

4.2 Documentación ER y AF para Analizador Léxico

Tabla de Tokens

Token	Tipo	Identificador	Explicación
HTML	Palabra clave	página	Define el inicio del documento HTML.
HEAD	Palabra clave	cabecera	Contiene información sobre el documento.
TITLE	Palabra clave	título	Define el título del documento.
BODY	Palabra clave	Cuerpo	Contiene todo el contenido visible de la página.
H1-H6	Palabra clave	e1-e6	Etiquetas de encabezado de diferentes niveles.
P	Palabra clave	P	Define un párrafo de texto.
DIV	Palabra clave	bloque	Define una sección de la página.
SPAN	Palabra clave	contenedor	Define una sección en línea de la página.
A	Palabra clave	Enlace	Define un enlace a otra página.
IMG	Palabra clave	Imagen	Inserta una imagen en la página.
SRC	Atributo	Src	Define la URL de origen de una imagen.
HREF	Atributo	Href	Define la URL de destino de un enlace.
ALT	Atributo	Alt	Proporciona un texto alternativo para una imagen.
ID	Atributo	Id	Define un identificador único para el elemento.
CLASS	Atributo	Clase	Asigna una o más clases CSS al elemento.
STYLE	Atributo	Estilo	Define estilos CSS directamente en el elemento.
WIDTH	Atributo	Ancho	Define el ancho de un elemento.
HEIGHT	Atributo	Alto	Define la altura de un elemento.
TITLE	Atributo	Título	Proporciona información adicional sobre el elemento.
EQUALS	Operador	=	Asigna valores a los atributos.
QUOTE	Símbolo	"	Símbolo de comilla simple.
DOUBLE_QUOTE	Símbolo	""	Símbolo de comillas dobles.
SEMICOLON	Símbolo	;	Símbolo de punto y coma.
LEFT_ANGLE	Símbolo	<	Símbolo de menor que.
RIGHT_ANGLE	Símbolo	>	Símbolo de mayor que.
COMMENT	Comentario	<!--	Inicio de un comentario en HTML.

Reglas de coincidencia

- **HTML:** <html>
- **HEAD:** <encabezado>
- **TITLE:** <titulo>
- **BODY:** <cuerpo>
- **H1-H6:** <e1>|<e2>|<e3>|<e4>|<e5>|<e6>
- **P:** <p>
- **DIV:** <bloque>
- **SPAN:** <contenedor>
- **A:** <enlace>
- **IMG:**
- **SRC:** src\s*=\s*"(.*)"|src\s*=\s*'(.*)'
- **HREF:** href\s*=\s*"(.*)"|href\s*=\s*'(.*)'
- **ALT:** alt\s*=\s*"(.*)"|alt\s*=\s*'(.*)'
- **ID:** id\s*=\s*"(.*)"|id\s*=\s*'(.*)'
- **CLASS:** clase\s*=\s*"(.*)"|clases\s*=\s*'(.*)'
- **STYLE:** estilos\s*=\s*"(.*)"|estilos\s*=\s*'(.*)'
- **WIDTH:** ancho\s*=\s*"(.*)"|ancho\s*=\s*'(.*)'
- **HEIGHT:** alto\s*=\s*"(.*)"|alto\s*=\s*'(.*)'
- **TITLE:** titulo\s*=\s*"(.*)"|titulo\s*=\s*'(.*)'
- **EQUALS:** =
- **QUOTE:** '.*?'
- **DOUBLE_QUOTE:** ".*?"
- **SEMICOLON:** ;
- **LEFT_ANGLE:** <
- **RIGHT_ANGLE:** >
- **COMMENT:** <!--.*?-->

Estas reglas representan patrones de texto que serán reconocidos por el analizador léxico para identificar cada token en el texto de entrada. Cada regla está diseñada para coincidir con la estructura específica que define cada tipo de token en el lenguaje HTML-es

Asociar acciones a las reglas

1. Palabras clave

```
HTML    -> <html    { return "PALABRA_CLAVE_HTML"; }
HEAD    -> <encabezado { return
"PALABRA_CLAVE_ENCABEZADO"; }
TITLE   -> <titulo   { return "PALABRA_CLAVE_TITULO"; }
BODY    -> <cuerpo    { return "PALABRA_CLAVE_CUERPO"; }
H1      -> <e1      { return "PALABRA_CLAVE_E1"; }
H2      -> <e2      { return "PALABRA_CLAVE_E2"; }
H3      -> <e3      { return "PALABRA_CLAVE_E3"; }
H4      -> <e4      { return "PALABRA_CLAVE_E4"; }
H5      -> <e5      { return "PALABRA_CLAVE_E5"; }
H6      -> <e6      { return "PALABRA_CLAVE_E6"; }
P       -> <p       { return "PALABRA_CLAVE_P"; }
DIV     -> <bloque   { return "PALABRA_CLAVE_BLOQUE"; }
SPAN    -> <contenedor { return
"PALABRA_CLAVE_CONTENEDOR"; }
A       -> <enlace   { return "PALABRA_CLAVE_ENLACE"; }
```

2. Atributos

```
SRC      -> src\s*=\s*("[^"]*"|'[']*') { return "ATRIBUTO_SRC"; }
HREF     -> href\s*=\s*("[^"]*"|'[']*') { return "ATRIBUTO_HREF"; }
ALT      -> alt\s*=\s*("[^"]*"|'[']*') { return "ATRIBUTO_ALT"; }
ID       -> id\s*=\s*("[^"]*"|'[']*') { return "ATRIBUTO_ID"; }
CLASS    -> clase\s*=\s*("[^"]*"|'[']*'){ return "ATRIBUTO_CLASE"; }
STYLE    -> estilos\s*=\s*("[^"]*"|'[']*'){ return "ATRIBUTO_ESTILO"; }
WIDTH    -> ancho\s*=\s*("[^"]*"|'[']*'){ return "ATRIBUTO_ANCHO"; }
HEIGHT   -> alto\s*=\s*("[^"]*"|'[']*'){ return "ATRIBUTO_ALTO"; }
TITLE    -> titulo\s*=\s*("[^"]*"|'[']*'){ return "ATRIBUTO_TITULO"; }
```

3. Operador

```
EQUALS    -> =    { return "OPERADOR_IGUAL"; }
```

4. Símbolos

```
QUOTE     -> '     { return "SIMBOLO_COMILLA_SENCILLA"; }
```

```
DOUBLE_QUOTE -> "     { return "SIMBOLO_COMILLA_DOBLE"; }
```

```
SEMICOLON -> ;     { return "SIMBOLO_PUNTO_COMA"; }
```

```
LEFT_ANGLE -> <     { return "SIMBOLO_MENOR_QUE"; }
```

```
RIGHT_ANGLE -> >     { return "SIMBOLO_MAYOR_QUE"; }
```

5. Comentario

```
COMMENT   -> <!--.*?--> { return "COMENTARIO"; }
```