

Teoría de Lenguajes

Trabajo Práctico N° 1

Departamento de computación,
Facultad de Ciencias Exactas y Naturales,
Universidad de Buenos Aires

Primer cuatrimestre de 2013

Pablo Herrero
Pablo Romano
Paula Verghelet

Gramática

1. $S \rightarrow E<html>EAE</html>E \mid E<html>ECE</html>E \mid E<html>E</html>E$
2. $A \rightarrow <head>EBE</head>ECE$
3. $B \rightarrow <title>T</title>EBE \mid <script>T</script>EBE \mid \lambda$
4. $C \rightarrow <body>EDE</body> \mid <body>TDT</body> \mid \lambda$
5. $D \rightarrow TD \mid \lambda \mid <div>EDE</div>EDE \mid <p>EDE</p> EDE \mid <h1>EDE</h1> EDE \mid
EDE \mid <div>TDT</div>TDT \mid <p>TDT</p>TDT \mid <h1>TDT</h1>TDT \mid
TDT$
6. $E \rightarrow \text{espacio}E \mid E<!--T-->E \mid \lambda$
7. $T \rightarrow (a|...|z|A|...|Z|0|...|9|...|&|;|\text{espacio})T \mid E \mid \lambda$

Descripción de la gramática

1. Todas las derivaciones del símbolo distinguido comienzan (terminan) con cero o más espacios y luego un tag de apertura (cierre) de html, la derivación que contiene A corresponde a un HTML con head mientras que la que contiene C es un HTML que no tiene head
2. La derivación de A corresponde a un head seguido de cero o más espacios, cero o un body y cero o más espacios
3. La derivación de B es el contenido que puede estar dentro de un head, estos son titles (conteniendo texto) o scripts (conteniendo texto), en ambos casos puede haber cero o más ocurrencias
4. La derivación de C permite que haya cero o un body conteniendo tags, ver EDE y TDT en el punto siguiente.
5. Derivaciones de D
 - a. TD , esto permite agregar texto a izquierda de los tags, como el tag puede derivar en λ esto también permite terminar con texto a derecha "quitando" el último D
 - b. Las derivaciones rodeadas de div, p y h1 permiten generar estos tags, en todos los casos contienen EDE que permite nuevos tags rodeados de cero o más espacios o derivar D a TD y generar texto
 - c. La derivación $
EDE$ permite generar el tag $
$
 - d. Las derivaciones que contienen TDT son todas redundantes por lo expuesto en 5.a y 5.b pero hacen que los ejemplos sean más compactos
6. Esta derivación permite generar cero o más espacios o comentarios
7. Esta derivación permite generar texto de longitud cero o mayor, se agregan los símbolos $\&$ y $;$ para poder codificar las entidades html y se permite derivar a E para poder incluir comentarios ya que esto siempre es válido

Decisiones tomadas

Se optó por un alfabeto sencillo (letras minúsculas, mayúsculas, espacio, números, ampersand, punto y coma, mayor, menor y barra separadora) pero abarcativo (contiene una variedad de símbolos distintos) para mostrar que puede ser rápidamente extendido a todo el Unicode, algo que consideramos innecesario en esta etapa del TP.

Optamos por no permitir los símbolos mayor y menor en el texto libre, de esta forma impedimos la apertura y cierre de tags en lugares inapropiados (por ejemplo dentro del title), esta decisión se fundamenta en el hecho de que también están prohibidos estos caracteres especiales en

HTML y para usarlos en texto libre se los debe codificar como > y <, mayor y menor respectivamente, al agregar los símbolos & y ; permitimos esta codificación.

