Matemáticas Discretas

Lenguajes y Gramáticas

Lenguajes y Gramáticas

Tenemos dos clases de lenguaje:

- Lenguaje Natural
- Lenguaje Formal

Cont…

- De acuerdo al diccionario Webster, un lenguaje es un cuerpo de palabras y métodos de combinación de palabras utilizado, y comprendido por una comunidad de tamaño considerable, tales lenguajes se llaman Lenguajes Naturales.
- Los Lenguajes Formales se utilizan para modelar los lenguajes naturales y para comunicarse con las computadoras.

Lenguaje Formal

- **Definición:**
  - Sea A un conjunto finito. Un lenguaje (formal) L sobre A es un subconjunto de A*, el conjunto de todas las cadenas sobre A.
- **Por Ejemplo**
  - Sea A = {a , b}.
    a) El conjunto **L** de todas las cadenas sobre **A** que contiene un número impar de letras a es un lenguaje sobre **A**.
    b) **L** es precisamente el conjunto de cadenas sobre **A** aceptadas por el autómata de estado finito.
    c) Una forma de definir un lenguaje es dar una lista de **reglas** (gramática) que éste debe cumplir.

Escribimos G=(N, T, P, σ)

Gramática Simple

- **Definición:**
  - Una gramática con estructura de frases (o simplemente una gramática) G consta de:
    - a) Un conjunto finito **N** de símbolos no terminales.
    - b) Un conjunto finito **T** de símbolos terminales, donde $N \cap T = \emptyset$.
    - c) Un subconjunto finito **P** de $[(N \cup T)^* - T^*] \times (N \cup T)^*$, llamado el conjunto de producciones.
    - d) Un símbolo inicial $\sigma \in N$.
  - Generalmente, una producción $(A,B) \in P$ se escribe:
    $A \rightarrow B$
  - A debe incluir al menos un símbolo no Terminal, mientras que B puede constar de cualquier combinación de símbolos terminales y no terminales.

Cont…

- **Ejemplo 1:**
  - Sean:
    - $N = \{\sigma, S\}$
    - $T = \{a, b\}$
    - $P = \{\sigma \rightarrow b\,\sigma,\ \sigma \rightarrow aS, S \rightarrow bS,\ S \rightarrow b\}$
  - Entonces $G = (N, T, P, \sigma)$ es una gramática.
  - Podemos construir un Lenguaje L(G) a partir de una gramática G.
    - Se usan las producciones para deducir las cadenas que forman el leguaje.
  - A partir del símbolo inicial se utilizan (remplazan) varias veces las producciones hasta obtener una cadena de símbolos terminales (T).
  - **El lenguaje L(G) es el conjunto de todas las cadenas que se obtienen.**

Derivación

- **Definición:**
  - Sea $G=(N,T,P,\sigma)$ una gramática.
    - Si $\sigma \rightarrow \beta$ es una produccion y $x\alpha y \in (N \cup T)^*$, decimos que $x\,\beta\,y$ se **deriva directamente** de $x\,\alpha\,y$ y escribimos

$$x\,\alpha\,y \Rightarrow x\,\beta\,y$$

- Si $\alpha_i \in (N \cup T)^*$ para i=1,.....,n y α_i+1 se deriva directamente de α_i para i=1,...,n-1, decimos que α_n se deriva de α_i y escribimos

$$\alpha_i \Rightarrow \alpha_n$$

- Decimos que α_n es la derivación de α_i

$$\alpha_1 \Rightarrow \alpha_2 \Rightarrow \ldots \Rightarrow \alpha_n$$

El lenguaje generado por G, denotado L(G), consta de todas las cadenas sobre T derivables de σ.

Cont…

- **Ejemplo 2:**
  - Sea G la gramática
    - $N=\{\sigma,S\}$
    - $T=\{a,b\}$
    - $P=\{\sigma \rightarrow b\,\sigma,\ \sigma \rightarrow aS, S \rightarrow bS,\ S \rightarrow b\}$
    - $G=\{N,T,P,\sigma\}$
- La cadena *abSbb* se deriva directamente de *aSbb,* lo cual se escribe

$$aSbb \Rightarrow abSbb$$

- La cadena bbab se deriva de σ:

$$\sigma \Rightarrow b\sigma \Rightarrow bb\sigma \Rightarrow bbaS \Rightarrow bbab$$

Cont…

- **Cont… Ejemplo 2:**
  - Las únicas derivaciones de σ son:

$$
\begin{aligned}
&\sigma \Rightarrow b\sigma \\
&\ldots \\
&\Rightarrow b^n\sigma \qquad n\geq 0 \\
&\Rightarrow b^n aS \\
&\ldots \\
&\Rightarrow b^n ab^{m-1}S \\
&\Rightarrow b^n ab^m S \qquad n\geq 0,\ m\geq 1
\end{aligned}
$$

