Teoría de la Computación

Clase 11: Forma normal de Chomsky

Mauro Artigiani

30 agosto 2021

Universidad del Rosario, Bogotá

Gramáticas regulares y DFAs

Gramáticas regulares

Una gramática es regular si y solo si sus reglas son de la forma

$$A \to aB$$

$$A \to a$$

$$A \to \varepsilon$$

Gramáticas regulares

Una gramática es regular si y solo si sus reglas son de la forma

$$A \to aB$$

$$A \to a$$

$$A \to \varepsilon$$

Teorema

Las gramáticas regulares y los `DFA` son equivalentes.

Gramáticas regulares

Una gramática es regular si y solo si sus reglas son de la forma

$$A \to aB$$
$$A \to a$$
$$A \to \varepsilon$$

Teorema

Las gramáticas regulares y los `DFA` son equivalentes.

La clase pasada hemos visto una implicación. Veamos la otra.

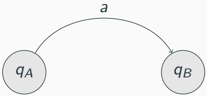
Gramáticas regulares

A partir de cada regla construimos un pedazo del autómata.

Gramáticas regulares

A partir de cada regla construimos un pedazo del autómata.

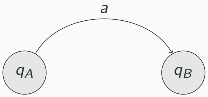
$$A \to aB$$



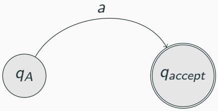
Gramáticas regulares

A partir de cada regla construimos un pedazo del autómata.

$$A \rightarrow aB$$



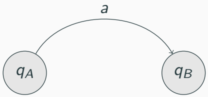
$$A \rightarrow a$$



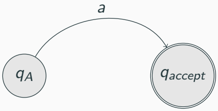
Gramáticas regulares

A partir de cada regla construimos un pedazo del autómata.

$$A \rightarrow aB$$



$$A \rightarrow a$$



$$A \rightarrow \varepsilon$$

Jerarquía de Chmosky

Jerarquía de Chomsky

El lingüista Noam Chomsky propuso en 1956 la siguiente jerarquía de lenguajes.

Jerarquía de Chomsky

El lingüista Noam Chomsky propuso en 1956 la siguiente jerarquía de lenguajes.

Tipo 3: Lenguajes regulares:

$$A \to \mathtt{a}B \mid \mathtt{a} \mid \varepsilon, \qquad \text{con } \mathtt{a} \in \Sigma,\ A, B \in V.$$

Jerarquía de Chomsky

El lingüista Noam Chomsky propuso en 1956 la siguiente jerarquía de lenguajes.

Tipo 2: Lenguajes independientes del contexto:

$$A \to x, \qquad \text{con } x \in (V \cup \Sigma)^*.$$

Tipo 3: Lenguajes regulares:

$$A \to \mathtt{a}B \,|\, \mathtt{a} \,|\, \varepsilon, \qquad \text{con } \mathtt{a} \in \Sigma,\ A, B \in V.$$

Jerarquía de Chomsky

El lingüista Noam Chomsky propuso en 1956 la siguiente jerarquía de lenguajes.

Tipo 1: Lenguajes sensibles al contexto:

$$xAy \to xuy, \qquad \text{con } x, u, y \in (V \cup \Sigma)^*.$$

Tipo 2: Lenguajes independientes del contexto:

$$A \to x, \qquad \text{con } x \in (V \cup \Sigma)^*.$$

Tipo 3: Lenguajes regulares:

$$A \to \mathtt{a}B \mid \mathtt{a} \mid \varepsilon, \qquad \text{con } \mathtt{a} \in \Sigma,\ A, B \in V.$$

Jerarquía de Chomsky

El lingüista Noam Chomsky propuso en 1956 la siguiente jerarquía de lenguajes.

