Autómatas y Lenguajes Formales 2017-1
Facultad de Ciencias UNAM
Nota de Clase 2

Favio E. Miranda Perea   A. Liliana Reyes Cabello   Lourdes González Huesca

18 de agosto de 2016

1. Lenguajes Regulares

El estudio de los lenguajes como conjuntos de cadenas debe ser susceptible a realizarse de manera formal mediante una abstracción conveniente.

Una forma para caracterizar lenguajes sencillos, de fácil descripción, es utilizar operaciones sobre cadenas, éstas permiten la *generación* de nuevos lenguajes. Los lenguajes regulares son los conjuntos de cadenas más simples.

Definición 1 *Un lenguaje $L \subseteq \Sigma^\star$ es regular si es generado a partir de los lenguajes básicos $\varnothing, \{\varepsilon\}, \{a\}$ (donde $a \in \Sigma$) y mediante las operaciones de unión, concatenación y estrella de Kleene. Recursivamente:*

- *$\varnothing$ y $\{\varepsilon\}$ son lenguajes regulares.*
- *Si $a \in \Sigma$ entonces $\{a\}$ es regular.*
- *Si L, M son regulares entonces $L \cup M, LM$ y $L^\star$ son regulares.*
- *Son todos.*

Ejemplos:

- Cualquier lenguaje *finito* es regular.
- En efecto si $L = \{w_1, \ldots, w_n\} \subseteq \Sigma^\star$ entonces $L = L_1 \cup L_2 \cup \ldots \cup L_n$ con $L_i = \{w_i\}$. Cada lenguaje L_i es regular puesto que si $w_i = a_1 \ldots a_{k_i}$ con $a_j \in \Sigma$ entonces $L_i = \{a_1\}\{a_2\} \ldots \{a_{k_i}\}$ y cada $\{a_j\}$ es regular por definición.
- En particular cualquier alfabeto Σ es un lenguaje regular.
- Sobre el alfabeto $\Sigma = \{a, b\}$ tenemos los siguientes lenguajes regulares *infinitos*:
  - $L_1 = \{w \in \Sigma^\star \mid w \text{ empieza con } b\} = \{b\}\Sigma^\star$
  - $L_2 = \{w \in \Sigma^\star \mid w \text{ contiene exactamente una } a\} = \{b\}^\star\{a\}\{b\}^\star$

- $L_3 = \{w \in \Sigma^\star \mid w \text{ contiene la subcadena } ba\} = \Sigma^\star\{ba\}\Sigma^\star$
- $L_4 = \{w \in \Sigma^\star \mid w \text{ contiene exactamente dos } b\} = \{a\}^\star\{b\}\{a\}^\star\{b\}\{a\}^\star$
- $L_5 = \{w \in \Sigma^\star \mid w \text{ contiene un número par de } a\} = \{b\}^\star \cup \left(\{b\}^\star\{a\}\{b\}^\star\{a\}\{b\}^\star\right)^\star$

- No todo lenguaje infinito es regular

Problemas de interés acerca de lenguajes regulares Dado un lenguaje regular $L \subseteq \Sigma^\star$ existen diversos problemas de decisión (es decir, problemas cuya respuesta es sí o no) acerca de L, enunciamos algunos de ellos que tienen solución algorítmica.

1. Problema de vacuidad: ¿Es L el lenguaje vacío?
2. Problema de finitud: ¿Es L finito ?
3. Problema de la pertenencia: Dada $w \in \Sigma^\star$, ¿$w \in L$?
4. Problema de la equivalencia: Dado $M \subseteq \Sigma^\star$, ¿$M = L$?

1.1. Expresiones Regulares

Los lenguajes regulares se pueden describir por medio de una expresión que generaliza la forma de las cadenas en el lenguaje, es decir una representación de las cadenas. Un mecanismo de suma importancia para denotar lenguajes regulares es por medio de las llamadas expresiones regulares, que son *fórmulas* explícitas de la cadena característica del lenguaje.