  - Así, *L(G)* consta de las cadenas sobre *{a,b}* que contienen precisamente una *a* y terminan con *b.*

Gramática BNF
(Forma normal de Backus)

- Es una forma alternativa de establecer las producciones de una gramática.
  - Los símbolos no terminales comienzan con “<” y terminan con”>”.
  - La producción $S \rightarrow T$ se escribe $S ::= T$
  - Se puede tener $S ::= T_1 | T_2 | T_3 | T_4 \ldots\ldots | T_n$
  - La barra | se lee “o”

Cont…

- **Ejemplo 3:**
  - Un entero se define como una cadena que consta de un signo opcional (+ o -) seguido de una cadena de dígitos (0 a 9). La siguiente gramática genera a todos los números enteros.

<dígito>	::=0|1|2|3|4|5|6|7|8|9
<entero>	::=<entero con signo>|<entero sin signo>
<entero con signo>	::=+<entero sin signo>|-<entero sin signo>
<entero sin signo>	::=<dígito>|<dígito><entero sin signo>

  - El símbolo inicial es <entero>
  - La derivación del entero -901
    - <entero> ⇒<entero con signo>
    - ⇒ -<entero sin signo>
    - ⇒ -<dígito><entero sin signo>
    - ⇒ -<dígito><dígito><entero sin signo>
    - ⇒ -<dígito><dígito><dígito>
    - ⇒ *-9*<dígito><dígito>
    - ⇒ *-90*<dígito>
    - ⇒ *-901.*

Cont…

- **Ejemplo 4:**
  - Escriba una gramática BNF para el conjunto de números con decimales (por ej: 487.34 o 8.3)
    - <digito>::=0|1|2|3|4|5|6|7|8|9
    - <numero>::=<digito>|<digito><numero>
    - <numero decimal>::=<numero>.<numero>
  - Para ésta gramática:
    - T = {0,1,2,3,4,5,6,7,8,9,.}
    - N = {digito, numero, numero decimal}

Cont…

- **Ejemplo 5:**
  - Escriba una gramática BNF para el conjunto de números enteros con signo.
    - <digito>::=0|1|2|3|4|5|6|7|8|9
    - <numero>::=<digito><numero> | <digito>
    - <signo>::=+ | -
    - <numero con signo>::=<signo><numero>
  - Para esta gramática:
    - T={0,1,2,3,4,5,6,7,8,9,+,-}
    - N={digito, numero, signo, numero con signo}

Gramáticas y Lenguajes

Tipos de Gramáticas

- Las gramáticas se clasifican por los tipos de producciones que las definen.
- **Definición:**
  - Sea G una gramática y λ la cadena nula.

  a) Si cada producción es de la forma
    - $\alpha A\beta \to \alpha\delta\beta$, donde $\alpha,\beta \in (N \cup T)^*$, $A \in N$, $\delta \in (N \cup T)^*$, - $\{\lambda\}$.
    - Entonces *G es una gramática* ***sensible al contexto (o de tipo 1).***

  b) Si cada producción es de la forma.
    - $A \to \delta$, donde $A \in N$, $\delta \in (N \cup T)^*$,
    - Entonces *G es una gramática* ***libre de contexto (o de tipo 2).***

  c) Si cada producción es de la forma.
    - $A \to \alpha$ o $A \to \alpha B$ o $A \to \lambda$ donde $A,B \in N$, $\alpha \in T$,
    - Entonces *G es una gramática* ***regular (o de tipo 3).***

Gramáticas y Lenguajes

Tipos de Gramáticas

- En una gramática **sensible al contexto**
  - Se puede remplazar *A* por *δ si A está en el contexto de α y β*
- En una gramática **libre de contexto**
  - Establece que podemos remplazar A por δ en cualquier momento.
- En una gramática **regular**
  - Se tienen reglas de sustitución particularmente sencillas:
    - Remplazamos un símbolo no terminal por
      - un símbolo terminal
      - un símbolo terminal seguido de un símbolo no terminal
      - la cadena nula

Observe que una gramática regular es una gramática libre de contexto y una gramática libre de contexto sin producciones de la forma $A \to \lambda$ es una gramática sensible al contexto

Gramáticas y Lenguajes

Cont…

- **Gramática sensible al contexto**
  - **Ejemplo 6:**
  - La gramática G definida por **T** = {a, b, c}, **N** = {σ, A, B, C, D, E}, símbolo inicial σ y con producciones

$$\sigma \to aAB,\ \sigma \to aB,\ A \to aAC,\ A \to aC,\ B \to Dc,$$
$$D \to b,\ CD \to CE,\ CE \to DE,\ DE \to DC,\ Cc \to Dcc,$$