Tipo 0: Lenguajes recursivamente enumerables (o con estructura de frase):

$$x \to y, \qquad \text{con } x, y \in (V \cup \Sigma)^*.$$

Tipo 1: Lenguajes sensibles al contexto:

$$xAy \to xuy, \qquad \text{con } x, u, y \in (V \cup \Sigma)^*.$$

Tipo 2: Lenguajes independientes del contexto:

$$A \to x, \qquad \text{con } x \in (V \cup \Sigma)^*.$$

Tipo 3: Lenguajes regulares:

$$A \to \mathtt{a}B \mid \mathtt{a} \mid \varepsilon, \qquad \text{con } \mathtt{a} \in \Sigma,\ A, B \in V.$$

Ambigüedad

Ambigüedad

A veces, dada una `CFG`, hay varias maneras distintas de producir la misma palabra. Dado que a cada derivación (o árbol de parsing) corresponde una interpretación de la palabra, esto puede ser un problema si estamos modelizando un lenguaje de programación.

Ambigüedad

A veces, dada una `CFG`, hay varias maneras distintas de producir la misma palabra. Dado que a cada derivación (o árbol de parsing) corresponde una interpretación de la palabra, esto puede ser un problema si estamos modelizando un lenguaje de programación. Esto pasa también con los lenguajes naturales, donde, a veces, la misma frase puede ser interpretada en más de una manera.

Ambigüedad

A veces, dada una CFG, hay varias maneras distintas de producir la misma palabra. Dado que a cada derivación (o árbol de parsing) corresponde una interpretación de la palabra, esto puede ser un problema si estamos modelizando un lenguaje de programación. Esto pasa también con los lenguajes naturales, donde, a veces, la misma frase puede ser interpretada en más de una manera.

Diremos que una palabra es derivada en manera ambigua si hay más de una derivación que la produce. De manera similar, si una gramática genera una palabra de manera ambigua diremos que la gramática misma es ambigua.

Un ejemplo de ambigüedad

Consideramos la gramática

$$\langle \texttt{expr} \rangle \to \langle \texttt{expr} \rangle + \langle \texttt{expr} \rangle | \langle \texttt{expr} \rangle \times \langle \texttt{expr} \rangle | \texttt{(} \langle \texttt{expr} \rangle \texttt{)} | \texttt{a}.$$

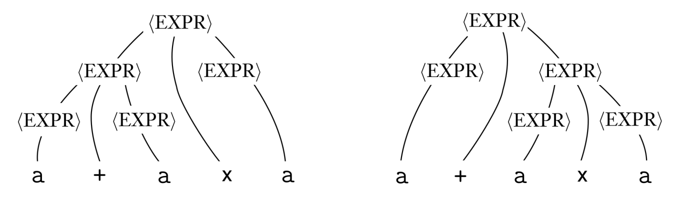
La gramática genera la expresión $\texttt{a} + \texttt{a} \times \texttt{a}$ en manera ambigua; es decir: no respecta el orden de las operaciones.

Un ejemplo de ambigüedad

Consideramos la gramática

$$\langle \texttt{expr} \rangle \to \langle \texttt{expr} \rangle + \langle \texttt{expr} \rangle | \langle \texttt{expr} \rangle \times \langle \texttt{expr} \rangle | \texttt{(}\langle \texttt{expr} \rangle\texttt{)} | \texttt{a}.$$

La gramática genera la expresión $\mathtt{a} + \mathtt{a} \times \mathtt{a}$ en manera ambigua; es decir: no respecta el orden de las operaciones.

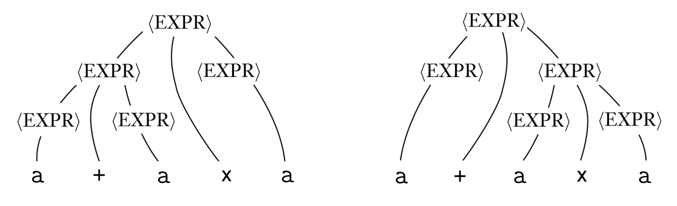


Un ejemplo de ambigüedad

Consideramos la gramática

$$\langle \texttt{expr} \rangle \to \langle \texttt{expr} \rangle + \langle \texttt{expr} \rangle | \langle \texttt{expr} \rangle \times \langle \texttt{expr} \rangle | \texttt{(}\langle \texttt{expr} \rangle\texttt{)} | \texttt{a}.$$

La gramática genera la expresión $\mathsf{a} + \mathsf{a} \times \mathsf{a}$ en manera ambigua; es decir: no respecta el orden de las operaciones.



