

# E79 Procesadores de lenguaje

## Examen de teoría (12 de julio de 2001)

PREGUNTA 1

(6 PUNTOS)

Para mantener tu página web, has decidido crear un pequeño lenguaje similar al  $\text{T}_{\text{E}}\text{X}$ . La idea es que el texto que escribas será procesado por tu programa,  $\text{W}_{\text{E}}\text{B}_{\text{T}}\text{E}_{\text{X}}$ , que generará el correspondiente fichero HTML. Los ficheros  $\text{W}_{\text{E}}\text{B}_{\text{T}}\text{E}_{\text{X}}$  consisten en una serie de párrafos de texto que pueden incluir algunas órdenes especiales. Para marcar un párrafo se utilizan una o más líneas en blanco (una línea en blanco se marca mediante un fin de línea seguido de espacios o tabuladores y otro fin de línea). En el fichero resultante, los párrafos se marcarán mediante el *tag*  $\langle P \rangle$ . (En HTML, los saltos de línea no son significativos y se considera igual un espacio que varios al separar palabras).

Algunos caracteres se escriben de forma especial en HTML. En  $\text{W}_{\text{E}}\text{B}_{\text{T}}\text{E}_{\text{X}}$  se escriben normalmente y en la salida aparece su equivalente. Sólo consideraremos:  $\langle ' \rangle \Rightarrow \langle \&lt; \rangle$ ;  $\langle ' \rangle \Rightarrow \langle \&gt; \rangle$ ;  $\langle \& \rangle \Rightarrow \langle \&amp; \rangle$ ; y  $\langle " \rangle \Rightarrow \langle \&quot; \rangle$ .

Las órdenes se escriben mediante una barra ( $\backslash$ ) seguida de una o más letras. Si la orden tiene parámetros, estos se escriben a continuación encerrados entre llaves (si el parámetro incluye alguna llave cerrada, se escribe escapándola ( $\backslash$ )). Puede haber espacios entre el nombre de la orden y sus parámetros. Algunas órdenes ya están predefinidas:

Orden	Parámetros	Significado	Tag
$\backslash\text{enlace}$	texto, enlace	crea el enlace con el texto dado	$\langle A \text{ href}=\text{"enlace"} \rangle, \langle /A \rangle$
$\backslash\text{HTML}$	texto	se copia el texto literalmente en la salida	—
$\backslash\text{italica}$	—	pone el texto en itálica	$\langle I \rangle, \langle /I \rangle$
$\backslash\text{negrita}$	—	pone el texto en negrita	$\langle B \rangle, \langle /B \rangle$
$\backslash\text{titulo}$	nivel, texto	pone un texto como título del nivel indicado (hasta 6)	$\langle Hn \rangle, \langle /Hn \rangle$

Las ordenes de cambio de tipo de letra, tienen efecto hasta el final del ámbito en el que aparecen. Los ámbitos se marcan mediante llaves. De esta manera, la entrada

```
{Texto normal, \italica texto en itálica,} texto normal
```

se traduciría por

```
Texto normal,  $\langle I \rangle$  texto en itálica,  $\langle /I \rangle$  texto normal
```

Las llaves no tienen ningún otro efecto, no aparecen en la salida y pueden anidarse. Si aparecen dentro del cuerpo de órdenes del usuario, deben estar correctamente anidadas.

Además, el usuario puede crear sus propias órdenes utilizando la orden  $\backslash\text{macro}$ . Esta orden tiene tres parámetros:

- El primero, opcional, entre corchetes y sin espacios, indica el número de parámetros (entre 1 y 9, ambos inclusive). En caso de no aparecer, la orden no tendrá parámetros.
- El segundo parámetro es el nombre de la orden.
- El tercero tiene el cuerpo de la orden. Los parámetros recibidos por la orden se indican mediante su número precedido por el símbolo  $\#$ .

Las órdenes del usuario funcionan como macros. En el punto de su aparición se sustituye la orden por la expansión de su cuerpo. En caso de haber parámetros, estos se expanden primero y después se sustituyen en el lugar correspondiente. Si en el cuerpo de la orden hay otras órdenes, estas se expanden *en el momento de la ejecución*.

Por ejemplo:

```
\macro[1]{\examen}{examen de la \universidad en la convocatoria de #1}  
\macro{\universidad}{Universitat Jaume I} \macro{\mes}{junio}  
Esto es un \examen{\mes};  
\macro{\universidad}{Universitat de València}  
y esto es un \examen{\mes}.
```

produciría:

```
Esto es un examen de la Universitat Jaume I en la convocatoria de junio;  
y esto es un examen de la Universitat de València en la convocatoria de junio.
```

Las macros definidas por el usuario tienen validez desde el momento de su definición en adelante, aunque pueden aparecer antes en el cuerpo de macros que no se expandan antes de la definición.