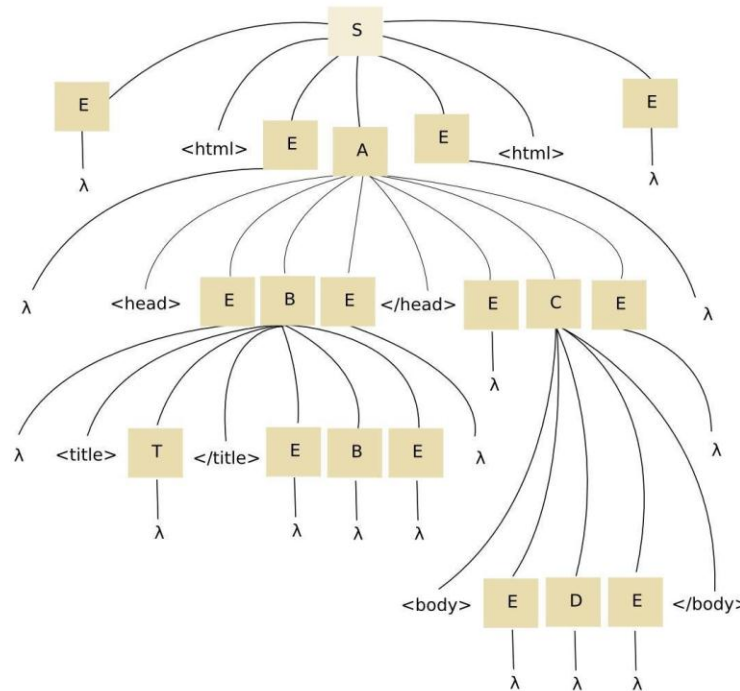
Finalmente tomamos una decisión un tanto controversial (al menos dentro de nuestro grupo), esta fue la incorporación a conciencia de redundancia en la gramática (ver Descripción de la gramática 5.d), esta redundancia, a pesar de ser prescindible, permite derivaciones más compactas permitiendo derivar $D \rightarrow EDE \rightarrow ETDE \rightarrow ETDTE \rightarrow TDTDE \rightarrow TDT$ en un único paso, es decir $D \rightarrow TDT$, con la ventaja académica de simplificar los árboles de derivación entregados en el punto siguiente y ecológica de imprimir menos hojas.

Ejemplo de árboles de derivación con entradas correctas

1) sin espacios en blanco

`<html><head><title></title></head><body></body></html>`

S -> **E<html>EAE</html>E**
 -> **λ<html>EAE</html>λ**
 -> **<html>EAE</html>**
 -> **<html>E<head>EBE</head>ECE</html>** (por A)
 -> **<html>λ<head>EBE</head>ECEλ</html>** (por E)
 -> **<html><head>E<title> T </title>EBEE</head>ECE</html>** (por B)
 -> **<html><head>λ<title> T </title>EBEλ</head>ECE</html>** (por E)
 -> **<html><head><title> T </title>EλE</head>ECE</html>** (por B)
 -> **<html><head><title>λ</title>EE</head>ECE</html>** (por T)
 -> **<html><head><title></title>λλ</head>E <body>EDE</body>E</html>** (por E)
 -> **<html><head><title></title></head>E <body>EλE</body>E</html>** (por D)
 -> **<html><head><title></title></head>E <body>λλ</body>E</html>** (por E)
 -> **<html><head><title></title></head>λ<body></body>λ</html>** (por E)
 -> **<html><head><title></title></head><body></body></html>**



Árbol de derivación para el ejemplo 1

2) del enunciado

`<html> <head><title>Esto es un título</title> <!-- esto es un comentario --> <script>print("Hola mundo")</script></head> <body> Esto es <p>una <h1>prueba</h1></p>
</body></html>`

S → **E<html>EAE</html>E**

(por E) → **λ<html>EAE</html>λ**

→ **<html>EAE</html>**

(por A) → **<html>E<head>EBE</head>ECE</html>**

(por E) → **<html>_<head>EBE</head>ECEλ</html>**

(por B) → **<html>_<head>E<title>T</title>EBEE</head>ECE</html>**

(por E) → **<html>_<head>λ<title>T</title>EBEλ</head>ECE</html>**

(por B) → **<html>_<head><title>T</title>E<script>T</script>EBEE</head>ECE</html>**

(por E) → **<html>_<head><title>T</title>_E<script>T</script>EBEE</head>ECE</html>**

(por E) → **<html>_<head><title>T</title>_E<!-- T -->E<script>T</script>EBEE</head>ECE</html>**

(por E) → **<html>_<head><title>T</title>_λ<!-- T -->_<script>T</script>EBEE</head>ECE</html>**

(por T) → **<html>_<head><title>T</title>_<!-- eT -->_<script>T</script>EBEE</head>ECE</html>**

(por T) → **<html>_<head><title>T</title>_<!-- esT -->_<script>T</script>EBEE</head>ECE</html>**