La clase pasada hemos visto otra gramática que genera el mismo lenguaje, pero en manera no ambigua.

Definición de ambigüedad

Hemos visto una gramática que genera un pedazo del inglés. Esa gramática, como el lenguaje inglés, es ambigua. Por ejemplo, hay dos maneras de generar la frase `the girl touches the boy with the flower`, lo que corresponde a los dos significados de la frase misma.

Definición de ambigüedad

Hemos visto una gramática que genera un pedazo del inglés. Esa gramática, como el lenguaje inglés, es ambigua. Por ejemplo, hay dos maneras de generar la frase `the girl touches the boy with the flower`, lo que corresponde a los dos significados de la frase misma.

A veces dos derivaciones difieren solo por el orden en que hacemos las varias substituciones, pero tienen el mismo árbol de parsing. En este caso, la palabra tiene un único significado.

Definición de ambigüedad

Hemos visto una gramática que genera un pedazo del inglés. Esa gramática, como el lenguaje inglés, es ambigua. Por ejemplo, hay dos maneras de generar la frase `the girl touches the boy with the flower`, lo que corresponde a los dos significados de la frase misma.

A veces dos derivaciones difieren solo por el orden en que hacemos las varias substituciones, pero tienen el mismo árbol de parsing. En este caso, la palabra tiene un único significado. Por eso decidimos un orden en las derivaciones: siempre substituimos la variable más a la izquierda en la palabra. Llamamos este orden *derivación a la izquierda*.

Definición de ambigüedad

Definición

Decimos que una palabra w es generada en manera ambigua en una gramática independiente del contexto G si hay por lo menos dos derivaciones a la izquierda de w. Una gramática G es ambigua si genera una palabra en manera ambigua.

Definición de ambigüedad

Definición

Decimos que una palabra w es generada en manera ambigua en una gramática independiente del contexto G si hay por lo menos dos derivaciones a la izquierda de w. Una gramática G es ambigua si genera una palabra en manera ambigua.

A veces, la ambigüedad deriva de la gramática que hemos elegido, como en el ejemplo de la aritmética. A veces, la ambigüedad es intrínseca en el lenguaje mismo y no en la gramática que lo describe.

Forma normal de Chomsky

Forma normal de Chomsky

Para hacer manipulaciones formales, es bueno tener una forma estándar para una gramática independiente del contexto.

Forma normal de Chomsky

Para hacer manipulaciones formales, es bueno tener una forma estándar para una gramática independiente del contexto.

Definición

Una `CFG` es en forma normal de Chomsky si todas las reglas son de la forma

$$A \to BC$$
$$A \to a,$$

donde a es un terminal y A, B, y C son variables. Además pedimos que ni B ni C sean la variable inicial y permitimos la regla $S \to \varepsilon$, si S es la variable inicial.

Forma normal de Chomsky

Para hacer manipulaciones formales, es bueno tener una forma estándar para una gramática independiente del contexto.

Definición

Una `CFG` es en forma normal de Chomsky si todas las reglas son de la forma

$$A \to BC$$
$$A \to a,$$

donde a es un terminal y A, B, y C son variables. Además pedimos que ni B ni C sean la variable inicial y permitimos la regla $S \to \varepsilon$, si S es la variable inicial.

Todas la gramáticas se pueden escribir en forma normal de Chomsky.

Ejemplo

Antes de hacer la demostración formal, hagamos un ejemplo.

Ejemplo

Antes de hacer la demostración formal, hagamos un ejemplo.

