



PRÁCTICA Nº 1: 1 sesión

(11, 12, 15, 23 y 24 de marzo de 2004)

TRADUCTOR DE CÓDIGO PSEUDOHTML

0. OBJETIVOS

Plantear una tarea de programación que fomente la capacidad de análisis de un problema, para a partir de un enunciado abstracto ser capaz de identificar las distintas fases que se deben abordar para alcanzar la solución.

1. INTRODUCCIÓN

El lenguaje HTML es un lenguaje de marcado de texto que es interpretado por los navegadores de internet y que permite la visualización de páginas web. Las marcas introducidas por este lenguaje indican fundamentalmente el formato con el que se va a ver la información (texto grande, texto pequeño, realzado, letra cursiva, etc...). Para crear una página web es preciso introducir en un archivo el texto que se desea visualizar junto con las marcas HTML para darle el formato deseado. Así por ejemplo, el siguiente texto incluye marcas HTML:

```
<html>

<head>
<title>Algoritmos y Estructuras de Datos</title>
</head>

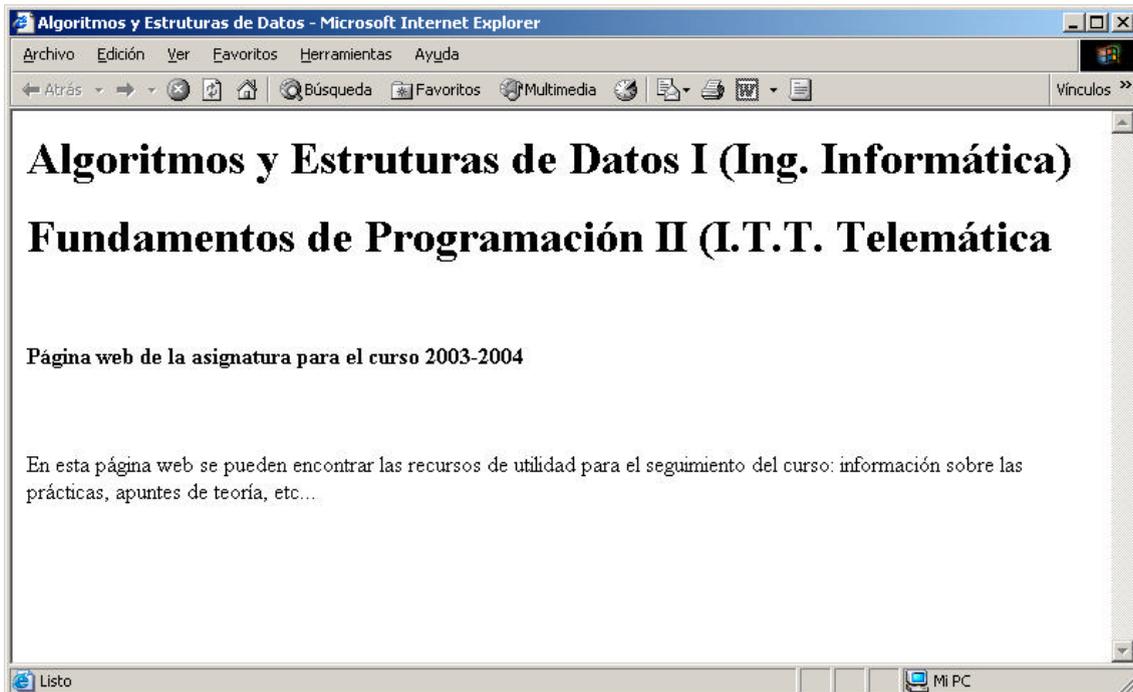
<body>
  <h1>Algoritmos y Estructuras de Datos I (Ing. Informática)</h1>
  <h1>Fundamentos de Programación II (I.T.T. Telemática)</h1>
  <p>&nbsp;</p>
  <p><b>Página web de la asignatura para el curso 2003-2004</b></b>
  <p>&nbsp;</p>
  <p>En esta página web se pueden encontrar los recursos de
  utilidad para el seguimiento del curso: información sobre las prácticas,
  apuntes de teoría, etc...</p>
  <p>&nbsp;</p>
</body>

</html>
```

