Tema 2: Lenguajes Independientes del Contexto

Lección 2.1

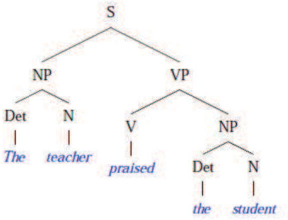
Gramáticas

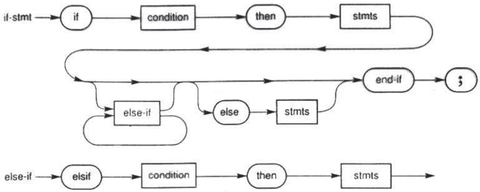
Jordi Bernad

Índice

Las gramáticas aparecen:

- Estructura formal del lenguaje natural: Chomsky
- Descripción formal de la sintaxis de lenguajes de programación: notación Backus-Naur .





Ejemplo de gramática

- Una gramática viene dada por reglas de sustitución:

$$S \rightarrow aSb$$
$$S \rightarrow \epsilon$$

- Utilizamos las reglas de sustitución para generar cadenas
- Empezamos por S

$$S \Rightarrow aSb \Rightarrow aaSbb \Rightarrow aaaSbbb \Rightarrow aaabbb$$

- ¿Qué cadenas generamos con la gramática?

Ejemplo de gramática

- Una gramática viene dada por reglas de sustitución:

$$S \rightarrow aSb$$
$$S \rightarrow \epsilon$$

- Utilizamos las reglas de sustitución para generar cadenas
- Empezamos por S

$$S \Rightarrow aSb \Rightarrow aaSbb \Rightarrow aaaSbbb \Rightarrow aaabbb$$

- ¿Qué cadenas generamos con la gramática?
- $a^n b^n$, con $n \geq 0$

Otro ejemplo

- Si tenemos la siguiente gramática

$$\begin{aligned} S &\rightarrow aS \\ S &\rightarrow T \\ S &\rightarrow Sa \\ T &\rightarrow bTc \\ T &\rightarrow bc \end{aligned}$$

- Podemos abreviar

$$\begin{aligned} S &\rightarrow aS|T|Sa \\ T &\rightarrow bTc|bc \end{aligned}$$

Ejemplo derivación

- $\Sigma = \{a, b, c\}$ son los símbolos de un alfabeto, símbolos terminales; y S, T, variables (no terminales)

$$S \to aS|T|Sa$$
$$T \to bTc|bc$$

- Empezamos por S (start): variable no terminal.
- A partir de S, podemos utilizar una regla de sustitución, por ejemplo, $S \to aS$

$$S \Rightarrow aS$$

- Aplicamos dos veces la regla $S \to Sa$

$$aS \Rightarrow aSa \Rightarrow aSaa$$

- Aplicamos $S \to T$

$$aSaa \Rightarrow aTaa$$

- Aplicamos las reglas $T \to bTc$ y $T \to bc$

$$aTaa \Rightarrow abTcaa \Rightarrow abbccaa$$

Ejemplo derivación

- Hemos visto que la cadena *abbccaa* se deriva de la gramática

$$
\begin{aligned}
S &\to aS|T|Sa \\
T &\to bTc|bc
\end{aligned}
$$

- Otra posible derivación

$$S \Rightarrow Sa \Rightarrow Saa \Rightarrow Taa \Rightarrow bTcaa \Rightarrow bbTccaa \Rightarrow bbbcccaa$$

- ¿Cómo son todas las cadenas de Σ^* que se derivan de la gramática?

Ejemplo derivación

- Hemos visto que la cadena *abbccaa* se deriva de la gramática

$$\begin{aligned} S &\rightarrow aS|T|Sa \\ T &\rightarrow bTc|bc \end{aligned}$$

- Otra posible derivación

$$S \Rightarrow Sa \Rightarrow Saa \Rightarrow Taa \Rightarrow bTcaa \Rightarrow bbTccaa \Rightarrow bbbcccaa$$

- ¿Cómo son todas las cadenas de Σ^* que se derivan de la gramática?
- $a^i b^n c^n a^j$, con $i, j \geq 0$ y $n \geq 1$

Definición formal de gramática

Definición

Una **gramática independiente del contexto (libre del contexto)** es una 4-tupla $G = (N, \Sigma, R, S)$ donde

- N es un alfabeto de **variables (no terminales)**.
- Σ es un alfabeto de **símbolos terminales**.
- R es un conjunto finito de reglas: $R \subset N \times (N \cup \Sigma)^*$.
- $S \in N$.

