

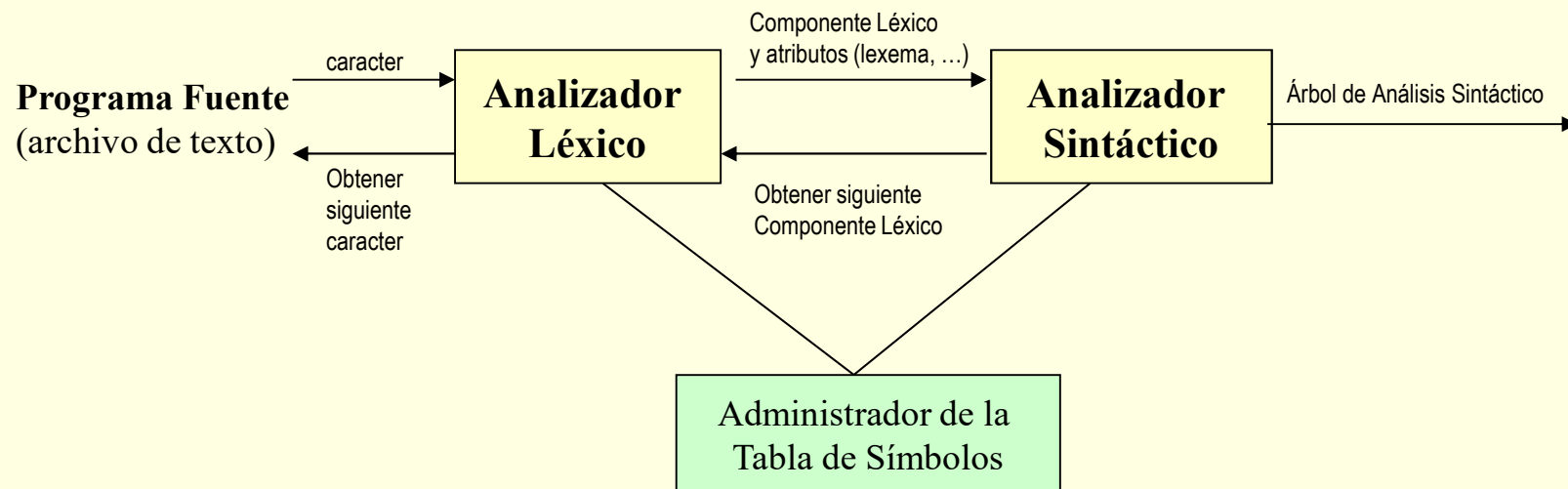
Análisis Léxico, Lenguajes Formales y Autómatas Finitos

Paradigmas y Lenguajes

Análisis léxico - Introducción

- Primera fase del proceso de compilación
- Analizador léxico = Scanner
- **Objetivos del Análisis Léxico**
 - Identificar Componentes Léxicos (y atributos)
 - Detectar Errores Léxicos

Comportamiento:



Análisis léxico - Introducción

■ Definiciones:

- **Componente léxico:** conjunto de cadenas que tienen un mismo significado desde el punto de vista sintáctico (es decir que pueden encontrarse en los mismos lugares desde el punto de vista de la sintaxis). Ejemplo: **identificador**
- **Lexema:** cada una de las cadenas pertenecientes a un componente léxico. Ejemplo: **tot1**
- **Patrón:** regla que define la estructura de los lexemas de un componente léxico (ley de conformación interna del conjunto). Ejemplo: una letra seguida de 0 o más letras o dígitos.

Introducción a los Lenguajes Formales

- **La Teoría de Lenguajes y Autómatas sirve como base para las tareas que deben realizarse dentro de las distintas fases de la compilación.**
- **Definiciones:**
 - **Símbolo:** ente abstracto sin definición formal
 - **Alfabeto:** conjunto finito (y no vacío) de símbolos
 - **Cadena** (palabra/word): secuencia finita de 0 o más símbolos de un alfabeto
 - **Longitud de una cadena:** sea w una cadena, la longitud de w se denota $|w|$ y es la cantidad de símbolos que posee.
 - **Cadena vacía:** se denota ε (epsilon) y $|\varepsilon|=0$