(por T) → **<html>_<head><title>T</title>_<!-- estT -->_<script>T</script>EBEE</head>ECE</html>**

(por T) → **<html>_<head><title>T</title>_<!-- estoT -->_<script>T</script>EBEE</head>ECE</html>**

(por T) → ... (por T)

(por T) → **<html>_<head><title>T</title>_<!-- esto es un comentario_T -->_**

<script>T</script>EBEE</head>ECE</html>

[illegible]

(por E) -> <html>_<head><title>Esto_es_un_título</title>_<!-- esto_es_un_comentario_ ->_<script>print("Hola_mundo")</script></head>_<body>_Esto_es_<p>λDλ</p>EDE</body></html>

(por D) -> <html>_<head><title>Esto_es_un_título</title>_<!-- esto_es_un_comentario_ ->_<script>print("Hola_mundo")</script></head>_<body>_Esto_es_<p>TD</p>EDE</body></html>

(por T) -> <html>_<head><title>Esto_es_un_título</title>_<!-- esto_es_un_comentario_ ->_<script>print("Hola_mundo")</script></head>_<body>_Esto_es_<p>una_TD</p>EDE</body></html>

(por T) -> <html>_<head><title>Esto_es_un_título</title>_<!-- esto_es_un_comentario_ ->_<script>print("Hola_mundo")</script></head>_<body>_Esto_es_<p>una_λD</p>EDE</body></html>

(por D) -> <html>_<head><title>Esto_es_un_título</title>_<!-- esto_es_un_comentario_ ->_<script>print("Hola_mundo")</script></head>_<body>_Esto_es_<p>una_<h1>EDE</h1>EDE</p>>EDE</body></html>

(por E) -> <html>_<head><title>Esto_es_un_título</title>_<!-- esto_es_un_comentario_ ->_<script>print("Hola_mundo")</script></head>_<body>_Esto_es_<p>una_<h1>λDλ</h1>λDλ</p>EDE</body></html>

(por D) -> <html>_<head><title>Esto_es_un_título</title>_<!-- esto_es_un_comentario_ ->_<script>print("Hola_mundo")</script></head>_<body>_Esto_es_<p>una_<h1>TD</h1>λ</p>EDE</body></html>

(por D) -> <html>_<head><title>Esto_es_un_título</title>_<!-- esto_es_un_comentario_ ->_<script>print("Hola_mundo")</script></head>_<body>_Esto_es_<p>una_<h1>Tλ</h1></p>EDE</body></html>

(por T) -> <html>_<head><title>Esto_es_un_título</title>_<!-- esto_es_un_comentario_ ->_<script>print("Hola_mundo")</script></head>_<body>_Esto_es_<p>una_<h1>pT</h1></p>EDE</body></html>

(por T) ->... (por T)

(por T) -> <html>_<head><title>Esto_es_un_título</title>_<!-- esto_es_un_comentario_ ->_<script>print("Hola_mundo")</script></head>_<body>_Esto_es_<p>una_<h1>pruebaT</h1></p>EDE</body></html>

(por T) -> <html>_<head><title>Esto_es_un_título</title>_<!-- esto_es_un_comentario_ ->_<script>print("Hola_mundo")</script></head>_<body>_Esto_es_<p>una_<h1>pruebaλ</h1></p>EDE</body></html>

(por D) -> <html>_<head><title>Esto_es_un_título</title>_<!-- esto_es_un_comentario_ ->_<script>print("Hola_mundo")</script></head>_<body>_Esto_es_<p>una_<h1>prueba</h1></p>E
EDEE</body></html>

(por D) -> <html>_<head><title>Esto_es_un_título</title>_<!-- esto_es_un_comentario_ ->_<script>print("Hola_mundo")</script></head>_<body>_Esto_es_<p>una_<h1>prueba</h1></p>E
EλEE</body></html>

(por E) -> <html>_<head><title>Esto_es_un_título</title>_<!-- esto_es_un_comentario_ ->_<script>print("Hola_mundo")</script></head>_<body>_Esto_es_<p>una_<h1>prueba</h1></p>E
λλλ</body></html>

(por E) -> `<html>_<head><title>Esto_es_un_título</title>_ <!-- esto_es_un_comentario_ -->_<script>print("Hola_mundo")</script></head>_<body>_Esto_es_<p>una_<h1>prueba</h1></p>_
</body></html>`

(Reemplazamos el “_” agregado por legibilidad por un espacio en blanco.)

