Ciencias de la Computación II

Lenguajes Regulares

Eduardo Contrera Schneider

Universidad de la Frontera

24 de agosto de 2016

Introducción

De ahora en adelante nos ocuparemos de la definición de lenguajes, es decir, especificar qué cadenas pertenecen a un lenguaje en particular. Sería natural entonces preguntarse cuántos lenguajes se pueden definir sobre un alfabeto Σ. A partir de este punto, la teoría de conjuntos nos prestará una ayuda invaluable.

Cardinalidad de Σ^*

Consideremos el alfabeto $\Sigma = \{a, b\}$. Es evidente que para todo $n \in \mathbb{N}$ existe un número finito de palabras con longitud n. Las cadenas formadas con los símbolos de Σ pueden ser ordenadas lexicográficamente (orden del diccionario). Así las cosas, podemos numerar al principio las palabras de longitud 0, después las de longitud 1 y así sucesivamente. De esta manera, podemos ver que, a pesar de ser de Σ^* es un conjunto infinito, es al menos numerable.
Por argumentos prestados de la teoría de conjuntos, podemos deducir que el conjunto de todos los lenguajes sobre Σ es infinito no numerable.

Lenguajes Regulares

Los lenguajes regulares son un tipo de lenguaje interesante por el hecho de que sirven para especificar las construcción de analizadores léxicos. Constituyen el menor conjunto de lenguajes sobre Σ que es cerrado con respecto a las operaciones de concatenación, cerradura de Kleene y unión de lenguajes.

Lenguaje Regular

Sea Σ un alfabeto. El conjunto de los leguajes regulares sobre Σ se define recursivamente como sigue:

1. $\emptyset$ es un lenguaje regular.
2. $\{\epsilon\}$ es un lenguaje regular.
3. Para todo $a \in \Sigma$, $\{a\}$ es un lenguaje regular.
4. Si A y B son lenguajes regulares, entonces $A \cup B$, $A \cdot B$ y A^* son lenguajes regulares.
5. Ningún otro lenguaje sobre Σ es regular.

Ejemplo

Dado $\Sigma = \{a, b\}$, son ciertas las siguiente afirmaciones:

- $\emptyset$ y $\{\epsilon\}$ son lenguajes regulares.
- $\{a\}$ y $\{b\}$ son lenguajes regulares.
- $\{a, b\}$ es regular.
- $\{a, ab, b\}$ es regular.
- $\{a^i | i \geq 0\}$ es regular.
- $\{a^i b^j | i \geq 0 \quad y \quad j \geq 0\}$ es regular.
- $\{(ab)^i | i \geq 0\}$ es regular.

Expresiones Regulares

Ahora respondamos la siguiente pregunta. ¿El lenguaje de todas las cadenas sobre $\{a, b, c\}$ que no tienen ninguna subcadena ac es un lenguaje regular? Si llamamos A a tal lenguaje, no es difícil establecer que $A = \{c\}^*(\{a\} \cup \{b\}\{c\}^*)^*$.
Así las cosas, podemos simplificar la especificación de un lenguaje regular mediante la introducción de un especie de abreviatura llamada **expresión regular**. Primero, se escribirá a en lugar de $\{a\}$. Por ende,

1. $a \cup b \equiv \{a\} \cup \{b\}$
2. $ab \equiv \{ab\}$
3. $a^* \equiv \{a\}^*$
4. $a^+ \equiv \{a\}^+$

Para expresiones regulares tenemos,

1. $\emptyset$ y ϵ son expresiones regulares.
2. Para todo $a \in \Sigma$, a es una expresión regular.
3. Si r y s son expresiones regulares, entonces $r \cup s$, $r \cdot s$ y r^* son expresiones regulares.
4. Ninguna otra secuencia de símboles es una expresión regular.

Es evidente que toda expresión regular r denota un lenguaje regular sobre Σ, para el cual usaremos $L(r)$. Dos expresiones regulares son equivalentes si denotan los mismos lenguajes. Por último, cabe destacar que hay muchas expresiones regulares para denotar un mismo lenguaje. A modo de ejemplo, $(a^*b)^*$ y $\epsilon \cup (a \cup b)^*b$.

Proposiciones

Algunas equivalencias son

Teorema

Sean r, s y t expresiones regulares sobre el mismo alfabeto Σ. Entonces:

1. $r \cup s = s \cup r$
2. $r \cup \emptyset = \emptyset \cup r = r$
3. $r \cup r$
4. $r \cup (s \cup t) = (r \cup s) \cup t$
5. $r\epsilon = \epsilon r = r$
6. $r\emptyset = \emptyset r = \emptyset$
7. $(rs)t = r(st)$
8. $r(sr)^* = r(sr)^*$

Diagramas de Transición

Si dada una cadena w que pertenece a un lenguaje regular A, se nos pregunta si tal cadena pertenece a A o no, debemos analizar no sólo sus caracteres que aparecen en w, sino también sus posiciones relativas. Podemos construir un diagrama que nos ayude con la tarea. Este diagrama tiene forma de un grafo dirigido con información adicional añadida, y se llama **diagrama de transición**, que tiene las siguientes características:

- Los nodos se llaman estados y se usan para señalar hasta qué lugar se ha analizado la cadena.
- Las aristas se etiquetan con caracteres del alfabeto y se llaman transiciones.
- Si el siguiente caracter a reconocer concuerda con alguna etiqueta de transción desde el estado actual, se desplaza al estado que lleva la arista correspondiente.

La pregunta que surge ahora es, ¿cuándo la cadena es legal? Para determinarlo, se marcan ciertos estados como de aceptación o estados finales, de manera que si al finalizar el análisis se queda en alguno de estos estados, entonces la cadena es legal. Veamos algunos ejemplos.

Ejemplos

- La expresión $a^k b$ para $k \geq 0$.
- El lenguaje $A = \{(ab)^i | i \geq 1\}$.
- La expresión $(ab)^+$.

Automáta Finito Determinista

Un autómata finito determinista M es una colección de cinco elementos.

1. Un alfabeto de entrada Σ.
2. Una colección finita de estados Q.
3. Un estado inicial s.
4. Una colección F de estados finitos o de aceptación.
5. Una función $\delta : Q \times \Sigma \to Q$ que determina el único estado siguiente para el par (q_i, σ) correspondiente al estado actual y la entrada.

El término autómata finito determinista se abrevia *AFD* y se usa la notación $M = (Q, \Sigma, s, F, \delta)$ para indicar todos los conjuntos indicados anteriormente.

Ejemplos

Cree los diagramas de transición para los siguientes AFD:

1. $Q = \{q_0, q_1\}$, $\Sigma = \{a, b\}$, $F = \{q_0\}$, $s = q_0$ y δ representado por

δ	a	b
q_0	q_0	q_1
q_1	q_1	q_0

2. $Q = \{q_0, q_1, q_2, q_3\}$, $\Sigma = \{a, b\}$, $F = \{q_0, q_1, q_2\}$, $s = q_0$ y δ representado por

δ	a	b
q_0	q_0	q_1
q_1	q_0	q_2
q_2	q_0	q_3
q_3	q_3	q_3