

Universidad de Buenos Aires
Facultad de
Ciencias Exactas y Naturales
Departamento de Computación

Teoría de Lenguajes
Primer Cuatrimestre de 2013

Trabajo práctico

Micro HTML Prettyprint

Integrante	LU	Correo electrónico
Cammi, Martín	676/02	<code>martincammi@gmail.com</code>
De Sousa Bispo, Mariano	389/08	<code>marian.sabianaa@hotmail.com</code>
Felisatti, Ana	335/10	<code>anafelisatti@gmail.com</code>

1. Introducción

En el presente trabajo práctico definiremos la gramática para generar un lenguaje que interprete una versión simplificada de HTML.

2. Asunciones

3. Gramática

$G = \langle V_N, V_T, P, \text{BEGIN} \rangle$ donde

$V_N = \{ \text{BEGIN, HTML, HEAD, LEER_HEAD, TITLE, SCRIPTS, MORE_SCRIPTS, BODY, LEER_BODY} \}$

$V_T = \{ \text{initHtml, endHtml, initHead, endHead, initTitle, endTitle, initScript, endScript, noScripts, initBody, endBody, initDiv, endDiv, initH1, endH1, initP, endP, br, noTags} \}$

y P está dada por:

BEGIN	→	initHtml HTML endHtml
HTML	→	HEAD BODY
HEAD	→	$\lambda \mid \text{initHead LEER_HEAD endHead}$
LEER_HEAD	→	SCRIPTS TITLE MORE_SCRIPTS TITLE MORE_SCRIPTS MORE_SCRIPTS
TITLE	→	initTitle noTags endTitle
SCRIPTS	→	initScript noScripts endScripts MORE_SCRIPTS
MORE_SCRIPTS	→	$\lambda \mid \text{initBody LEER_BODY endBody}$
BODY	→	$\lambda \mid \text{initBody LEER_BODY endBody}$
LEER_BODY	→	$\lambda \mid \text{noTags LEER_BODY} \mid \text{br LEER_BODY} \mid$ $\text{initDiv LEER_BODY endDiv LEER_BODY} \mid$ $\text{initP LEER_BODY endP LEER_BODY} \mid$ $\text{initH1 LEER_BODY endH1 LEER_BODY}$

4. Expresiones regulares

<i>Terminal</i>	<i>Expresión regular</i>
initHtml	<html>
endHtml	</html>
initHead	<head>
endHead	</head>
initBody	<body>
endBody	</body>
initTitle	<title>
endTitle	</title>
initScript	<script>
endScript	</script>
initDiv	<div>
endDiv	</div>
initH1	<h1>
endH1	</h1>
initP	<p>
endP	</p>
br	
noTags	$\wedge((?!(\text{TAGS_GRAMATICA})).)*\$$
noScripts	$\wedge((?!(<\text{script}> </\text{script}>)).)*\$$

Donde TAGS_GRAMATICA = <html>|</html>|<head>|</head>|<body>|</body>|<title>|</title>|<script>|</script>|<div> |</div>|<p>|</p>|<h1>|</h1>|

El tag *noTags* se representa con una expresión regular que identifica cualquier texto sin tags definidos por la gramática.

Por otro lado, la expresión regular para identificar los comentarios es la siguiente:

$$comment = <!(.*?)->$$

Los diferentes símbolos utilizados en las expresiones denotan:

- | \longrightarrow Indica disyunción entre dos expresiones.
- \wedge \longrightarrow Matchea la expresión al comienzo de una línea.
- $\$$ \longrightarrow Matchea la expresión al final de una línea.
- ?! \longrightarrow No matchea la aparición denotada a su derecha.
- * \longrightarrow Indica una repetición de 0 o más veces.

5. Árboles de derivación