre* $Q \times \Sigma^*$*:*
$(q, \alpha) \vdash (p, \beta)$ *si* $(\delta(q, a) = p \wedge \alpha = a\beta)$.

De lo anterior tenemos que $(q, \alpha\beta) \overset{*}{\vdash} (p, \beta)$ si y sólo si se puede pasar del estado q al estado p consumiendo la cadena α.

Lema. *Sea el AFD* $M =< Q, \Sigma, \delta, q_0, F >$
Para todo $q \in Q$ *y* $\alpha, \beta \in \Sigma^*$,

$$\text{si } \ (q, \alpha\beta) \overset{*}{\vdash} (q, \beta) \ \text{ entonces } \forall i \geq 0, \ \left(q, \alpha^i \beta\right) \overset{*}{\vdash} (q, \beta)\,.$$

Demostración. Queremos probar, $q \in Q$ y $\alpha, \beta \in \Sigma^*$: Si $\ (q, \alpha\beta) \overset{*}{\vdash} (q, \beta)\ $ entonces $\forall i \geq 0, \ \left(q, \alpha^i \beta\right) \overset{*}{\vdash} (q, \beta)$.

Fijemos $\alpha \in \Sigma^*$ y $q \in Q$. Asumamos $\forall \beta \in \Sigma^*$, $(q, \alpha\beta) \overset{*}{\vdash} (q, \beta)$.

Caso base $(k = 0)$ $\left(q, \alpha^0 \beta\right) \overset{0}{\vdash} (q, \beta)$
Caso inductivo. Supongamos que vale para k, es decir,
Si $(q, \alpha\beta) \overset{*}{\vdash} (q, \beta)$ entonces $\left(q, \alpha^k \beta\right) \overset{*}{\vdash} (q, \beta)$.
Veamos que vale para $k + 1$.

Por definición,	$\left(q, \alpha^{k+1}\beta\right) = \left(q, \alpha\alpha^k \beta\right)$
Por la suposición arriba,	$\left(q, \alpha\alpha^k \beta\right) \overset{*}{\vdash} \left(q, \alpha^k \beta\right)$,
Por hipótesis inductiva,	$\left(q, \alpha^k \beta\right) \overset{*}{\vdash} (q, \beta)$.
Por lo tanto,	$\left(q, \alpha^{k+1}\beta\right) \overset{*}{\vdash} (q, \beta)$.

□

Lema ("Pumping", Scott & Rabin 19659, Bar Hillel, Perles & Shamir 1961). *Sea* L *un lenguaje regular. Existe entonces una longitud* n *tal que*

todas las cadenas z *de* L *de longitud mayor o igual que* n *pueden ser escritas de la forma*

$$z = uvw,$$

donde

$$|uv| \leq n,$$

$$|v| \geq 1,$$

$$\forall i \ \geq 0, \, uv^i w \in L.$$

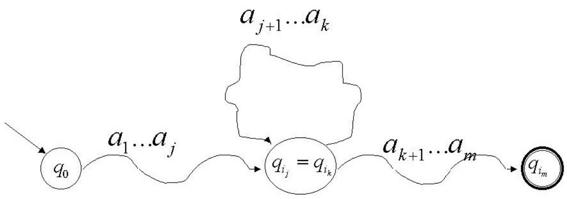
Demostración. Sea AFD M tal que $\mathcal{L}(M) = L$. Sea n su cantidad de estados. Sea z una cadena de longitud $m \geq n$, $z = a_1 \cdots a_m,$.

Para aceptar z el autómata debe usar tantas transiciones como la longitud de z, es decir, m transiciones.

Por lo tanto , debemos considerar $m + 1$ estados (estado inicial, estado luego de consumir el primer símbolo, estado luego de consumir el segundo símbolo, etc).

Como $m + 1 > n$, para aceptar z el autóma debe pasar DOS ó más veces por un mismo estado.