Definición formal de gramática

Definición

Una **gramática independiente del contexto (libre del contexto)** es una 4-tupla $G = (N, \Sigma, R, S)$ donde

- N es un alfabeto de **variables (no terminales)**.
- Σ es un alfabeto de **símbolos terminales**.
- R es un conjunto finito de reglas: $R \subset N \times (N \cup \Sigma)^*$.
- $S \in N$.

- Notar que la parte de la izquierda de las reglas está formada por una única variable.

Definición de derivación

Definición

Sea $G = (N, \Sigma, R, S)$ una gramática independiente del contexto. Si $u, v \in (N \cup \Sigma)^*$, y $A \to w$ es una regla de R, denotamos por $uAv \Rightarrow uwv$ a la sustitución de la variable A por w.
Decimos que hay una **derivación** de u a v y se denota por $u \Rightarrow^* v$, si existen $u_i \in (N \cup \Sigma)^*$, $1 \leq i \leq k$, tal que

$$u \Rightarrow u_1 \Rightarrow u_2 \cdots \Rightarrow u_k \Rightarrow v$$

Lenguaje generado por gramática

Definición

Sea $G = (N, \Sigma, R, S)$ una gramática independiente del contexto. El lenguaje generado por un gramática G es el conjunto de cadenas de Σ^* que se derivan de la variable inicial S

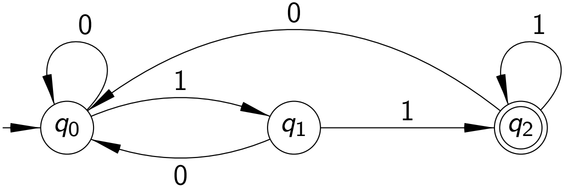
$$L(G) = \{w \in \Sigma^* \mid S \Rightarrow^* w\}$$

Diremos que $L(G)$ es un **lenguaje independiente del contexto o incontextual (LIC)**

Lenguajes regulares y LICs

- Hemos visto que $L = \{a^n b^n \mid n \in \mathbb{N}\}$ es un lenguaje independiente del contexto.
- También vimos que L no es regular.
- Por tanto, no todos los LICs son lenguajes regulares.
- ¿Los lenguajes regulares son LICs?
- Si L es un lenguaje regular, ¿existe una gramática independiente del contexto que genera L?

Construir gramática de un AFD o AFnD



- Utilizamos una variable por cada estado: Q_0, Q_1, Q_2
- La variable inicial S es la variable correspondiente a el estado inicial: $S = Q_0$
- Una regla por cada transición:

$$
\begin{aligned}
Q_0 &\rightarrow 0Q_0|1Q_1 \\
Q_1 &\rightarrow 0Q_0|1Q_2 \\
Q_2 &\rightarrow 0Q_0|1Q_2|\epsilon
\end{aligned}
$$

- En los estados finales se añade la regla $Q_f \rightarrow \epsilon$

- Observar que todas las reglas que aparecen al dar la gramática que genera el mismo lenguaje que un AFD (AFnD) son de la forma

$$NO\_TERMINAL \rightarrow TERMINAL\ NO\_TERMINAL$$
$$NO\_TERMINAL \rightarrow \epsilon$$

- Todas las reglas $r_i \in N \times (\Sigma N \cup \{\epsilon\})$
- Todas los lenguajes regulares son generados por gramáticas de este tipo

Definición gramática regular

Definición

Una gramática $G = (N, \Sigma, R, S)$ se dice **regular** si

$$R \subset N \times (\Sigma N \cup \{\epsilon\})$$

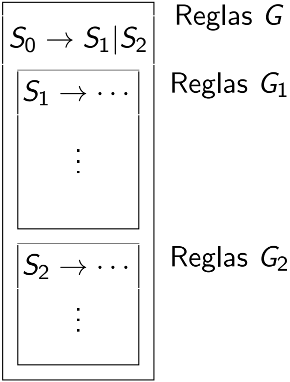
Los lenguajes generados por gramáticas regulares son regulares. Además, todo lenguaje regular es generado por una gramática regular.