-> `<html> <head><title>Esto es un título</title> <!-- esto es un comentario --> <script>print("Hola mundo")</script></head> <body> Esto es <p>una <h1>prueba</h1></p>
</body></html>`

Ejemplo 3) un html sin head, con comentario y tags anidados

`<html> <body><!-- pinto TP--> <div>aaa<div>bbb</div>ccc<p>ddd</p>eee</div>fff</body></html>`

Derivación

Por 1:

$S \rightarrow E<html>ECE</html>$

Por 4:

$E<html>ECE</html> \rightarrow E<html>E<body>TDT</body>E</html>$

Por 6:

$E<html>E<body>TDT</body>E</html> \rightarrow E<html> E<body>TDT</body>E</html>$

Por 6:

$E<html> <body>TDT</body>E</html>$

Por 6:

$<html> <body>TDT</body>E</html>$

Por 7:

$<html> <body>EDT</body>E</html>$

Por 6:

$<html> <body>E<!-- T -->EDT</body>E</html>$

Por 6:

$<html> <body><!-- T -->EDT</body>E</html>$

Por 6:

$<html> <body><!-- T -->DT</body>E</html>$

Por 5:

$<html> <body><!-- T -->TDT</body>E</html>$

Por 6:

$<html> <body><!-- T --> TDT</body>E</html>$

Por 6:

$<html> <body><!-- T --> DE</body>E</html>$

Por 7 varias veces:

$<html> <body><!-- pinto TP --> DE</body>E</html>$

Por 6:

$<html> <body><!-- pinto TP --> DE</body></html>$

Por 6:

$<html> <body><!-- pinto TP --> D</body></html>$

Por 5:

$<html> <body><!-- pinto TP --> <div>TDT</div>TDT</body></html>$

Por 7 varias veces:

$<html> <body><!-- pinto TP --> <div>TDT</div>TDffT</body></html>$

Por 5:

3)

```
<html> <body><!-- pinto TP--> <div>aaa<div>bbb</div>ccc<p>ddd</p>eee</div>fff</body></html>
```

Derivación

S → E<html>ECE</html>

(por C) → E<html>E<body>EDE</body>E</html>

(por E) → **Λ**<html>_<body>**EDΛ**</body>**Λ**</html>

→ <html>_<body>ED</body></html>

(por E) → <html> <body>E<!-- T -->ED</body></html>

(por E) → <html> <body><!-- T -->D</body></html>

(por T) → <html> <body><!-- pinto TP --> D</body></html>

(por D)

→ <html> <body><!-- pinto TP --> <div>EDE</div>EDE</body></html>

(por E)

→ <html> <body><!-- pinto TP --> <div>D</div>D</body></html>

(por D)

→ <html> <body><!-- pinto TP --> <div>TD</div>TD</body></html>

(por D)

→ <html> <body><!-- pinto TP --> <div>T<div>EDE</div>EDE</div>TD</body></html>

(por E)

→ <html> <body><!-- pinto TP --> <div>T<div>D</div>D</div>TD</body></html>

(por T)

→ <html> <body><!-- pinto TP --> <div>aaa<div>D</div>D</div>TD</body></html>

(por D)

→ <html> <body><!-- pinto TP --> <div>aaa<div>TD</div>D</div>TD</body></html>

(por T)

→ <html> <body><!-- pinto TP --> <div>aaa<div>bbbTD</div>D</div>TD</body></html>

(por T y D)

→ <html> <body><!-- pinto TP --> <div>aaa<div>bbb</div>D</div>TD</body></html>

(por D)

→ <html> <body><!-- pinto TP --> <div>aaa<div>bbb</div>TD</div>TD</body></html>

(por D)

→ <html> <body><!-- pinto TP --> <div>aaa<div>bbb</div>T<p>EDE</p> EDE</div>TD</body></html>

(por E)

→ <html> <body><!-- pinto TP --> <div>aaa<div>bbb</div>T<p>D</p>D</div>TD</body></html>

(por T)

→ <html> <body><!-- pinto TP --> <div>aaa<div>bbb</div>cccT<p>D</p>D</div>TD</body></html>

(por D)