Sea $q_{\ell_0}, q_{\ell_1}, \cdots, q_{\ell_m}$, con $q_{\ell_0} = q_0$ y q_{ℓ_m} un estado final, la sucesión de estados hasta aceptar z.



Como $m+1 > n$, existen j y k, con $0 \leq j < k \leq n$, tales que $q_{\ell_j} = q_{\ell_k}$.
Partamos la cadena z en tres cadenas u, v y w de la siguiente manera:

$$u = \begin{cases} a_1 \cdots a_j & \text{si } j > 0 \\ \lambda & \text{si } j = 0 \end{cases}$$

$$v = a_{j+1} \cdots a_k$$

$$w = \begin{cases} a_{k+1} \cdots a_m & \text{si } k < m \\ \lambda & \text{si } k = m. \end{cases}$$

Entonces, $|uv| \leq n$, $|v| \geq 1$ y

$$(q_0, uvw) \overset{*}{\vdash} \left(q_{\ell_j}, vw\right) \overset{*}{\vdash} (q_{\ell_k}, w) \overset{*}{\vdash} (q_{\ell_m}, \lambda)\,.$$

Pero, como $q_{\ell_j} = q_{\ell_k}$,$\forall i \geq 0$, $\quad \left(q_{\ell_j}, v^i w\right) \overset{*}{\vdash} \left(q_{\ell_j}, w\right) = (q_{\ell_k}, w)\,.$
Por lo tanto, $uv^i w \in L$. $\qquad \square$.

Decidir Verdadero o Falso

Sea AFD $M =< Q, \Sigma, \delta, q_0, F >$, con $|Q| = n$.

1. $\mathcal{L}(M)$ es no vacío si y solo si existe w en Σ^* tal que

   $\widehat{\delta}(q_0, w) \in F$ y $|w| < n$.

   Es verdadero

2. $\mathcal{L}(M)$ es infinito si y solo si existe w en Σ^* tal que

   $\widehat{\delta}(q_0, w) \in F$ y $n \leq |w| < 2n$.

   Es Verdadero

Veamos la demostración de las afirmaciones.
Demostración.
1. Debemos ver que $\mathcal{L}(M)$ es no vacío si y solo si existe w en Σ^* tal que $\widehat{\delta}(q_0, w) \in F$ y $|w| < n$, donde $n = |Q|$.

$\Rightarrow$). Supongamos $\mathcal{L}(M)$ es no vacío. Sabemos que es regular.
Sea z en $\mathcal{L}(M)$ de longitud mayor o igual a n y supongamos que no hay ninguna más corta en $\mathcal{L}(M)$.

El Lema de Pumping garantiza que hay u, v, w apropiados tal que $|v| \geq 1$, $z = uvw$ y $\forall i \geq 0$, $uv^i w$ en $\mathcal{L}(M)$.

Entonces uw está en $\mathcal{L}(M)$. Pero uw es más corta que z, lo que contradice nuestra suposición de que z era la más corta.

Concluimos que en $\mathcal{L}(M)$ hay cadenas más cortas que n.

$\Leftarrow$). Es obvio que L es no vacío.

2. Debemos ver que $\mathcal{L}(M)$ es infinito si y solo si existe w en Σ^* tal que $\widehat{\delta}(q_0, w) \in F$ y $n \leq |w| < 2n$, donde $n = |Q|$.

$\Rightarrow$). Supongamos $\mathcal{L}(M)$ es infinito. Supongamos que no hay ninguna cadena en $\mathcal{L}(M)$ de longitud entre n y $2n-1$. Sin pérdida de generalidad, sea z en $\mathcal{L}(M)$ de longitud $2n$ (si la longitd es otra aplica el mismo argumento, usandolo tantas veces como haga falta hasta llegar a la contradicción buscada).