Definición 2 *Las expresiones regulares están definidas recursivamente como sigue:*

- $\varnothing$ *es una expresión regular.*
- ε *es una expresión regular.*
- *Si* $a \in \Sigma$ *entonces* a *es una expresión regular.*
- *Si* α, β *son expresiones regulares entonces* $\alpha\beta,\ \alpha + \beta,\ \alpha^\star$ *son expresiones regulares.*
- *Son todas.*

Dada una expresión regular α, su significado $L(\alpha)$ es un lenguaje regular definido como sigue:

- $L(\varnothing) = \varnothing$
- $L(\varepsilon) = \varepsilon$
- $L(a) = \{a\} \quad \text{si } \alpha \in \Sigma$
- $L(\alpha\beta) = L(\alpha)L(\beta)$
- $L(\alpha + \beta) = L(\alpha) \cup L(\beta)$
- $L(\alpha^\star) = L(\alpha)^\star$

De los significados de las expresiones, se puede ver a una expresión regular como una representación alternativa de un lenguaje. Así podemos pensar en la equivalencia entre lenguajes regulares y expresiones regulares:

Proposición 1 *Un lenguaje $M \subseteq \Sigma^\star$ es regular si y sólo si existe una expresión regular α tal que $M = L(\alpha)$.*

Demostración. La dirección $\Leftarrow$ es directa. Para la otra dirección hacemos inducción estructural.

Se puede estudiar la equivalencia de expresiones regulares:

Definición 3 *Dadas dos expresiones regulares α, β decimos que son equivalentes, al escribir $\alpha = \beta$, si y sólo si α y β tienen el mismo significado, es decir, si y sólo si $L(\alpha) = L(\beta)$.*

Propiedades de las Expresiones Regulares

$$
\begin{array}{cc}
\alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma & \alpha + \beta = \beta + \alpha \\
\alpha + \varnothing = \alpha & \alpha + \alpha = \alpha \\
\alpha\varepsilon = \alpha & \alpha\varnothing = \varnothing \\
\alpha(\beta\gamma) = (\alpha\beta)\gamma & \\
\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma & (\beta + \gamma)\alpha = \beta\alpha + \gamma\alpha \\
\varepsilon^\star = \varepsilon & \varnothing^\star = \varepsilon \\
\alpha\alpha^\star = \alpha^\star\alpha & \alpha^\star = \alpha^\star\alpha^\star = (\alpha^\star)^\star \\
\alpha^\star = \varepsilon + \alpha\alpha^\star & (\alpha + \beta)^\star = (\alpha^\star + \beta^\star)^\star \\
\text{Si } L(\alpha) \subseteq L(\beta) \text{ entonces } \alpha + \beta = \beta &
\end{array}
$$

Propiedades de cerradura Las propiedades de cerradura nos permiten construir nuevos lenguajes regulares a partir de lenguajes ya conocidos por medio de algunas operaciones entre lenguajes. Si L, M son lenguajes regulares entonces:

- $L \cup M$ es regular.
- LM es regular.
- $L^\star$ es regular.
- L^+ es regular.
- $\overline{L}$ es regular
- $L \cap M$ es regular.
- $L - M$ es regular.

Aplicaciones de las expresiones regulares

- Búsqueda y sustitución en editores de texto (vi, emacs, etc.)
- Caza de patrones (pattern matching) y procesamiento de textos y conjuntos de datos, por ejemplo en minería de datos (grep, sed, awk, perl, etc.)

- Implementación de lenguajes de programación (fase de análisis léxico): conversión del programa fuente en una sucesión de elementos léxicos (lexemas o tokens), en esta sucesión se identifican las distintas componentes de un programa, como son palabras reservadas, identificadores, tipos de datos, etc.(lex, flex, etc.)

No todos los lenguajes son regulares, por ejemplo:

$$L = \{a^n b^n \mid n \geq 1\}$$

no es definible mediante una expresión regular.

- ¿Cómo decidir cuando un lenguaje no es regular?
- Mediante propiedades de cerradura.
- Mediante el lema del bombeo (pendiente).

En el siguiente tema discutiremos más formas de decidir cuándo un lenguaje es regular o no.