La estructura básica del documento es: dos etiquetas `<html>...</html>` que engloban toda la información, y dentro de estas etiquetas dos bloques: `<head>...</head>` con la información de la cabecera, y `<body>...</body>` con la información del cuerpo del documento. Dentro del apartado 'body' se incluye la información que deseamos visualizar.



La visualización del documento en un navegador será:



Como se puede observar en el ejemplo, las etiquetas HTML se introducen en el texto identificadas por los símbolos “<” y “>”. Además, es preciso indicar el ámbito de aplicación de la etiqueta, es decir, donde comienza a tener efecto la etiqueta (<etiqueta>) y donde termina (</etiqueta>).

Las etiquetas del lenguaje HTML utilizan palabras inglesas o abreviaturas de ellas para identificar las tareas a realizar. Se puede considerar que esto hace especialmente complicado acordarse de la etiqueta exacta que hay que introducir en cada caso. Por eso, se ha pensado en la posibilidad de crear un vocabulario propio para identificar las etiquetas de formato que debemos introducir en las páginas web y que utilice como base el castellano o el valenciano, indistintamente. De manera que se pueda indicar <titulo> o <titol> en lugar <title>, o bien, <parrafo> o <paragraf> en lugar de <p>. Obviamente, un navegador estándar no entendería un documento escrito con estas nuevas etiquetas, pero también es evidente (?) que éstas son más fáciles de recordar. Para resolver el problema, se ha pensado en realizar un programa que traduzca cualquier archivo escrito con el nuevo conjunto de etiquetas a otro archivo en el que se utilicen exclusivamente las etiquetas HTML estándar. De manera, que se puedan escribir páginas web empleando el juego de etiquetas en castellano o en valenciano y después se aplique el programa traductor y se genere el archivo HTML que se debe visualizar en la web.

En esta primera práctica se pide escribir el programa traductor.

2. REALIZACIÓN DE LA PRÁCTICA

Se debe desarrollar un programa que lea un archivo de texto que especifique lo que será una página web, pero utilizando etiquetas pertenecientes al vocabulario propio (no HTML). Este programa debe proceder a generar un nuevo archivo de texto cuyo contenido sea igual al archivo de entrada pero donde se haya sustituido cada etiqueta de formato por la etiqueta equivalente en HTML estándar. Para ello, se debe saber que la tabla de equivalencias (conversión) entre los dos conjuntos de etiquetas se encuentra almacenada en el archivo “tabla.txt”, que se puede descargar de la página web de la asignatura. Además, como ayuda al proceso de traducción se proporciona una función que se puede utilizar para leer las etiquetas del archivo original,

```
string leerEtiqueta (ifstream & f);
```



Esta función puede ser llamada tras encontrar un símbolo '<' y devuelve la etiqueta completa (hasta el siguiente '>') o incompleta (si tras el '<' no hay una etiqueta válida: existen espacios en blanco en la etiqueta o se acaba el fichero).

3. ENTREGA DE PROGRAMAS

Al comienzo de la siguiente sesión de prácticas se entregará al profesor un fichero:

1) Fichero de traducción (pr01##.cpp)

dónde ## es el número asignado a la pareja (01, 02,...)

Nota Muy Importante

Antes de poder empezar a realizar cualquiera de las prácticas es necesario presentar las hojas de especificación de programas (documentación de programas) con las tareas que se van a realizar en la práctica, explicando brevemente como se van a solucionarse los problemas que se plantean.

ENTREGA DE PROGRAMAS:

Al comenzar la sesión 2 de prácticas (22, 25, 26, 30 y 31 de Marzo de 2004)