Por el Lema de Pumping, hay u, v, w tal que $z = uvw$, con $|uv| \leq n$, $|v| \geq 1$ y $\forall i \geq 0$ $uv^i w$ está en $\mathcal{L}(M)$. Entonces uw está en $\mathcal{L}(M)$.

Como $1 \leq |v| \leq n$, la longitud de uw es $n \leq |uv| \leq 2n-1$, contradiciendo que no había ninguna en $\mathcal{L}(M)$ de esta longitud.

$\Leftarrow$). Supongamos z pertenece a $\mathcal{L}(M)$ y $n \leq z < 2n$.

Por el Lema de Pumping $z = uvw$ y para todo $i \geq 0$, $uv^i w$ esta en $\mathcal{L}(M)$. Luego $\mathcal{L}(M)$ es infinito.

Teorema. *El conjunto de lenguajes regulares incluido en Σ^* es cerrado respecto de la unión.*

Demostración. Sean L_1 y L_2 lenguajes regulares. Debemos pobar que $L_1 \cup L_2$ es un lenguaje regular.

Sean $M_1 =< Q_1, \Sigma, \delta_1, q_1, F_1 >$ y $M_2 =< Q_2, \Sigma, \delta_2, q_2, F_2 >$, donde $Q_1 \cap Q_2 = \varnothing$ tales que $L_1 = \mathcal{L}(M_1)$ y $L_2 = \mathcal{L}(M_2)$.

Definimos $M =< Q_1 \cup Q_2 \cup \{q_0\}, \Sigma, \delta, q_0, F >$, con $\{q_0\} \cap Q_1 = \varnothing$ y $\{q_0\} \cap Q_2 = \varnothing$, donde

- q_0 es un nuevo estado.
- *si* $\lambda \notin L_1$ y $\lambda \notin L_2$ $F = F_1 \cup F_2$. Sino, $F = F_1 \cup F_2 \cup \{q_0\}$.
- $\delta(q_0, a) = \delta(q_1, a) \cup \delta(q_2, a)$.
- $\forall q \in Q_1, \forall a \in \Sigma,\ \delta(q, a) = \delta_1(q, a)$.
- $\forall q \in Q_2, \forall a \in \Sigma,\ \delta(q, a) = \delta_2(q, a)$.

El autómata M así construido es no-determinístico.

Se demuestra por inducción en el largo $i \geq 1$ de la cadena w: $(q_0, w) \overset{i}{\underset{M}{\vdash}} (q, \lambda)$ si y solo si

$$\left(q \in Q_1 \wedge (q_1, w) \overset{i}{\underset{M_1}{\vdash}} (q, \lambda)\right) \vee \left(q \in Q_2 \wedge (q_2, w) \overset{i}{\underset{M_2}{\vdash}} (q, \lambda)\right).$$

□

Teorema. *El conjunto de lenguajes regulares incluido en Σ^* es cerrado respecto de la concatenación.*

Teorema. *El conjunto de lenguajes regulares incluido en Σ^* es cerrado respecto de la complementación.*

Demostración. Sea $L \subseteq \Delta^*$ regular, con $\Delta \subseteq \Sigma$.

Sea $M =< Q, \Delta, \delta, q_0, F >$ tal que $L = \mathcal{L}(M)$.

El AFD $M' =< Q, \Delta, \delta, q_0, Q - F >$ acepta $\Delta^* - \mathcal{L}(M)$.
Es decir, $\mathcal{L}(M') = \Delta^* - \mathcal{L}(M)$.

En caso de que $\Delta = \Sigma$, $\overline{\mathcal{L}(M)} = \mathcal{L}(M')$, por lo tanto es regular.

