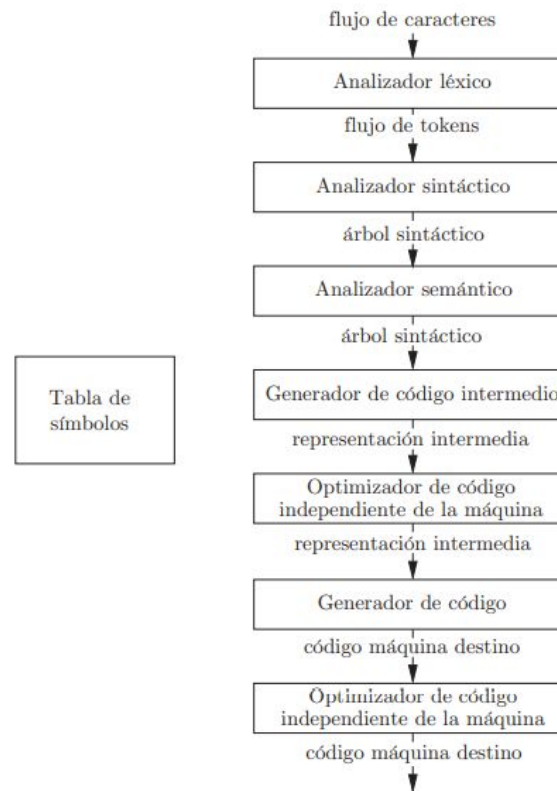


Compiladores

Análisis Léxico

Fases de un compilador



Función del Analizador Léxico

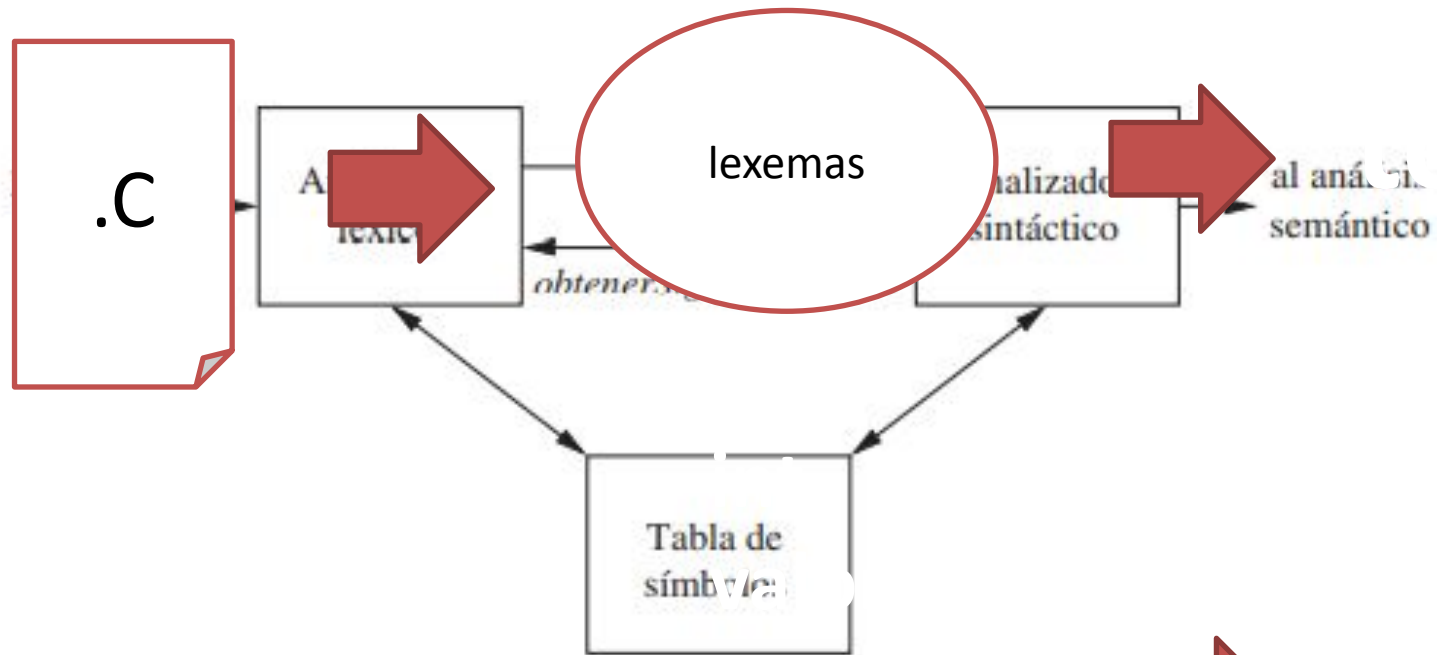


Figura 3.1: Interacción entre el analizador léxico y el analizador sintáctico

Proceso

- Preprocesamiento
- Eliminación de
 - Comentarios
 - Espacios en blanco (32)
 - Tabuladores (7)
 - Nueva línea (13)
 - Caracteres de separación