→ <html> <body><!-- pinto TP --> <div>aaa<div>bbb</div>cccT<p>TD</p>TD</div>TD</body></html>

(por D)

→ <html> <body><!-- pinto TP --> <div>aaa<div>bbb</div>cccT<p>T</p>T</div>T</body></html>

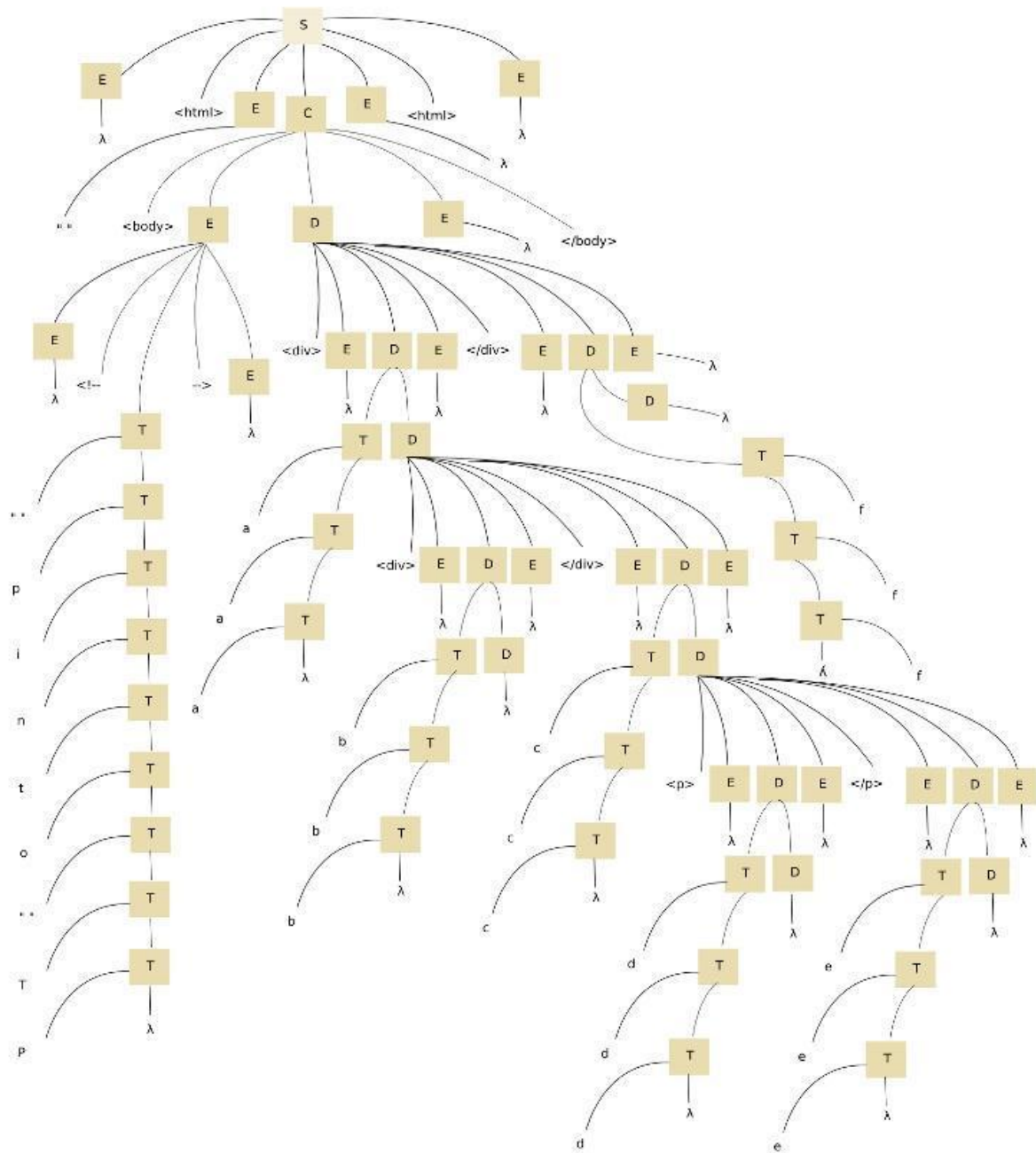
(por T)