En caso de que $\Delta \subset \Sigma$, $\overline{\mathcal{L}(M)} = \mathcal{L}(M') \cup \Sigma^* (\Sigma - \Delta) \Sigma^*$,
que es la unión de dos lenguajes regulares y por lo tanto regular. □

Teorema. *El conjunto de lenguajes regulares incluido en Σ^* es cerrado respecto de la intersección.*

Demostración. Sean L_1 y L_2 regulares.

$$L_1 \cap L_2 = \overline{\overline{L_1 \cap L_2}} = \overline{\overline{L_1} \cup \overline{L_2}}.$$

Dado que los lenguajes regulares están cerrdos por unión y complementación, concluimos que $L_1 \cap L_2$ es regular. □

Una demostración alternativa de que ell conjunto de lenguajes regulares está cerrado por intersección.

Demostración. Dados M_1 y M_2 AFDs, definiremos M' tal que $\mathcal{L}(M') = \mathcal{L}(M_1) \cap \mathcal{L}(M_2)$.

Sea $M^{'} =< Q', \Sigma, \delta', q_0', F' >$ con

- $Q' = Q_1 \times Q_2$
- $\delta'((q, r), a) = (\delta_1(q, a), \delta_2(r, a))$ para $q \in Q_1$ y $r \in Q_2$
- $q_0' = (q_{0_1}, q_{0_2})$
- $F' = F_1 \times F_2$

entonces

$$\begin{aligned}
\alpha \in \mathcal{L}(M^{'}) \quad &\Leftrightarrow \quad \widehat{\delta'}((q_{0_1}, q_{0_2}), \alpha) \in F' \\
&\Leftrightarrow \quad \left(\widehat{\delta_1}(q_{0_1}, \alpha), \widehat{\delta_2}(q_{0_2}, \alpha)\right) \in F_1 \times F_2 \\
&\Leftrightarrow \quad \left(\widehat{\delta_1}(q_{0_1}, \alpha) \in F_1\right) \wedge \left(\widehat{\delta_2}(q_{0_2}, \alpha) \in F_2\right) \\
&\Leftrightarrow \quad \alpha \in \mathcal{L}(M_1) \wedge \alpha \in \mathcal{L}(M_2).
\end{aligned}$$

□

De los teoremas anteriores podemos afirmar:

El conjunto de los lenguajes regulares incluidos en Σ^* es un álgebra Booleana de conjuntos.

Teorema. *La unión finita y la intersección finita de lenguajes regulares dan por resultado un leguaje regular.*

Demostración. Debemos ver que

$$\forall n \in \mathbb{N}, \bigcup_{i=1}^{n} L_i \text{ es regular, y } \forall n \in \mathbb{N}, \bigcap_{i=1}^{n} L_i \text{ es regular.}$$

Por inducción en n.

- Caso base $n = 0$: $\bigcup_{i=1}^{0} L_i = \emptyset$ es regular.
- Caso inductivo: Supongamos que para $n > 0$, $\bigcup_{i=1}^{n} L_i$ es regular. Veamos que vale para $n + 1$.

  $\bigcup_{i=1}^{n+1} L_i = \bigcup_{i=1}^{n} L_i \cup L_{n+1}$ es regular, por ser la union de dos regulares.

La demostración para $\cap$ es similar. □

Observación. *Los lenguajes regulares no están clausurados por unión infinita.*

Demostración. Damos un contraejemplo.

Para cada $i \geq 1$ sea el lenguaje regular $L_i = \{a^i b^i\}$.

Si los lenguajes regulares estuvieran clausurados por unión infinita, $\bigcup_{i=1}^{\infty} L_i$ debería ser regular.

$$\bigcup_{i=1}^{\infty} L_i = \bigcup_{i=1}^{\infty} \{a^i b^i\} = \{a^k b^k : k \in \mathbb{N}\}$$

Usando el el Lema de Pumping se demuestra que no es regular. □

Teorema. *Todo lenguaje finito es regular.*

Demostración. Sea L un lenguaje finito, con n cadenas, $L = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$.

Para cada $i = 1, 2, \ldots n$, sea $L_i = \{\alpha_i\}$.

Entonces $L = \bigcup_{i=1}^{n} \{\alpha_i\}$.