Token, patrones y lexemas

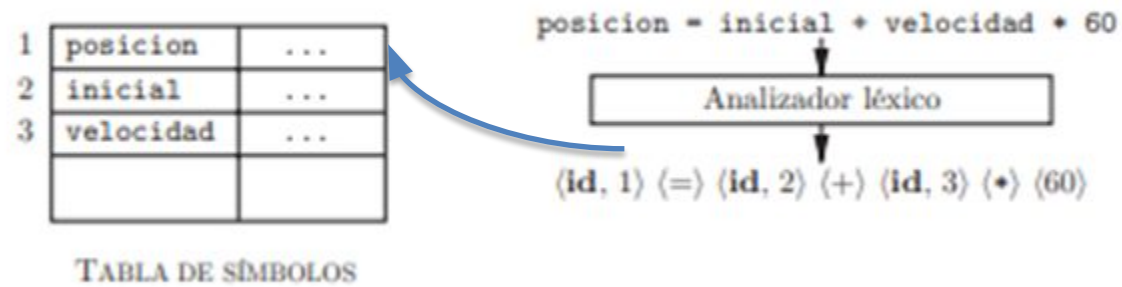
- Token
- Patrón
- Lexema

Tokens y lexemas

TOKEN	DESCRIPCIÓN INFORMAL	LEXEMAS DE EJEMPLO
if	caracteres i, f	if
else	caracteres e, l, s, e	Else
comparacion	< o > o <= o >= o == o !=	<=, !=
id	letra seguida por letras y dígitos	pi, puntuacion, D2
numero	cualquier constante numérica	3.14159, 0, 6.02e23
literal	cualquier cosa excepto ", rodeada por "'s	"core dumped"

Figura 3.2: Ejemplos de tokens

Atributo de un token



Errores léxicos

Recuperación de errores

- Eliminar un carácter
- Insertar un carácter
- Sustituir un carácter
- Transponer dos caracteres

```
void mani()  
{  
    intt año=0;  
    rerurn año;  
}
```

Ejercicio

- Obtenga los lexemas y los valores que se deberían asociar a él.

He aquí una foto de mi casa:

<P>

Vea Más Imágenes si
le gustó ésa.<P>

- Modifique los valores de la tabla y obtenga el resultado en
html

Búfer

Para lectura de archivos

es

Doble búfer

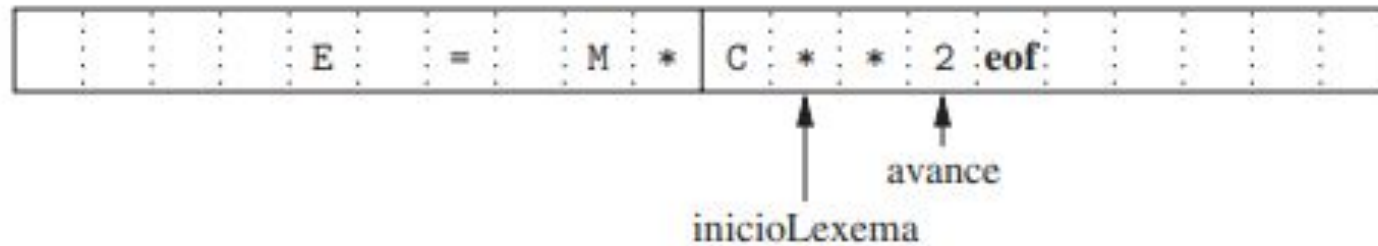
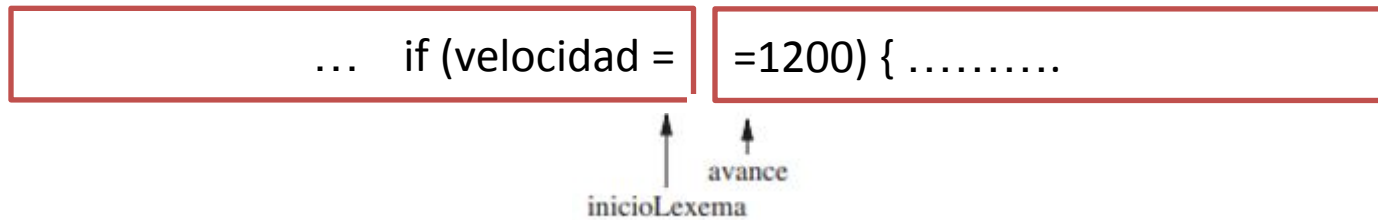