Consideramos la gramática:

$$S \to ASA \mid \mathtt{a}B$$
$$A \to B \mid S$$
$$B \to \mathtt{b} \mid \varepsilon$$

Ejemplo

Antes de hacer la demostración formal, hagamos un ejemplo. Consideramos la gramática:

$$S \to ASA \mid \mathtt{a}B$$
$$A \to B \mid S$$
$$B \to \mathtt{b} \mid \varepsilon$$

1. Para asegurar que la variables inicial no esté en la derecha de ninguna regla, añadimos una nueva variable inical S_0:

$$S_0 \to S$$
$$S \to ASA \mid \mathtt{a}B$$
$$A \to B \mid S$$
$$B \to \mathtt{b} \mid \varepsilon$$

Ejemplo

2. Ahora quitamos todas las reglas $A \to \varepsilon$, ajustando las demás reglas para no cambiar el lenguaje.

$$
\begin{aligned}
S_0 &\to S \\
S &\to ASA \mid \mathtt{a}B \mid \mathtt{a} \\
A &\to B \mid S \mid \varepsilon \\
B &\to \mathtt{b} \mid \varepsilon
\end{aligned}
$$

$$
\begin{aligned}
S_0 &\to S \\
S &\to ASA \mid \mathtt{a}B \mid \mathtt{a} \mid SA \mid AS \mid S \\
A &\to B \mid S \mid \varepsilon \\
B &\to \mathtt{b}
\end{aligned}
$$

Ejemplo

2. Ahora quitamos todas las reglas $A \to \varepsilon$, ajustando las demás reglas para no cambiar el lenguaje.

$$\begin{aligned} S_0 &\to S \\ S &\to ASA \mid \mathtt{a}B \mid \mathtt{a} \\ A &\to B \mid S \mid \varepsilon \\ B &\to \mathtt{b} \mid \varepsilon \end{aligned} \qquad \begin{aligned} S_0 &\to S \\ S &\to ASA \mid \mathtt{a}B \mid \mathtt{a} \mid SA \mid AS \mid S \\ A &\to B \mid S \mid \varepsilon \\ B &\to \mathtt{b} \end{aligned}$$

3a. Quitamos las reglas $S \to S$ y $S_0 \to S$

$$\begin{aligned} S_0 &\to S \\ S &\to ASA \mid \mathtt{a}B \mid \mathtt{a} \mid SA \mid AS \mid S \\ A &\to B \mid S \\ B &\to \mathtt{b} \end{aligned} \qquad \begin{aligned} S_0 &\to S \mid ASA \mid \mathtt{a}B \mid \mathtt{a} \mid SA \mid AS \\ S &\to ASA \mid \mathtt{a}B \mid \mathtt{a} \mid SA \mid AS \\ A &\to B \mid S \\ B &\to \mathtt{b} \end{aligned}$$

Ejemplo

3b. Quitamos las reglas $A \to B$ y $A \to S$

$$\begin{aligned} S_0 &\to ASA \mid \mathtt{a}B \mid \mathtt{a} \mid SA \mid AS \\ S &\to ASA \mid \mathtt{a}B \mid \mathtt{a} \mid SA \mid AS \\ A &\to B \mid S \mid \mathtt{b} \\ B &\to \mathtt{b} \end{aligned}$$

$$\begin{aligned} S_0 &\to ASA \mid \mathtt{a}B \mid \mathtt{a} \mid SA \mid AS \\ S &\to ASA \mid \mathtt{a}B \mid \mathtt{a} \mid SA \mid AS \\ A &\to S \mid \mathtt{b} \mid ASA \mid \mathtt{a}B \mid \mathtt{a} \mid SA \mid AS \\ B &\to \mathtt{b} \end{aligned}$$