Propiedades de clausura

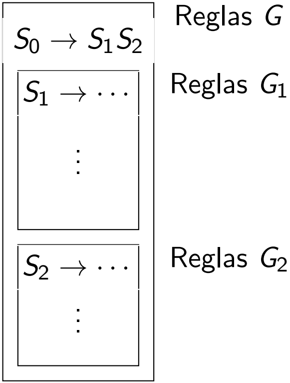
- Recordad que la unión, concatenación, estrella de Kleene, complemento, intersección, reverso de lenguajes regulares es regular.
- ¿Qué propiedades de clausura se cumplen para LICs?
- Tenemos que comprobar que existe una gramática independiente de contexto que genere la unión (concatenación, etc) de dos LICs.

Unión de LIC

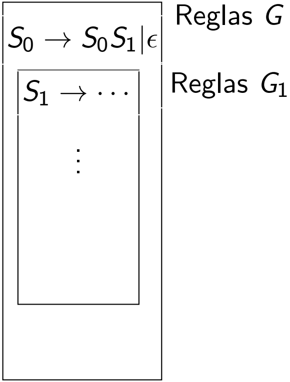
- Dados dos LICs generados por dos gramáticas $G_1 = \{N_1, \Sigma, R_1, S_1)$ y $G_2 = \{N_2, \Sigma, R_2, S_2)$
- La gramática que genera la unión $G = \{N_1 \cup N_2, \Sigma, R, S_0)$: tiene una variable inicial S_0, y las reglas de G son la unión de las reglas de G_1 y G_2 junto con la regla $S_0 \rightarrow S_1|S_2$



- Dados dos LICs generados por dos gramáticas $G_1 = \{N_1, \Sigma, R_1, S_1)$ y $G_2 = \{N_2, \Sigma, R_2, S_2)$
- La gramática que genera la concatenación $G = \{N_1 \cup N_2, \Sigma, R, S_0)$: tiene una variable inicial S_0, y las reglas de G son la unión de las reglas de G_1 y G_2 junto con la regla $S_0 \rightarrow S_1 S_2$



- Dado un LICs generado por la gramática $G_1 = \{N_1, \Sigma, R_1, S_1)$
- La gramática que genera la estrella de Kleene $G = \{N_1, \Sigma, R, S_0)$: tiene una variable inicial S_0, y las reglas de G son las reglas de G_1 junto con la regla $S_0 \rightarrow S_0 S_1 | \epsilon$



Reverso de LIC

- Dado un LICs generado por la gramática $G_1 = \{N_1, \Sigma, R_1, S_1)$
- La gramática que genera el reverso $G = \{N_1, \Sigma, R, S_1)$: si $A \rightarrow w$ es una regla de G_1, entonces $A \rightarrow w^R$ es una regla de G

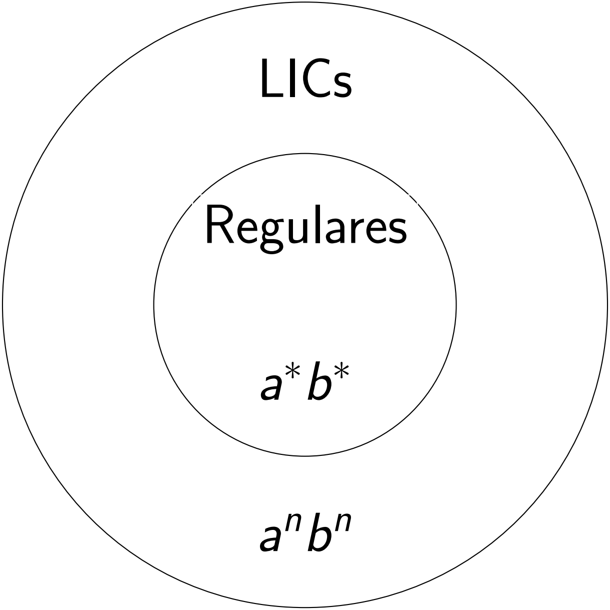
Otras propiedades de clausura

- No es cierto que el complemento de un LIC sea LIC
- Tampoco es cierto que la intersección de LICs sea LIC.
- En la siguiente lección veremos ejemplos.
- Sobre la intersección, tenemos el siguiente resultado positivo

Lema

Si L_1 es un LIC y L_2 es regular, entonces $L_1 \cap L_2$ es LIC.

Clasificación de los lenguajes



Referencias

- Sipser: Capítulo 2, sección 2.1 (págs. 100-105)