



Esta obra está bajo una Licencia Creative Commons Atribución 4.0



# INTRODUCCIÓN A LAS COLECCIONES DE DATOS (V 1.1)

Prof. José Fager  
Montevideo enero de 2020

*fic*

IPAD II

A horizontal bar at the top of the slide, divided into a red section on the left and a teal section on the right.

# Colecciones de Datos

# ¿Qué es una colección de datos?

- ▣ Una "**colección de datos**" (CD) es un conjunto de datos que: responden a una "**definición**", con una "**estructura**" determinada, una serie de "**restricciones**", que pueden ser "**persistidos**" y sobre los cuales se pueden ejecutar "**consultas**".

# ¿Cómo está formada una CD?

- ▣ La “**definición**” de una CD es un texto escrito en lenguaje natural, que describe la realidad que se intenta modelar.
- ▣ La “**estructura**” de una CD, es un artefacto que representa la forma en que se relacionan, los distintos elementos que integran la estructura. Se representa con un diagrama con forma de árbol.
- ▣ Las “**restricciones**” de una CD, son una serie reglas que la estructura de la CD debe cumplir. Se explicitan en un archivo de texto llamado archivo de restricciones escrito con sintaxis DTD, la extensión usada para estos archivos es “.dtd”.
- ▣ La “**persistencia**” de una CD, son los datos propiamente dichos que posee la CD, son contenidos en un archivo de texto usando un lenguaje formal, la extensión usada para estos archivos es “.xml”.
- ▣ Las “**consultas**” se pueden ejecutar sobre una instancia de la persistencia de la CD, usando un lenguaje formal llamado Xquery.

## Ejemplo: CD "discoteca"

Por medio de un ejemplo sencillo se intentará introducir los conceptos más relevantes sobre colecciones de datos.

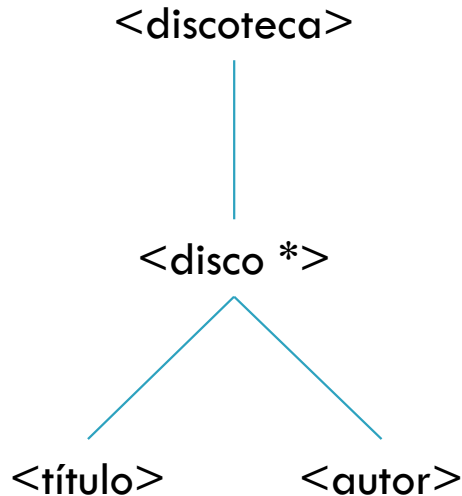
# Definición de la CD “discoteca”

Se quiere modelar una discoteca que contiene discos.

De cada disco: se sabe su título y se sabe el autor.

# Estructura de la CD “discoteca”

## <<DIAGRAMA DE LA ESTRUCTURA>>



- Los elementos de la estructura son: discoteca, disco, título y autor. Siempre se escriben en singular aunque en la definición estén en plural.
- El nombre de la CD siempre es la raíz del árbol, en este caso "discoteca".
- La raíz del árbol siempre tiene un y solo un hijo que es el elemento que se quiere representar, en este caso "disco".
- El símbolo "\*" en el elemento que se quiere representar (disco) indica la cardinalidad. En este caso significa que: "la discoteca puede tener ninguno, uno o varios discos".

# Restricciones de la CD “discoteca”

<<ARCHIVO DE RESTRICCIONES >>

<!ELEMENT discoteca (disco\*)>

<!ELEMENT disco (titulo, autor)>

<!ELEMENT titulo (#PCDATA)>

<!ELEMENT autor (#PCDATA)>

- Se restringen los nombre de los elementos que pueden estar en la CD.
- Se indica la relación taxonómica entre sus elementos.
- Se especifica la cardinalidad de cada elemento.
- Se usa la sintaxis DTD (Document Type Definition) para escribir las restricciones en un archivo que llamamos “archivo de restricciones”.
- La extensión usada para el archivo de restricciones es “.dtd”.



# Persistencia de la CD “discoteca” (I)

<<ARCHIVO DE PERSISTENCIA>>

<discoteca>

<disco>

<titulo>Let it be</titulo>

<autor>The Beatles</autor>

</disco>

<disco>

<titulo>Let it bleed</titulo>

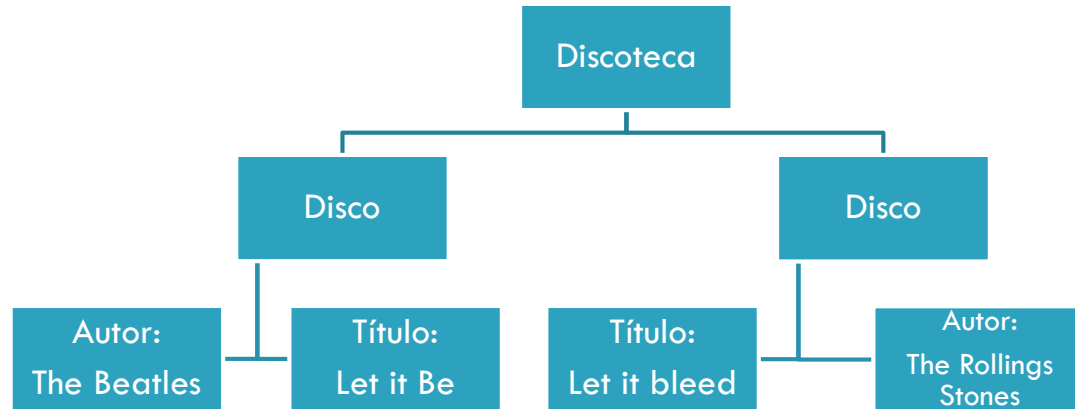
<autor>The Rolling Stones</autor>

</disco>

</discoteca>

Usa la  
extensión  
.xml

Realidad que  
modela el archivo de  
persistencia



# Persistencia de la CD “discoteca” (II)

<discoteca>

<disco>

<titulo>Let it be</titulo>

<autor>The Beatles</autor>

</disco>

<disco>

<titulo>Let it bleed</titulo>

<autor>The Rolling Stones</autor>

</disco>

</discoteca>

- Contiene los datos de instancia puntual de la CD.
- Se usa un lenguaje formal para escribir la persistencia en un archivo que llamamos “archivo de persistencia”.
- En el ejemplo se muestra una instancia de la CD “discoteca” que tiene dos discos (dos registros).
- El archivo de persistencia lo identificamos con la extensión “.xml”.

# Consultas sobre la CD “discoteca”

<<CONSULTA>>

```
for $r in doc('discoteca.xml') //disco/autor
return $r
```

<<SALIDA>>

```
<autor>The Beatles </autor>
<autor>The Rollings Stones</autor>
```

- Se debe tener una instancia puntual de la CD (una persistencia).
- Se usa un lenguaje formal para escribir la consulta llamado Xquery.
- Luego de ejecutar la consulta se obtiene un conjunto resultado.
- En el ejemplo se muestra una consulta que retorna todos los autores de los discos de la CD “discoteca”.