Ejemplo

3b. Quitamos las reglas $A \to B$ y $A \to S$

$$\begin{aligned} S_0 &\to ASA \mid \mathtt{a}B \mid \mathtt{a} \mid SA \mid AS \\ S &\to ASA \mid \mathtt{a}B \mid \mathtt{a} \mid SA \mid AS \\ A &\to B \mid S \mid \mathtt{b} \\ B &\to \mathtt{b} \end{aligned}$$

$$\begin{aligned} S_0 &\to ASA \mid \mathtt{a}B \mid \mathtt{a} \mid SA \mid AS \\ S &\to ASA \mid \mathtt{a}B \mid \mathtt{a} \mid SA \mid AS \\ A &\to S \mid \mathtt{b} \mid ASA \mid \mathtt{a}B \mid \mathtt{a} \mid SA \mid AS \\ B &\to \mathtt{b} \end{aligned}$$

4. Añadimos variables y reglas para completar el trabajo.

$$\begin{aligned} S_0 &\to AA_1 \mid UB \mid \mathtt{a} \mid SA \mid AS \\ S &\to AA_1 \mid UB \mid \mathtt{a} \mid SA \mid AS \\ A &\to \mathtt{b} \mid AA_1 \mid UB \mid \mathtt{a} \mid SA \mid AS \\ A_1 &\to SA \\ U &\to \mathtt{a} \\ B &\to \mathtt{b} \end{aligned}$$

Formal normal de Chomsky

El procedimiento del ejemplo nos permite demostrar el siguiente resultado.

Teorema

Cualquier lenguaje independiente del contexto se puede generar a través de una gramática independiente del contexto en formal normal de Chomsky.

Demostración

Vamos a describir en manera algorítmica lo que hicimos antes.

1. Añadimos una nueva variable inicial S_0 con la regla $S_0 \to S$. Así la variable inicial no comparece en la derecha de ninguna regla.

Demostración

Vamos a describir en manera algorítmica lo que hicimos antes.

1. Añadimos una nueva variable inicial S_0 con la regla $S_0 \to S$. Así la variable inicial no comparece en la derecha de ninguna regla.
2. Removemos todas las reglas $A \to \varepsilon$. Para no cambiar el lenguaje, añadimos una nueva regla para cada regla que removimos: si hay una regla $R \to uAv$, con u y v terminales, añadimos la regla $R \to uv$. Lo mismo hacemos para cualquier ocurrencia de A. Si había la regla $R \to A$, añadimos la regla $R \to \varepsilon$ si ya no habíamos eliminado la misma. Repetimos hasta eliminar todas las reglas $A \to \varepsilon$, con $A \neq S_0$.

Demostración

3. Eliminamos todas las reglas unitaria de la forma $A \to B$. Para no cambiar el lenguaje, cada vez que había una regla $B \to u$, donde u es una cadena de terminales y variables, añadimos la regla $A \to u$. Repetimos hasta eliminar todas las reglas unitarias.

Demostración

3. Eliminamos todas las reglas unitaria de la forma $A \to B$. Para no cambiar el lenguaje, cada vez que había una regla $B \to u$, donde u es una cadena de terminales y variables, añadimos la regla $A \to u$. Repetimos hasta eliminar todas las reglas unitarias.
4. Por último, convertimos las reglas que quedan en forma estándar. Remplazamos cada regla de la forma $A \to u_1 u_2 \cdots u_k$, donde cada u_i es una variable o un terminal, con las reglas:

$$A \to u_1 A_1, \quad A_1 \to u_2 A_2, \ldots, A_{k-2} \to u_{k-1} u_k.$$

Los A_i son variables nuevas. También remplazamos todos los terminales u_i con nuevas variables U_i y añadimos las reglas $U_i \to u_i$. $\square$

Resumen

Resumen

Hoy aprendimos:

- Reconocer las gramáticas que hacen parte de la Jerarquía de Chomsky;
- Qué es gramática independiente del contexto ambigua;
- Qué es la forma normal de Chomsky de una `CFG`;
- Cómo dar una gramática en forma normal de Chomsky.