Figura 3.3: Uso de un par de búferes de entrada



Consideraciones

```
texto = "En un lugar de la Mancha, de cuyo nombre no quiero acordarme, "\n\n"no ha mucho tiempo que vivía un hidalgo de los de lanza en astillero, "\n\n"adarga antigua, rocín flaco y galgo corredor."
```

Búfer de 4k

```
... if (velocidad = Eob =1200) { ..... Eof
```

Tomado de: [Cadenas de texto en Python - Línea de Código \(lineadecodigo.com\)](#)

Ejercicio

- Leer de un archivo de texto para contar:
- Párrafos
- Palabras
- Caracteres

Toke

Conceptos
ns

Conceptos

- Alfabeto
 - Conjunto finito de símbolos $\{0,1\}$
- Cadena
 - Secuencia finita de símbolos de un alfabeto
 - Sea S una cadena
 - Longitud de cadena $|S|$
 - Cadena vacía ϵ

Conceto

- Lengua
- Con
- Caden
- Pref
- Sufij
- Subc
- Subs

Término	Definición
Cadena	Secuencia finita de símbolos de un alfabeto
Prefijo	Cadena que se obtiene de eliminar uno o mas elementos de la derecha de una cadena
Sufijo	Cadena que se obtiene de eliminar uno o mas elementos de la derecha de una cadena
Subcadena	Cadena que se obtiene de eliminar un prefijo y un sufijo
Subsecuencia	Cadena obtenida de la eliminación de símbolos no necesariamente contiguos

ϵ



$\langle \rangle S$ y $\langle \rangle \epsilon$

$\{ \epsilon \}$ y \emptyset (conjunto vacío) son lenguajes abstractos

Conceptos

- Sea x e y dos cadenas
 - xy es la concatenación de ambas cadenas
- Por lo tanto
 - $x \varepsilon = \varepsilon x = x$
- Si xx puede representarse como s^2 entonces s^0 es ε

Operaciones en los lenguajes

- Unión
- Concatenación
- Cerradura de Kleene
 - Sea L un lenguaje, L^* es el conjunto de cadenas que se obtiene de concatenar 0 o más veces
- Cerradura positiva L^+
 - Equivale a $L^* - \epsilon$

Operaciones en los lenguajes

OPERACIÓN	DEFINICIÓN Y NOTACIÓN
<i>Unión de L y M</i>	$L \cup M = \{ s \mid s \text{ está en } L \text{ o } s \text{ está en } M \}$
<i>Concatenación de L y M</i>	$LM = \{ st \mid s \text{ está en } L \text{ y } t \text{ está en } M \}$
<i>Cerradura de Kleene de L</i>	$L^* = \bigcup_{i=0}^{\infty} L^i$
<i>Cerradura positivo de L</i>	$L^+ = \bigcup_{i=1}^{\infty} L^i$

Figura 3.6: Definiciones de las operaciones en los lenguajes

Ejercicio

- Suponga que L es el conjunto de letras $\{A..Z, a..z\}$ y D el conjunto de dígitos $\{0..9\}$, obtenga 5 ejemplos de:
 - $L \cup D$
 - LD
 - L^4
 - L^*
 - $L(L \cup D)^*$
 - D^+

Ejercicio 2

- Obtenga el patrón que describe cada conjunto:

- $L \cup D$

1. $L \cup D$ es el conjunto de letras y dígitos; hablando en sentido estricto, el lenguaje con 62 cadenas de longitud uno, y cada una de las cadenas es una letra o un dígito.

- LD

2. LD es el conjunto de 520 cadenas de longitud dos, cada una de las cuales consiste en una letra, seguida de un dígito.

- L^4

3. L^4 es el conjunto de todas las cadenas de 4 letras.

- L^*

4. L^* es el conjunto de todas las cadenas de letras, incluyendo ϵ , la cadena vacía.

- $L(L \cup D)^*$

5. $L(L \cup D)^*$ es el conjunto de todas las cadenas de letras y dígitos que empiezan con una letra.

- D^+

6. D^+ es el conjunto de todas las cadenas de uno o más dígitos.

Toke

Expresiones regulares (p 147)

ns

Referencias

- Aho,
- Imágenes
 - Sanchez,J.J.(2018). Diferencias entre private, protected y public.
<https://janpierrsanchez.medium.com/diferencias-entre-private-protected-y-public-fb3ddb72c5e0>
 - Aho/Sethi text book