

Estructuras de Datos

Guía 6: XML y XPath

1. Tenemos el siguiente documento XML:

```
<Pedidos>
  <Pedido>
    <Cliente>Ana</Cliente>
    <PC>pc500</PC>
  </Pedido>
  <Pedido>
    <Cliente>Juan</Cliente>
    <PC>pc600</PC>
  </Pedido>
</Pedidos>
```

- a. Dibujar el árbol del documento.
- b. Calcular las respuestas a las siguientes consultas XPath:
 - i. `/Pedidos/Pedido/PC`
 - ii. `//Cliente//*`
 - iii. `/Pedidos/Pedido[Cliente='Ana']`
 - iv. `//Pedido[Cliente and PC]/PC`
 - v. `b.5 //*//Pedido[Cliente='Ana']/Cliente`
 - vi. `b.6 //Pedido[Cliente='Juan'][PC != 'pc600']`
 - vii. `b.7 //Pedido[PC='pc600']//`

2. El archivo `courses-noID - castellano.xml` tiene una lista de algunas materias de varios departamentos de la Universidad de Stanford. Con ese archivo:

- a. Encontrar todos los nodos Título (tanto de departamentos como de materias).
- b. Encontrar los apellidos de todos los directores de departamento.
- c. Listar todas las materias con más de 500 vacantes.
- d. Listar los nombres (Títulos) de departamentos que tienen una materia con correlativa "CS106B".
- e. Apellidos de todos los docentes que firman con segundo nombre.
- f. Apellidos de todos los docentes que tienen materias con vacantes mayores o iguales a 100.

3. El archivo shaks200.zip contiene varias obras de Shakespeare en formato .xml. Dentro del .zip se encuentra el archivo playlist con la lista de obras contenidas y play.dtd con la definición del esquema .xml.
 - a. Analizar el archivo dtd y graficar en forma de árbol la estructura de una obra. Recordar que en la definición * significa que el bloque se repite 0 o más veces, + que se repite 1 o más veces, ?: 0 o 1 vez y #PCDATA se refiere a texto
 - b. Escribir una programa que recorra todas las obras disponibles (leyendo el archivo playlist) y liste todos los grupos de personas que participan en alguna obra con el siguiente formato:

Título de la Obra

 Descripción del grupo

 Persona

 Persona
 - c. Escribir un programa que lea todos los personajes de todas las obras y arme una especie de índice invertido indicando los títulos de las obras en que participa cada personaje. (Recorrer el archivo playlist para obtener el listado de obras disponibles, para que si en el futuro se agregan o quitan obras o si se modifican los nombres de los archivos, no queden los nombres referenciados en el código). Listar todos los personajes ordenados de mayor a menor por la cantidad de obras en las que aparecen con el siguiente formato (si varios personajes aparecen en la misma cantidad de obras listarlos ordenados alfabéticamente):

Persona

 Título de la obra

 ...

 Título de la obra
 - d. Escribir un programa y calcular quien dice más líneas en la obra Romeo y Julieta, Romeo o Julieta

Se recomienda usar la librería lxml (<http://lxml.de/>). Se adjunta un fragmento de código con ejemplos de uso.