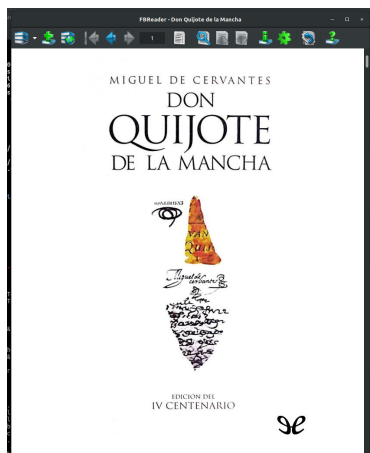


METODOLOGÍA DE LA PROGRAMACIÓN DGIIM - UNIVERSIDAD DE GRANADA Curso 2020-2021

PROCESANDO UN FICHERO EPUB



Un fichero EPUB está preparado para leerse en un lector de libros electrónicos¹ y necesitaríamos un programa como **fbreader** para poder abrirlo y leerlo. Este fichero tiene apariencia de binario, pero no es del todo cierto. Si cambiamos la extensión .epub por .zip veremos que el fichero se puede abrir en el descompresor y muestra su estructura interna. En particular, en la carpeta **/OEBPS/Text/** podemos encontrar una serie de ficheros .xhtml que son, en realidad, el contenido del libro. Por



ejemplo, si abrimos el fichero **sinopsis.xhtml** Veremos que se puede cargar en un navegador. Este examen de programación² lo vamos a realizar precisamente sobre esos ficheros .xhtml los cuales son ficheros de texto codificados en el conocido lenguaje html y son, los que, en último término, visualizan los lectores de EPUB en la pantalla.

```
<?xml version="1.0" encoding="utf-8" standalone="no"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.1//EN"
"http://www.w3.org/TR/xhtml11/DTD/xhtml11.dtd">

<html xmlns="http://www.w3.org/1999/xhtml">
<head>
  <title></title>
  <link href="../../Styles/style.css" rel="stylesheet" type="text/css" />
</head>

<body>
  <div class="sinopsis">
    <p class="salto10">Poco puede decirse del Quijote que no se haya dicho ya. Obra culmen de
la literatura universal y primera novela moderna, Cervantes ideó la historia de un hidalgo
aldeano que enloquece de tanto leer novelas de caballerías y, como buen caballero andante, sale
a los caminos con el noble afán de ayudar a los necesitados. Así, en compañía del afable y
crédulo Sancho Panza, don Quijote participa en una serie de delirantes aventuras que provocarán
la hilaridad del lector, ya que la vida que reproduce Cervantes es alegre, graciosa y
dramática, a la vez que sana y optimista, a pesar de las vicisitudes de su existencia. Sin
embargo, el Quijote es mucho más, pues constituye una lección magistral sobre la grandeza y la
miseria de la condición humana.</p>
```

¹ Cervantes Saavedra, Miguel de - Don Quijote de la Mancha (IV CENTENARIO) [4765] (r1.4).epub

² Descargar de Prado el proyecto de NetBeans "ExamenLaboratorio2021.zip" y descomprimirlo.

EJERCICIO DE EXAMEN

Para ello, vamos a extraer todos los ficheros xhtml en la carpeta “./data/”. Vamos a abrir uno de estos ficheros xhtml y vamos a buscar una cadena clave, por ejemplo “<p class=” que delimita un campo de html, descrito entre comillas <p class=“salto10”> . y vamos a seleccionar el contenido que hay entre el carácter de color verde y el de color rojo. Eso es lo que buscamos: salto10 Hay cuatro variantes diferentes de examen que se describen en la siguiente tabla

Variante	Patrón a buscar XXXXXXXX	EJEMPLO
<p class=“	< p c l a s s = “ X X X X X X X “ >	<p class=“salto10”> salto10
<a href=“	< a h r e f = “ X X X X X X X “ >	 ../Text/notas2.59.xhtml
<sup>	< s u p > X X X X X X X < / s u p >	^[2] [2]
<h3>	< h 3 > X X X X X X X < / h 3 >	<h3>CAPÍTULO LIX</h3> CAPÍTULO LIX
	0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19	

Esto lo repetimos línea por línea, buscando sólo la primera ocurrencia del patrón, si es que hubiese alguna, y nos da una secuencia de cadenas que deberemos guardar en otro fichero de texto “./data/salida.dat”. Ya está. Ahora paso a paso

- Consulta el fichero “./data/valida.dat” y busca tu nombre. A continuación aparece qué fichero tienes que abrir y qué variante debes implementar. Introducir tu nombre (tal y como aparece en este fichero) en la llamada a Introduccion(“nombre completo”). Introducir el nombre del fichero de trabajo en la variable namecapitulo.
- El programa debe hacer lo siguiente.
 - Abrir el fichero “./data/index.dat”. En cada línea hay un número entero X y una cadena S. Si la cadena S coincide con namecapitulo (el fichero de trabajo), entonces X es el número de líneas del fichero de trabajo. Cerrar el fichero.
 - Abrir el fichero de trabajo, cada uno el que tenga asignado. Reservar memoria para el vector *capitulo y leer X líneas en el vector (usar getline para leer líneas completas).
 - Aplicar la variante que te ha tocado
 - Busca la variante y calcula la posición del corte verde.
 - Sigue buscando en la cadena hasta encontrar el carácter del corte rojo.
 - Recorta la cadena entre el corte verde y el rojo.
 - Generar una secuencia de cadenas resultante y almacenarlas en el vector *seleccion.
 - Guardar el vector *seleccion en el fichero “./data/salida.dat”. En el fichero “./data/valida.dat”, junto a tu nombre, aparece la secuencia de cadenas que te debe salir de resultado.
- Cuando termines, haz el zip del proyecto como siempre y lo subes a Prado

int string::find(const string &busca); Devuelve string::npos si no está. En caso contrario, devuelve la posición de la primera ocurrencia	string s=“BLABLA^{HOLA}BLA”; x = s.find(“<sup>”) → 6 y= s.find(“BLO”) → string::npos
string string::substr(int pos, int largo); Devuelve el corte de la cadena que empieza en pos y tiene el largo especificado	string s=“BLABLA^{HOLA}BLA”; x = s.substr(11,4) → “HOLA”