Un ejemplo de texto y su correspondiente traducción sería:

<pre>\macro{\prueba}{\italica\negrita prueba} \titulo{2}{Esta es mi página de \prueba}  Como todas las páginas de \prueba, incluye \enlace {enlaces}     {http://mimaquina.es/mislinks.htm} a otras páginas de {\negrita interés}.  \macro[1]{\dos}{#1#1}\macro{\tres}{\dos{\aaa}\aaa} \macro{\aaa}{&lt;algo&gt;} Ahora vienen {\negrita tres repeticiones} de \aaa:  \tres  \macro{\aaa}{otro}y ahora se repite \aaa:  \tres  \macro[1]{\rojo} {\HTML{&lt;FONT color="red"&gt;}#1\HTML{&lt;/FONT&gt;}} \rojo{Esto está en rojo}</pre>	<pre>&lt;H2&gt;Esta es mi página de &lt;I&gt;&lt;B&gt; prueba&lt;/B&gt;&lt;/I&gt;&lt;/H2&gt; &lt;P&gt; Como todas las páginas de &lt;I&gt;&lt;B&gt; prueba&lt;/B&gt;&lt;/I&gt;, incluye &lt;A href="http://mimaquina.es/mislinks.htm"&gt;enlaces&lt;/A&gt; a otras páginas de &lt;B&gt; interés&lt;/B&gt;. &lt;P&gt; Ahora vienen &lt;B&gt; tres repeticiones&lt;/B&gt; de &amp;lt;algo&amp;gt;; &lt;P&gt; &amp;lt;algo&amp;gt;&amp;lt;algo&amp;gt;&amp;lt;algo&amp;gt;&amp;lt;algo&amp;gt; &lt;P&gt; y ahora se repite &amp;quot;otro&amp;quot;; &lt;P&gt; &amp;quot;otro&amp;quot;&amp;quot;otro&amp;quot;&amp;quot;otro&amp;quot;&amp;quot;otro&amp;quot; &lt;P&gt; &lt;FONT color="red"&gt;Esto está en rojo&lt;/FONT&gt;</pre>
--	---

Se pide:

- Especificación del analizador léxico, incluyendo, al menos: categorías léxicas, sus expresiones regulares, sus atributos y las acciones correspondientes.
- Especificación del analizador sintáctico mediante una gramática incontextual.
- Transformación de la gramática anterior en una gramática LL(1), RLL(1), LR(0), SLR, LR(1) o LALR. Demostración de que la gramática transformada pertenece a alguna de estas familias.
- Relación (no implementación) de las comprobaciones semánticas que debe realizar  $\text{WEBTEX}$ .
- Gramática atribuida (basada en la original o en la transformada) con las acciones semánticas para:
  - traducir la entrada de  $\text{WEBTEX}$  a HTML; o
  - crear una representación intermedia de la entrada. En este caso, debe explicarse cómo se realizaría el paso de la representación intermedia al texto HTML.

**Cualquier decisión de diseño que se tome, debe ser indicada explícitamente.**

PREGUNTA 2

(2 PUNTOS)

¿Cuáles de las siguientes restricciones a los números de un lenguaje de programación pueden expresarse en el analizador léxico mediante expresiones regulares y cuáles deben expresarse mediante las acciones asociadas? Justifica las respuestas.

- El número tiene que ser primo.
- El número tiene que ser múltiplo de siete.
- El número tiene que ser capicúa.
- El número tiene que ocupar menos de 16bits.

PREGUNTA 3

(2 PUNTOS)

Sea  $G$  una gramática con dos no terminales  $\langle A \rangle$  y  $\langle B \rangle$  tales que hay un terminal  $a$  que es primero de ambos. ¿Cuáles de las siguientes afirmaciones son ciertas? Justifica las respuestas.

- Si existen las reglas  $\langle C \rangle \rightarrow \langle A \rangle \alpha | \langle B \rangle \beta$ , la gramática no puede ser LL(1).
- Si existe la regla  $\langle C \rangle \rightarrow \langle A \rangle \alpha \langle B \rangle \beta$ , la gramática no puede ser LL(1).
- Si existen las reglas  $\langle C \rangle \rightarrow \langle A \rangle \alpha | \langle B \rangle \beta$ , la gramática no puede ser LR(1).
- Si existe la regla  $\langle C \rangle \rightarrow \langle A \rangle \alpha \langle B \rangle \beta$ , la gramática no puede ser LR(1).