    - En la producción CE → DE podemos reemplazar C con D si C va seguida de E y
    - En la producción Cc → Dcc podemos reemplazar C con Dc si C va seguida de c.
    - Podemos derivar DC de C, pues $CD \Rightarrow CE \Rightarrow DE \Rightarrow DC$
    - La cadena $a^3b^3c^3$ esta en **L(G)**, pues tenemos

$$\sigma \Rightarrow aAB \Rightarrow aaACB \Rightarrow aaaCCDc \Rightarrow aaaDCCc \Rightarrow aaaDCDcc$$
$$\Rightarrow aaaDDCcc \Rightarrow aaaDDDccc \Rightarrow aaabbbccc$$

    - *Se puede mostrar que* **L(G)= { $a^nb^nc^n$ | n =1,2,…}**

Cont…

- **Lenguaje sensible al contexto**
  - **Definición:**
    - Un lenguaje **L** es sensible al contexto (respectivamente Libre de contexto, regular) si existe una gramática G sensible al contexto tal que L = L(G).
  - **Continuación ejemplo 6:**
    - De acuerdo con el ejemplo anterior, el lenguaje

      $L = L(G) = \{ a^n b^n c^n \mid n = 1,2,\ldots\}$

    - Se puede mostrar que no existe una gramática libre de contexto G tal que L = L(G); por tanto, L no es un lenguaje libre de contexto.

Cont…

- **Gramática libre al contexto**
  - **Ejemplo 7:**
    - La gramática G definida por **T** = {a,b}, **N** = {σ}, símbolo inicial σ y con producciones

      $\sigma \to a\sigma b,\ \sigma \to ab,$

    - Las únicas derivaciones de σ son

      $\sigma \Rightarrow a\sigma b$

      ….

      $\Rightarrow a^{n-1}\, \sigma\, b^{n-1}$

      $\Rightarrow a^{n-1}\, ab\, b^{n-1} = a^n b^n$

    - Así, **L(G)** consta de las cadenas sobre {a,b} de la forma $a^n b^n$, n= 1,2,….
    - Este lenguaje es libre de contexto.

Cont…

- **Gramática regular**
  - **Ejemplo 8:**
    - Sean:
      - N = {σ, S}
      - T = {a, b}
      - P = {σ → b σ, σ → aS, S → bS, S → b}
    - Entonces G = (N, T, P ,σ) es una gramática.
    - **L(G)** = $\{b^n a b^m \mid n = 1,2,\ldots;\ m = 1,2,\ldots\}$
    - Este lenguaje, generado por la gramática regular, es regular.

Cont…

- **Definición:**
  - Las gramáticas G y G' son equivalentes si L(G) = L(G').
- **Ejemplo 9:**
  - Si definimos una relación R sobre un conjunto de gramáticas mediante la regla de *G R G'* si *G* y *G'* son equivalentes, R es una relación de equivalencia. Cada clase de equivalencia consta de un conjunto de gramáticas equivalentes entre sí.

Gramática de Lindenmayer

- **Definición 1:**
  - Una gramática de Lindenmayer interactiva libre de contexto consta de:
    - Un conjunto finito N de símbolos no terminales.
    - Un conjunto finito T de símbolos terminales, donde $N \cap T = \emptyset$
    - Un conjunto finito P de producciones $A \rightarrow B$, donde A Є *(N U T) y B Є (N U T)*.*
    - *Un símbolo inicial σ Є N.*
  - La diferencia entre una gramática de Lindermayer interactiva libre de contexto y una gramática libre de contexto es que la primera permite el uso de producciones de la forma $A \rightarrow B$ donde A es un símbolo terminal o no terminal.

Cont…

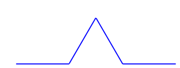
- **Definición 2:**
  - Sea G =(N, P, T,σ) una gramática de Lindenmayer interactiva libre de contexto. Si
    $\alpha = x_1 \ldots x_n$ Y existen producciones $x_i \rightarrow \beta_i$
  - *En P, para i = 1,…,n, escribimos*
    $\alpha \Rightarrow \beta_i,\ldots,\beta_n$
  - Decimos que *$\beta_i,\ldots,\beta_n$ es derivable de manera directa de α. Si α_{i+1} es derivable de manera directa de α_i para i = 1,…,n-1, decimos que α_n es derivable de α_1 y escribimos*
    $\alpha_1 \Rightarrow \alpha_n$
  - *Decimos que $\alpha_1 \Rightarrow \alpha_2 \Rightarrow \ldots \Rightarrow \alpha_n$*
    - *es la derivación de α_n* (a partir de α_1). El lenguaje generado por G, denotado por L(G), consta de todas las cadenas de sobre T derivables a partir de *α.*