→ <html> <body><!-- pinto TP -->

<div>aaa<div>bbb</div>cccT<p>dddT</p>eeeT</div>fffT</body></html>

(por T)

→<html> <body><!-- pinto TP --> <div>aaa<div>bbb</div>ccc<p>ddd</p>eee</div>fff</body></html>



Ejemplo 4) ambigüedad

<html><body>Hola Mundo!</body></html>

Una derivación posible

S → **E<html>ECE</html>E**

(por E) → **λ<html>ECE</html>λ**

→ **<html>ECE</html>**

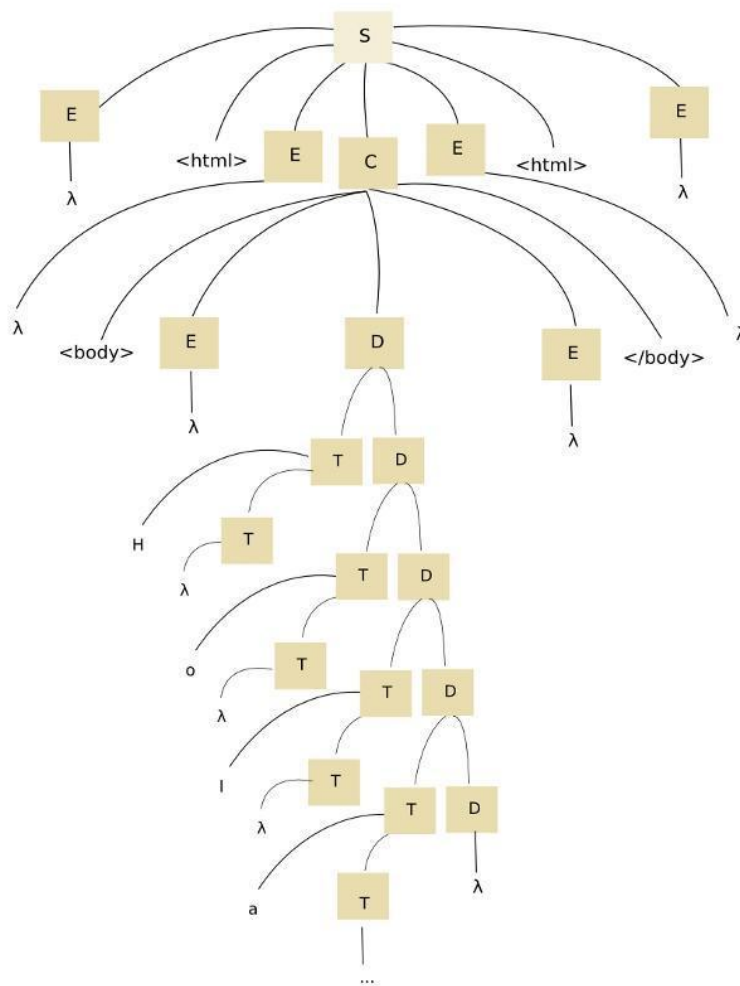
(por E) → **<html>λCλ</html>**

(por C) -><html><body>EDE</body></html>
 (por D) -><html><body>TD</body></html>
 (por D) -><html><body>TTD</body></html>
 (por D) -><html><body>TTTD</body></html>
 (por D) -><html><body>TTTTD</body></html>
 (por D) -><html><body>TTTTλ</body></html>
 (por T) -><html><body>HTTTT</body></html>
 (por T) -><html><body>HλTTT</body></html>
 (por T) -><html><body>HoTTT</body></html>
 (por T) -><html><body>HoλTT</body></html>
 (por T) -><html><body>HoITT</body></html>
 (por T) -><html><body>HoIλT</body></html>
 (por T) -><html><body>HolaT</body></html>

....

Continúa del mismo modo hasta finalizar el texto

(por T) -><html><body>Hola Mundo!λ</body></html>
 <html><body>Hola Mundo!</body></html>



S $\rightarrow E \langle \text{html} \rangle ECE \langle / \text{html} \rangle E$

-> <html>ECE</html>

(por C) -><html><body>EDE</body></html>

(por T) -><html><body>HTD</body></html>

(por T) -><html><body>HoITD</body></html>

...

```
(por T) -><html><body>Hola Mundo!\lambda D</body></html>
```

```
<html><body>Hola Mundo!</body></html>
```