Introducción a los Lenguajes Formales

■ Operaciones sobre cadenas

- **Concatenación:** Sean w y x dos cadenas, su concatenación se escribe wx y es la cadena compuesta por la secuencia de símbolos de w seguidos por la secuencia de símbolos de x .
- **Prefijo de una cadena w :** es cualquier cadena s , donde $w = sx$ $|x| \geq 0$
- **Sufijo de una cadena w :** es cualquier cadena s , donde $w = xs$ $|x| \geq 0$
- **Subcadena de w :** es cualquier cadena s , donde $w = xsy$ $|x| \geq 0, |y| \geq 0$
- **Prefijo propio, sufijo propio y subcadena propia:** es un prefijo, sufijo, subcadena distinta a la original
- **Potencia:** sea w una cadena e $i \geq 0$ un número natural, se escribe w^i y es:
$$w^0 = \varepsilon$$
$$w^i = w.w^{i-1} \quad i > 0$$

Introducción a los Lenguajes Formales

- **Lenguaje (formal):** conjunto de cadenas formadas a partir de un alfabeto
- **Operaciones sobre lenguajes**
 - **Unión:** sean L y M dos lenguajes, la unión de L con M se escribe $L \cup M$ y es:
$$L \cup M = \{w \mid (w \text{ en } L) \vee (w \text{ en } M)\}$$
 - **Concatenación:** sean L y M dos lenguajes, su concatenación se escribe LM y es:
$$L.M = \{wx \mid (w \text{ en } L) \wedge (x \text{ en } M)\}$$
 - **Potencia:** sea L un lenguaje e $i \geq 0$ un natural, la potencia se escribe L^i y es:
$$L^0 = \{\varepsilon\}$$
$$L^i = L.L^{i-1} \quad i > 0$$
 - **Clausura de Kleene o clausura transitiva:** sea L un lenguaje, su clausura se escribe L^* y es:
$$L^* = \bigcup_{i=0.. \infty} L^i$$
 - **Clausura positiva:** sea L un lenguaje, su clausura se escribe L^+ y es:
$$L^+ = \bigcup_{i=1.. \infty} L^i$$

Lenguajes Formales y Autómatas Finitos

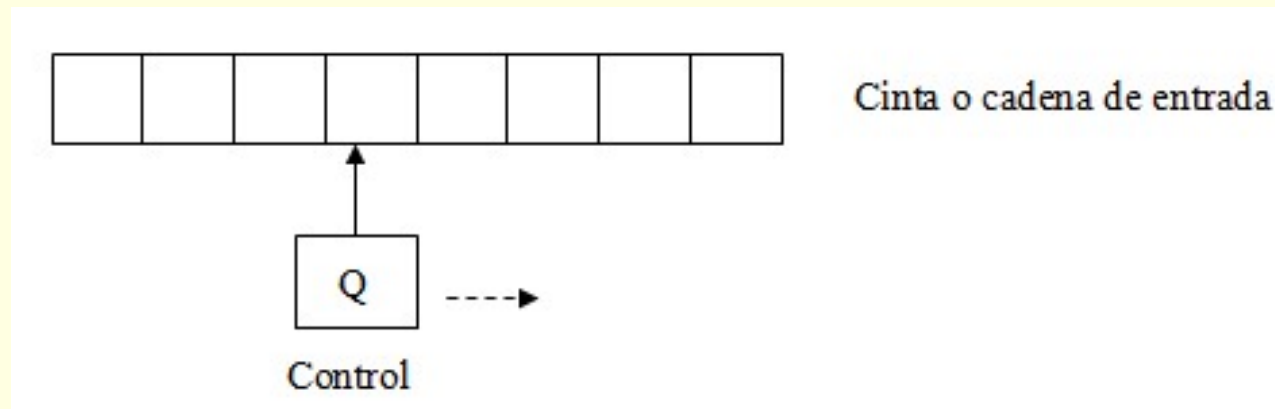
- **Sistema de estados finitos:** es un sistema que tiene las siguientes características:
 - Posee entradas y salidas discretas
 - Posee una cantidad finita de configuraciones llamadas “estados”
 - El sistema en cada momento se encuentra en un estado determinado
 - Cada estado resume la información de los anteriores, pero el sistema no recuerda (registra) los estados en que estuvo

- **Autómata finito:** modela un sistema de estados finitos
 - Es un modelo matemático definido por una 5-upla $(Q, \Sigma, \delta, q_0, F)$, donde:
 - Q : conjunto finito de estados
 - Σ : alfabeto de entrada
 - $q_0 \in Q$: estado inicial (nota: “en” se utiliza como “pertenece”)
 - $F \subseteq Q$: conjunto de estados “Finales” o de “Aceptación”
 - δ : Función de Transición
 - $\delta: Q \times \Sigma \rightarrow Q$