Cont…

- **Ejemplo 10:**
  - Sean
    - N = {D}
    - T = {d,+,-}
    - P = {$D \rightarrow D - D + + D - D$, $D \rightarrow d$, $+ \rightarrow +$, $- \rightarrow -$}
  - Consideramos a G (N, T, P, D) como una gramática Lindenmayer interactiva libre de contexto.
  - Como un ejemplo de derivación D, tenemos
    $D \Rightarrow D - D + + D - D \Rightarrow d - d + + d - d$
  - Así, d – d + + d – d Є *L(G).*

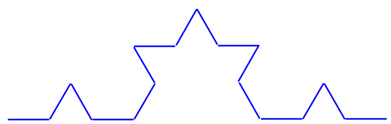
Cont…

- **Cont…**
  - Ahora daremos un significado a las cadenas de L(G).
  - Interpretamos el símbolo:
    - *d* como una instrucción para trazar una línea recta de una longitud fija en la dirección actual;
    - + como una instrucción para girar 60° hacia la derecha;
    - – como una instrucción para girar 60° hacia a la izqui erda.
      - Si comenzamos del lado izquierdo y el primer movimiento es horizontal y hacia la derecha, la interpretación de la cadena d-d++d-d produce la siguiente curva fractal



Cont…

- **Cont…**
  - La siguiente cadena más larga en L(G) es
    d - d + + d - d - d - d + + d - d + + d - d + + d - d - d - d + + d - d



Copos de nieve de von Kock

Gramáticas para Expresiones

- **Ejemplo 10:**
  - Si S es un conjunto de “palabras” entonces S* es la colección de todas las “oraciones” posibles formadas por las palabras de S.
  - Estas “oraciones” no necesariamente tienen un sentido o estructura aparente.
  - Un lenguaje es una especificación completa:
    1. Deben existir un conjunto S con todas las palabras que se consideran parte del lenguaje.
    2. Hay que designar un subconjunto de S* como el conjunto de las “oraciones con construcción adecuada” en el lenguaje.
    3. Hay que determinar cuales de las oraciones con construcción adecuada tienen significado y cual es este.

Cont…

- **Cont…**
  - La oración siguiente es una cadena en S*, pero no tiene una construcción adecuada (la disposición de los sustantivos y los verbos no es valida).

    “Iba a la tienda Juan Jorge a cantar”

  - La siguiente oración tiene una construcción adecuada pero carece de sentido.

    “Los sonidos azules se sientan y cruzan la pierna bajo la cima de la montaña”

  - En un lenguaje:
    - **Sintaxis:** regula la construcción adecuada de las oraciones.
    - **Semántica:** se encarga del significado de las oraciones.
  - En un lenguaje de programación lo que se enseña es la sintaxis.

Cont…

- Gramática para estructura de oraciones
  - G = (V, S , v_o, →), donde
    - V es un conjunto finito,
    - S un subconjunto de V,
    - v_0 Є V- S y
    - → es una relación finita en V*
  - **S** es el conjunto de todas las “palabras” permitidas en el lenguaje, y
  - **V** consta de S además de algunos otros símbolos.
  - El elemento **v_o** de V es un punto de partida para las sustituciones.
  - Por ultimo la relacio → sobre V* especifica los reemplazos permisibles.

Gramáticas y Lenguajes

Cont…

- **Ejemplo 11:**
  - Si G = (V, S , vo, →)
    - S es el conjunto de símbolos terminales y
    - N = V – S es el conjunto de símbolos no terminales
  - Sea
    - S = { Juan, Julia, maneja, corre, descuidadamente, rápido, frecuentemente }
    - N = {oración, sujeto, predicado, verbo, adverbio }
    - V = S U N .
    - v_o = oración y

Gramáticas y Lenguajes

Cont…

- **Cont…**
  - supóngase que la relación → en V* queda descrita enumerando todas las producciones como sigue
    - Oración → sujeto predicado
    - Sujeto → Juan
    - Sujeto → Julia
    - Predicado → verbo adverbio
    - Verbo → maneja
    - Verbo → corre
    - Adverbio → descuidadamente
    - Adverbio → rápido
    - Adverbio → frecuentemente

Gramáticas y Lenguajes

Cont…

- **Cont…**
  - El conjunto S contiene todas las palabras permitida en el lenguaje.
  - Se afirma que la oración **" Julia maneja frecuentemente "** que esta denotada por w, es una oracion permisible o con sintaxis correcta, deacuerdo con las reglas de este lenguaje.
  - Para demostrar esto, considerese la siguiente serie de cadenas en V*.
    - Oración
    - Sujeto predicado
    - Julia predicado
    - Julia verbo adverbio
    - Julia maneja adverbio
    - Julia maneja frecuentemente
  - Por definición w tiene sintaxis correcta, ya que en este ejemplo v_o es una oración.
  - En las gramáticas para estructura de oraciones, la búsqueda de una sintaxis correcta se refiere solo al proceso mediante el cual al formar una oración se procura que esta sea correcta gramaticalmente hablando, y nada mas.