

# Grado en Ingeniería Informática

## Procesadores de lenguajes - Análisis léxico

### Práctica complementaria

#### **Objetivos de la práctica**

- Desarrollar un traductor simple con JFlex.

#### **Entrega**

La entrega será individual y consistirá **únicamente** en la especificación jflex desarrollada (fichero **.flex**). La fecha de entrega será el domingo **19 de febrero a las 23:55** debiéndose usar la entrega facilitada en el campus virtual “Entrega práctica complementaria – Analizador léxico”.

#### **Enunciado**

Se desea realizar un visualizador de documentos escritos con el formato [Markdown](#). Markdown permite generar de forma rápida documentos formateados mediante marcas de texto sencillas, no tan complejas como LaTeX o HTML. El objetivo de la práctica es generar un documento HTML a partir del contenido formateado con Markdown. No se pretende cubrir la totalidad del lenguaje Markdown, tan solo los aspectos aquí descritos clasificados según tres niveles de calificación.

Para todos los niveles, el comienzo del fichero HTML a generar será el siguiente:

Comienzo HTML
<pre>&lt;!DOCTYPE html&gt; &lt;html&gt; &lt;head&gt;   &lt;title&gt;Markdown2HTML&lt;/title&gt; &lt;style&gt;   .bold {font-weight: bold;}   .ital  {font-style:  italic;} &lt;/style&gt; &lt;/head&gt; &lt;body&gt;</pre>

Y el final del fichero HTML será:

Fin HTML
<pre>&lt;/body&gt; &lt;/html&gt;</pre>

## Nivel aprobado

Para llegar al aprobado (Nota = 5) se deben proporcionar todas las siguientes características (nótese que no se pueden anidar ninguno de estos aspectos ni con ellos mismos ni con los otros):

- Encabezados: se permiten definir encabezados usando símbolos “#” seguidos de un espacio, el resto del contenido de la línea es el texto del encabezado. Un solo símbolo ”# ” equivale a un encabezado de primer nivel (elemento H1 de HTML), “## ” equivaldría a un H2 y así sucesivamente hasta H6.

Código Markdown	Código HTML a generar
# Encabezado de primer nivel ## Encabezado de segundo nivel #### Encabezado de cuarto nivel	<H1>Encabezado de primer nivel</H1> <H2>Encabezado de segundo nivel</H2> <H4>Encabezado de cuarto nivel</H4>

- Texto en negrita o cursiva: todo texto rodeado por dos parejas de dos asteriscos “\*\*” o dos guiones bajos “\_\_” se formateará como texto en negrita. Además, todo texto rodeado por dos parejas de un asterisco “\*” o guion bajo “\_” se formateará como texto en cursiva.

Código Markdown	Código HTML a generar
el texto en negrita se formatea con dos **asteriscos** o __guiones bajos__	el texto en negrita se formatea con dos <SPAN class="bold">asteriscos</SPAN> o <SPAN class="bold">guiones bajos</SPAN>
el texto en cursiva se formatea con un *asterisco* o _guiones bajos_	el texto en cursiva se formatea con un <SPAN class="ital">asterisco</SPAN> o <SPAN class="ital">guiones bajos</SPAN>

- Las citas se especificarán con el símbolo “>” un espacio y el texto de la cita.

Código Markdown	Código HTML a generar
> y una cita dentro del texto	<blockquote> y una cita dentro del texto </blockquote>

- Las líneas separadoras horizontales se especifican con alguna de estas tres secuencias de caracteres: "\*\*\*" , "---" o "\_\_". Deben aparecer solas en la línea.

Código Markdown	Código HTML a generar
antes de la linea horizontal --- despues de la linea horizontal	antes de la linea horizontal <HR/> despues de la linea horizontal

- Para especificar bloques de código, donde normalmente se respetan las indentaciones incluidas en el texto y se usa una fuente de estilo código, se rodea el contenido con dos parejas de la siguiente secuencia de caracteres "~~~" en las líneas anterior y posterior al contenido.