Lenguajes Formales y Autómatas Finitos

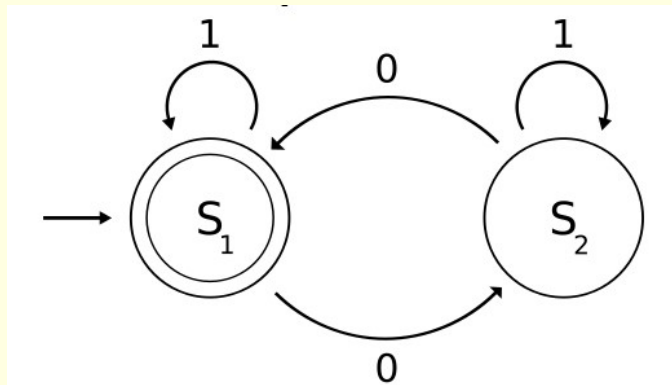
Comportamiento de un Autómata finito:

1. Al inicio del reconocimiento el autómata se encuentra en su estado inicial y el control está apuntando a la primer posición de la cadena.
2. Luego se utiliza la función de transición para calcular el nuevo estado (en base al estado actual y al símbolo al que apunta el control).
3. Posteriormente se avanza el control, una posición a la derecha.
4. Si no se alcanzó el fin de la cadena, se vuelve al paso 2.
5. Al finalizar, si el estado en el que quedó pertenece al conjunto F de estados finales, la cadena es aceptada (reconocida), y en caso contrario, es rechazada.



Lenguajes Formales y Autómatas Finitos

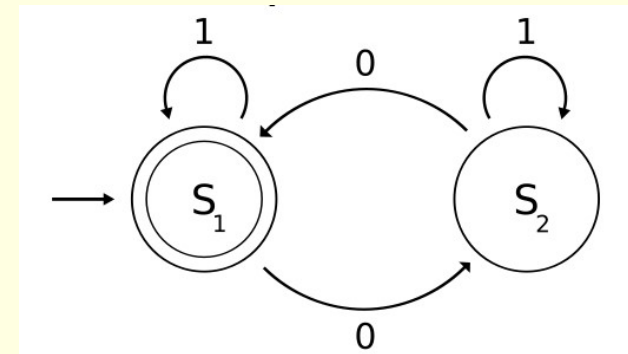
- **Un Autómata Finito define exactamente un Lenguaje (formal)**
- **Diagrama de transiciones:** es una representación gráfica de un AF, a través de un grafo dirigido y etiquetado, donde:
 - Los nodos del grafo representan los estados del autómata
 - Los arcos del grafo representan la función de transición, de tal manera de que el arco (x, y) posee la etiqueta a si, y solo si, $\delta(x, a)=y$
 - Los estados que se grafican con doble círculo son estados finales
 - El estado inicial se indica mediante una flecha sin nodo de partida



Lenguajes Formales y Autómatas Finitos

- **Tabla de Transiciones:** es una representación tabular de la función de transición, donde las filas son estados, las columnas son los símbolos de entrada y las celdas definen el estado que se alcanza desde el estado de la fila mediante el símbolo de la columna.

δ	0	1
S_1	S_2	S_1
S_2	S_1	S_2



- **Estado “Muerto”:** es todo estado de un AF que no es de aceptación y en el cual todas las transiciones de salida son ciclos hacia él mismo.
 - En símbolos
q es estado “muerto” si, y solo si : $(q \in (Q - F)) \wedge (\forall a \in \Sigma (\delta(q, a) = q))$

Lenguajes Formales y Autómatas Finitos

- **Lenguaje reconocido por un Autómata Finito**

- Función de transición extendida: $\hat{\delta}$

$$\hat{\delta}: Q \times \Sigma^* \rightarrow Q$$

- Definición:

$$\hat{\delta}(q, \varepsilon) = q$$

$$\hat{\delta}(q, aw) = \hat{\delta}(\delta(q, a), w) \quad \begin{array}{l} a \text{ en } \Sigma \\ w \text{ en } \Sigma^* \end{array}$$

- Dado un AF $M = (Q, \Sigma, \delta, q_0, F)$, el lenguaje que reconoce se denota L_M y es:

$$L_M = \{w \mid (w \text{ en } \Sigma^*) \text{ y } (\hat{\delta}(q_0, w) \text{ en } F)\}$$

- **La Clase de Lenguajes definidos por los Autómatas Finitos se denomina “Clase de los Lenguajes Regulares”**

Lenguajes Formales y Autómatas Finitos

Fin primera parte Análisis Léxico, Lenguajes
Formales y Autómatas Finitos