

TEMA

Técnicas de Representación y Organización en Internet



UNIVERSIDAD CARLOS III DE MADRID



Sistemas Avanzados de Recuperación de Información (SARI)
2008-2009

Jorge Morato Lara
Sonia Sánchez-Cuadrado



UNIVERSIDAD CARLOS III DE MADRID



Índice

1. Sistemas de Gestión de Contenidos
2. Representación de información estructurada
 1. XML
 2. METADATOS
 3. RDF y OWL



UNIVERSIDAD CARLOS III DE MADRID





CMS (Content Management Systems)

- ❑ Gestores de contenidos
<http://www.cmsenespanol.com>
- ❑ En sentido general se trata de portales que gestionan una colección de recursos informativos.
- ❑ Bajo esta denominación podría entrar la Wikipedia, Blogs, foros, etc.

Actualmente destacan unas pocas aplicaciones para este apartado:

- ❑ Wikipedia: Dokuwiki, MediaWiki
- ❑ Portales: Joomla/Mambo
- ❑ Moodle





UNIVERSIDAD CARLOS III DE MADRID





CMS (Content Management Systems)

Tipología	Ejemplos
Genéricos	Joomla, Plone, MODX, OpenCMS, Alfresco?
Específicos para ONG`s	Common Knowledge Content Server, Content Management
Foros	phpBB, SMF, MyBB
Blogs	WordPress, Movable Type, Drupal
Wikis	MediaWiki, TikiWiki
eCommerce	osCommerce
Portal	PHP-Nuke, GXPortal, Joomla MS SharePoint
Galeria	Gallery, Dragonfly CMS
E-learning	Moodle, Blackboard, eCollege, SumTotal
Publicaciones digitales	ePrints, SamdhaNews, Thinkindot CMS



UNIVERSIDAD CARLOS III DE MADRID





Índice

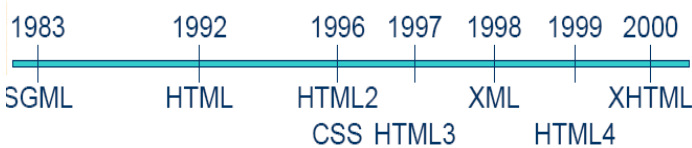
1. Sistemas de Gestión de Contenidos
2. Representación de información estructurada
 1. XML
 2. METADATOS
 3. RDF y OWL



UNIVERSIDAD CARLOS III DE MADRID



Introducción Histórica



Timeline of web technologies:

Year	Technology
1983	SGML
1992	HTML
1996	HTML2, CSS
1997	HTML3
1998	XML
1999	HTML4
2000	XHTML

Extensible Markup Language metalenguaje extensible de etiquetas de marcado desarrollado por W3C.

- Su predecesor es SGML y se diferencia por
 - por su **sencillez**
- Al igual que HTML, es portable a cualquier plataforma
 - Se diferencia de HTML por su flexibilidad (el número de etiquetas que puede incluir un documento XML es ilimitado)



UNIVERSIDAD CARLOS III DE MADRID

Objetivo de XML

- ❑ Directamente utilizable en Internet
- ❑ Soporte para una amplia variedad de aplicaciones para transferencia de datos
- ❑ Compatible con SGML
- ❑ Posibilidad de crear sencillos procesadores de XML
- ❑ Documentos XML legibles y medianamente claros (depende de la definición)
- ❑ Diseño rápido del lenguaje
- ❑ Simple, pero perfectamente formalizado
- ❑ Documentos XML fáciles de crear



UNIVERSIDAD CARLOS III DE MADRID



HTML vs. XML

Características HTML

- ❑ HTML carece de un chequeo sintáctico. Páginas con errores son mostradas en los navegadores
- ❑ HTML carece de estructura rígida
- ❑ HTML no es orientado a objetos
- ❑ HTML mezcla contenido y representación
- ❑ Por todo esto:
 - HTML no puede ser fácilmente leído por una máquina
 - HTML nunca será un estándar de intercambio de datos

```
<HTML>
<HEAD><TITLE>Libros de mi
infancia</TITLE></HEAD>
<BODY>
<P><i><B>Don Quijote de la
Mancha</B></i>
<P><i><B>Miguel de Cervantes</B></i>
<HR>
<P><i><B>La vida es sueño</B></i>
<P><i><B>Calderón de la Barca</B></i>
<HR>
</BODY>
</HTML>
```



UNIVERSIDAD CARLOS III DE MADRID



Características de XML

- ❑ Es un subconjunto del lenguaje SGML
- ❑ Al igual que él, se utiliza para representar datos de forma **estructurada** (Jerárquica)
- ❑ Se basa en una **gramática de obligado cumplimiento**. Esto facilita el desarrollo de *parsers* y, por lo tanto, su utilización masiva
- ❑ La estructura interna de un documento XML puede reflejarse en otro documento denominado **DTD** (*Document Type Definition*)
- ❑ A diferencia de HTML, separa radicalmente la **semántica** del documento, de su representación gráfica