Código Markdown
<pre> ~~~ System.out.print("Primera linea de codigo"); System.out.print("... mas codigo ..."); System.out.print("ultima linea de codigo"); ~~~ </pre>
Código HTML a generar
<pre> &lt;code&gt; &lt;pre&gt; System.out.print("Primera linea de codigo"); System.out.print("... mas codigo ..."); System.out.print("ultima linea de codigo"); &lt;/pre&gt; &lt;/code&gt; </pre>

### Nivel notable

Para llegar al notable (Nota = 8) hay que completar el aprobado junto con las siguientes características:

- Enlaces externos indicando el texto entre corchetes “[ ... ]” e inmediatamente después la URL entre paréntesis “(“ ... “)”. No es necesario comprobar que la URL tenga una sintaxis apropiada:

Código Markdown
<pre> [Esto deberia ser un enlace] (thphp://NO-es-una-URL_VALIDA) </pre>
Código HTML a generar
<pre> &lt;A HREF="thphp://NO-es-una-URL_VALIDA"&gt; Esto deberia ser un enlace(URL enlace aparentemente incorrecta) &lt;/A&gt; </pre>

- Permitir anidar una sola vez negritas y cursivas usando distinto carácter básico para los dos estilos:

Código Markdown
<pre> anidar <b><i>italica</i></b> en negrita anidar <b><i>*italica*</i></b> en negrita anidar <i><b>negrita</b></i> en italica anidar <i><b>*negrita*</b></i> en italica </pre>
Código HTML a generar
<pre> anidar &lt;SPAN class="bold"&gt;&lt;SPAN class="ital"&gt;italica&lt;/SPAN&gt; en negrita&lt;/SPAN&gt; anidar &lt;SPAN class="bold"&gt;&lt;SPAN class="ital"&gt;italica&lt;/SPAN&gt; en negrita&lt;/SPAN&gt; anidar &lt;SPAN class="ital"&gt;&lt;SPAN class="bold"&gt;negrita&lt;/SPAN&gt; en italica&lt;/SPAN&gt; anidar &lt;SPAN class="ital"&gt;&lt;SPAN class="bold"&gt;negrita&lt;/SPAN&gt; en italica&lt;/SPAN&gt; </pre>

## Nivel sobresaliente

Para llegar al sobresaliente (Nota = 10) hay que completar el notable junto con las siguientes características:

- Controlar que la URL de un hipervínculo tenga el formato apropiado. Es decir empezar por las cadenas “http://” o “https://” e ir seguidas por al menos dos ristas de letras y/o números siempre separadas por un punto (“.”). Si la URL no siguiera este formato se añadiría al final del texto del enlace el siguiente contenido “(URL enlace aparentemente incorrecta)”.

Código MarkDown
<pre>[Esto debería ser un enlace] (thphp://NO-es-una-URL_VALIDA) [Esto debería ser un enlace] (http://www.urjc.es)</pre>
Código HTML a generar
<pre>&lt;A HREF="thphp://NO-es-una-URL_VALIDA"&gt;Esto debería ser un enlace (URL aparentemente incorrecta)&lt;/A&gt; &lt;A HREF="http://www.urjc.es"&gt;Esto debería ser un enlace&lt;/A&gt;</pre>

- Listas no ordenadas, cuyos ítems comienzan por un guion y un espacio “- “ y termina al final de esa línea. La lista termina con una línea en blanco.

Código MarkDown	Código HTML a generar
<pre>- primer ítem - segundo ítem  - solo un ítem</pre>	<pre>&lt;UL&gt; &lt;LI&gt;primer ítem&lt;/LI&gt; &lt;LI&gt;segundo ítem&lt;/LI&gt; &lt;/UL&gt;  &lt;UL&gt;&lt;LI&gt;solo un ítem&lt;/LI&gt;&lt;/UL&gt;</pre>

- Los saltos de línea que deben aparecer en el código HTML deben ser los mismos que en el código MarkDown.

## Sobre la especificación a entregar

- Entregar un fichero jflex, con extensión “.flex”, no un fichero de texto con extensión “.txt” (**extensión por defecto** del Notepad de Windows)
- El **nombre de la clase** a generar será la concatenación de los dos **apellidos** sin acentos ni espacio en blanco.
- No se debe implementar en las acciones funcionalidades de reconocimiento que puedan ser proporcionadas por los patrones de JFlex.
- El código HTML se puede generar directamente por la salida estándar (imprimiendo en pantalla y redirigiendo la salida a un fichero) o en un fichero que se cree en el mismo directorio de la aplicación y cuyo nombre será **salida.html**.