Como cada $\{\alpha_i\}$ es regular, entonces L también lo es. □

Definición. *Un conjunto A de números naturales es decidible si hay un algoritmo que para cualquier número natural responde si pertenece o no al conjunto A.*

La definición de decidibilidad se extiende a otros conjuntos que los naturales, y a otros problemas que la pertenecia. En cada caso significa la existencia de un algoritmos que resuelve la pregunta por sí o por no.

Sea L un lenguaje regular sobre Σ. Decidir cuales de los siguientes problemas sobre lenguajes regulares son decidibles.

1. (Pertenencia) Para toda $\alpha \in \Sigma^*$, ¿pertenece α a L?

   Sí es decidible

2. (Finitud) ¿es L finito?

   Sí es decidible

3. (Vacuidad)¿es L vacío?

   Sí es decidible

4. (Equivalencia) Dados los lenguajes regulares L_1 y L_2,

   ¿son L_1 y L_2 equivalentes?

   Sí es decidible

Demostremos estas afirmaciones.
Demostración
1. Pertenencia: Dado el lenguaje regular L,
- se construye su AFD M tal que $\mathcal{L}(M) = L$.
- *si* α es aceptada, *entonces* pertenece a L y *sino* no.

2. Finitud: El lenguaje regular L es infinito si y solo si cada AFD $M =< Q, \Sigma, \delta, q_0, F >$ tal que $\mathcal{L}(M) = L$ acepta al menos una cadena de longitud ℓ donde

$$n \leq \ell < 2|Q|.$$

3. Vacuidad: Dado el lenguaje regular L,
- se construye su AFD M tal que $\mathcal{L}(M) = L$
- se determina el conjunto A de estados alcanzables.
- *si* $F \cap A = \emptyset$ *entonces* el lenguaje L es vacío y *sino* no.

4. Equivalencia: Dados los lenguajes regulares L_1 y L_2, aceptados por los autómatas M_1 y M_2 respectivamente, si el lenguaje regular

$$\left(L_1 \cap \overline{L_2}\right) \cup \left(\overline{L_1} \cap L_2\right)$$

es vacío entonces L_1 y L_2 son equivalentes, sino no lo son.

Ejercicios

1. Demostrar que los siguientes lenguajes no son regulares

   $L = \{a^i b^j : i \neq j\}$

   $L_a = \{xay : x \in \Sigma^*, y \in \Sigma^*, |x| = |y|\}$, donde a es un elemento prefijado de Σ.

2. Sea L un lenguaje regular, y sea n la constante del Lema de Pumping para L. Indicar Verdadero o Falso y justificar.

   - Para cada cadena z en L, con $|z| \geq n$, la descomposición de z en uvw, con $|v| \geq 1$ y $|uv| \leq n$, es única.
   - Cada cadena z en L, con $|z| \geq n$, es de la forma uv^iw para algún u, v, w, con $|v| \geq 1$ y $|uv| \leq n$ y algun i.
   - Hay lenguajes no regulares que satisfacen la condición afirmada en el Lema de Pumping.

3. Indicar Verdadero o Falso y justificar:

   Sean L_1 y L_2 lenguajes sobre el alfabeto Σ, tal que $L_1 \cup L_2$ es regular. Entonces, tanto L_1 como L_2 son regulares.

Ejercicios

4. Sea $\mathcal{C}$ el mínimimo conjunto que contiene a todos los lenguajes finitos, y está cerrada por unión finita, intersección, complemento y concatenación. ¿ Cuál es la relación entre $\mathcal{C}$ y el conjunto de todos los lenguajes regulares?

5. Dar un algoritmo de decisión que determine si el lenguaje aceptado por un autómata finito es el conjunto de todas las cadenas del alfabeto.

6. Dar un algoritmo de decisión que determine si el lenguaje aceptado por un autómata finito es cofinito.