UNIVERSIDAD CARLOS III DE MADRID



Documento HTML vs. Documento XML

```
<HTML>
<HEAD><TITLE>Libros de mi
infancia</TITLE></HEAD>
<BODY>
<P><B>Don Quijote de la
Mancha</B>
<P></>Miguel de Cervantes</>
<HR>
<P><B>La vida es sueño</B>
<P></>Calderón de la Barca</>
<HR>
</BODY>
</HTML>
```



```
<?xml version="1.0" encoding="utf-8"
standalone="yes" ?>
<Libros>
  <Libro>
    <Titulo>Don Quijote de la
Mancha</Titulo>
    <Autor>Miguel de
Cervantes</Autor>
  </Libro>
  <Libro>
    <Titulo>La vida es sueño</Titulo>
    <Autor>Calderón de la
Barca</Autor>
  </Libro>
</Libros>
```



UNIVERSIDAD CARLOS III DE MADRID



Documento HTML vs. Documento XML

En apariencia, el documento **HTML** anterior es correcto, sin embargo:

- Existen etiquetas que nunca se cierran: <P>
- Algunas etiquetas no están bien anidadas: el primer <I> nunca se cierra
- Para un lector no humano, no se sabe qué es un libro y qué es un autor

Reglas Generales de XML

- Un **único** elemento raíz
- Todo elemento debe tener etiquetas de **apertura y cierre**
- Distinción entre mayúsculas/minúsculas
- **Anidamiento** perfecto entre elementos
- Los **valores** de **atributos** siempre van entre **comillas**
- Los espacios en blanco se conservan



UNIVERSIDAD CARLOS III DE MADRID

Documentos Bien Formados y Válidos

Bien formado	Válido
■ Cumple con todas las reglas anteriormente expuestas ■ Contiene uno o más elementos ■ Hay un único elemento raíz (elemento documento) ■ Si el documento consta de más de una parte, todas están bien formadas ■ No se encuentran caracteres prohibidos en el texto	■ bien formado ■ cumple con las especificaciones semánticas expuestas en su plantilla (DTD o XML Schema)



UNIVERSIDAD CARLOS III DE MADRID

Componentes para XML



Comentarios:

- `<!-- Esto es un comentario, y no puedo incluir un doble guión-->`

Instrucciones de procesamiento:

- `<? Instrucción ?>`
- La instrucción no puede incluir los caracteres `?>`

Secciones CDATA:

- `<![CDATA[Este texto no será tratado, puede incluir "cualquier" &carácter < >]]>`
- No son tratadas por el *parser*
- Pueden incluir cualquier carácter prohibido (`"`, `'`, `&`, `>`, `<`).
- No puede incluir la cadena `]]>`



UNIVERSIDAD CARLOS III DE MADRID



Componentes para XML



Prólogo:

```
<?xml version="1.0" encoding="utf-8" standalone="yes"
?><?xml version="1.0" encoding="ISO-8859-1" ?>
```

- Es una instrucción de procesamiento obligatoria
- **Version:** indica la versión de XML que se está utilizando (1.0 en la actualidad). Es obligatoria
- **Encoding:** indica cómo se codificó el documento, y no es obligatoria (por defecto UTF-8). Válido para otros juegos de caracteres
- **Standalone:** "yes" indica que el documento es independiente; "no" va acompañado de DTDs externas



UNIVERSIDAD CARLOS III DE MADRID



Componentes para XML



□ DTD:

```
DOCTYPE: <!DOCTYPE MiDTD SYSTEM "C:\MiDTD.dtd">
```

- Indica la referencia (URI) a la DTD, así como el nombre (MiDTD) del elemento raíz de la misma
- La DTD podría ir incorporada en el propio documento XML, sin requerir otro fichero aparte
- El documento XML deberá cumplir con el contenido del DTD



UNIVERSIDAD CARLOS III DE MADRID



Componentes para XML



□ Etiquetas:

- Deben ir correctamente anidadas: apertura y cierre
- Etiqueta de apertura: comienza por <, más el nombre de la etiqueta y terminan por > Ejemplo <Libro>
- Etiqueta de cierre: </Libro>
- Etiqueta vacía: <Libro />
- No puede iniciar el nombre con ".", ":", "-", números
- Después de la primera letra pueden colocarse ".", números, "-"
- El nombre debe comenzar por una letra o un "_"
- No puede comenzar por "xml"



UNIVERSIDAD CARLOS III DE MADRID



Componentes para XML: elementos y atributos



Elemento:

- ❑ Es el conjunto de la etiqueta (marcador) de apertura, su contenido y la de cierre
- ❑ Por ejemplo: <Libro>Don Quijote de la Mancha</Libro>
- ❑ Hay algunos caracteres reservados (prohibidos):
 - Signo de mayor: >
 - Signo de menor: <
 - Ampersand: &
 - Apóstrofe: '
 - Comilla: "
- ❑ Estos caracteres prohibidos se reemplazan por entidades o se incluyen en secciones CDATA

Atributos:

- Cada elemento puede contener 0 ó más atributos
- Su valor debe ir siempre entrecomillado
- Sólo pueden aparecer en etiquetas de apertura o vacías
- El mismo atributo no puede aparecer repetido en la misma etiqueta
- Si el documento incluye DTD, cada atributo debe estar definido como atributo del presente elemento
- No puede contener ninguna referencia a entidad externa
- Son siempre tratados como cadenas de texto



UNIVERSIDAD CARLOS III DE MADRID



Componentes para XML



```
- <Libro>
  <Titulo>Don Quijote de la Mancha</Titulo>
  <Autor>Miguel de Cervantes</Autor>
  <Precio>1.123 euros</Precio>
  <Editorial> Santillana </Editorial>
</Libro>
```

(Sin atributos)

```
- <Libro Precio="1.123 euros" Editorial="Santillana">
  <Titulo>Don Quijote de la Mancha</Titulo>
  <Autor>Miguel de Cervantes</Autor>
</Libro>
```

(Dos elementos son atributos)



UNIVERSIDAD CARLOS III DE MADRID







Universidad
Carlos III de Madrid
www.uc3m.es

Sonia Sánchez Cuadrado
Assistant Professor
Computer Science Department

Superior Polytechnic School • Av. de la Universidad, 30
28911 Leganés • Madrid • Spain
Location: 2.1B.27 • Phone: +34 91 624 88 46 • Fax: +34 91 624 91 29
sonia.sanchez.cuadrado@uc3m.es



Papel ecológico



UNIVERSIDAD CARLOS III DE MADRID



Componentes para XML: DTDs

```
<!DOCTYPE Libros SYSTEM "Libros1.dtd">
<Libros>
  <Libro>
    <Titulo>Don Quijote de la Mancha</Titulo>
    <Autor>Miguel de Cervantes</Autor>
  </Libro>
  <Libro>
    <Titulo>La vida es sueño</Titulo>
    <Autor>Calderon de la Barca</Autor>
  </Libro>
</Libros>
```

```
view-source: - Código fuente de:file:///C:/Documents%20and%
Archivo  Editar  Ver
<!DOCTYPE Libros [
  <!ELEMENT Libros (Libro)+>
  <!ELEMENT Libro (Titulo, Autor)>
  <!ELEMENT Titulo (#PCDATA)>
  <!ELEMENT Autor (#PCDATA)>
]>
<Libros>
  <Libro>
    <Titulo>Don Quijote de la Mancha</Titulo>
    <Autor>Miguel de Cervantes</Autor>
  </Libro>
  <Libro>
    <Titulo>La vida es sueño</Titulo>
    <Autor>Calderon de la Barca</Autor>
  </Libro>
</Libros>
```



UNIVERSIDAD CARLOS III DE MADRID



DTDs

- Toda DTD debe tener uno y sólo un elemento raíz (también conocido como elemento documento)
- Este documento raíz debe coincidir con el nombre que aparece a continuación del DOCTYPE
- Un documento DTD puede contener:
 - Declaraciones de elementos
 - Declaraciones de atributos para un elemento
 - Declaraciones de entidades
 - Declaraciones de notaciones
 - Instrucciones de procesamiento
 - Comentarios
 - Referencias a entidades de parámetro

Ejemplo

```
<!DOCTYPE Libros [  
<!ELEMENT Libros (Libro)+>  
<!ELEMENT Libro (Titulo, Autor)>  
<!ELEMENT Titulo (#PCDATA)>  
<!ELEMENT Autor (#PCDATA)>  
>
```

A partir del elemento raíz, pueden opcionalmente depender (de forma jerárquica) otros elementos



UNIVERSIDAD CARLOS III DE MADRID



DTDs: (Contenido de los Elementos)

Contenido de un elemento:

- **EMPTY**: el elemento está vacío (puede contener atributos).
`<!ELEMENT IMAGEN EMPTY>`
- **ANY**: el elemento puede contener a cualquier otro elemento o incluso contenido textual.
`<!ELEMENT IMAGEN ANY>`
- **Otros elementos**: un elemento puede contener uno o más elementos hijos en una cierta secuencia (Ej. Libro)
- **#PCDATA**: texto *parseado*.
`<!ELEMENT LIBRO (#PCDATA)>`
- **Mixto**: el elemento puede incluir secuencias de caracteres opcionalmente mezcladas con elementos hijos.
`<!ELEMENT LIBRO (#PCDATA | AUTOR)*>`

```
<!ELEMENT LIBRO (Autor, Editorial)>  
<!ELEMENT Autor (#PCDATA)>  
<!ELEMENT PELICULA (Actor|Actriz|Director)+>  
<!ELEMENT PELICULA ((Actor | Actriz)*, Director, Maquillaje?)>  
<!ELEMENT PELICULA (#PCDATA | Actor)*>  
<!ELEMENT PELICULA (Titulo, Genero, (Actor | Actriz | Narrador)*)>  
<!ELEMENT FICHA (Nombre+, Apellido+, Direccion*, foto?,  
TelFijo*|TelMovil*)>
```



UNIVERSIDAD CARLOS III DE MADRID



- Secuencia:

-
- Diagram illustrating XML code snippets and their annotations:
- Snippet 1:**

```
<ELEMENT chiste
  (basilio+, antonio, aplauso?)>
```

 - Annotation: **1 o más veces** (pointing to `chiste`)
 - Annotation: **secuencia** (pointing to the content model)
 - Annotation: **0 o 1 veces** (pointing to `apluso?`)
 - Annotation: **0 o más veces** (pointing to `apluso?`)
 - Snippet 2:**

```
<ELEMENT basilio (#PCDATA | quote)*>
<ELEMENT antonio (#PCDATA | quote)*>
<ELEMENT quote (#PCDATA)*
  <ELEMENT aplauso EMPTY>
```

 - Annotation: **alternativa** (pointing to the content model in the `quote` element)
 - Snippet 3:**

```
<ATTLIST chiste
  name ID #REQUIRED
  label CDATA #IMPLIED
  status (funny|notfunny) 'funny'>
```

 - Annotation: **Valor por defecto** (pointing to the default value `'funny'`)



- <!ATTLIST Elemento Atributo Tipo Modificador>**

- Tipo de un atributo:**

-

DTDs (Modificadores de Atributos)



Modificadores:

- **#REQUIRED**: este atributo debe ser obligatoriamente introducido.
`<!ATTLIST Pelicula Titulo CDATA #REQUIRED>`
- **#IMPLIED**: indica que el atributo es *opcional*
- **ValorPredeterminado**: si se omitiese el atributo, los procesadores recogerían este valor por defecto
`<!ATTLIST Pelicula Genero (Ficcion | Terror | Humor) "Humor">`
`<!ATTLIST Autor Nacionalidad CDATA "Española">`
- **#FIXED**: se incluya o no se incluya el atributo, los procesadores siempre obtendrán este mismo valor
`<!ATTLIST Autor Nacionalidad CDATA #FIXED "Espanola">`



UNIVERSIDAD CARLOS III DE MADRID



DTDs (Recomendaciones para modelado de Atributos)



- Frecuentemente un mismo objeto se puede diseñar como un atributo o un elemento sin pérdida de semántica pero existen criterios para decantar...
- **Atributos**
 - Normalmente se trata de objetos cuya existencia no tiene sentido fuera del objeto al que describen (p.e. adjetivos), metadatos, identificadores únicos, el idioma, ...
 - En general, todo aquello por lo que existe mayor interés en filtrarlo que en mostrarlo

Ventajas

- Más fácil de procesar por el software (mayor eficiencia)
- Más legible, los atributos están próximos al elemento al que pertenecen



UNIVERSIDAD CARLOS III DE MADRID



DTDs

Recomendaciones para modelado de Elementos

- ❑ Elementos, se debería optar por ellos...
 - Siempre que se quiera definir sub-elementos (ya sean partes del elemento principal o sus específicos)
 - Permiten crear vínculos
 - Siempre que queramos repetir el mismo elemento con distinto valor (los atributos tienen un único valor como máximo)
 - Siempre que el contenido sea mayor que una palabra (oraciones, párrafos, ...) y sobre todo si se quiere mostrar el texto en cuestión
 - Tienen entidad propia independientemente del resto de elementos
- Los documentos que priman a los atributos pueden ser más breves al no tener una etiqueta propia que abrir y cerrar



UNIVERSIDAD CARLOS III DE MADRID



DTDs (Problemas)

- ❑ Una DTD no sigue el formato de un documento XML estándar. Esto representa un problema para los *parsers*
- ❑ No se soportan distintos tipos de datos al estilo de los lenguajes de programación (CDATA, #PCDATA)
- ❑ No se pueden crear tipos de datos personalizados
- ❑ No se soportan los espacios de nombres (namespaces)
- ❑ El número de ocurrencias no se puede controlar al 100% (Ej. 2 ocurrencias)



Por estas y otras razones, surgen los XML Schemas (Esquemas)



UNIVERSIDAD CARLOS III DE MADRID



Referencias a Caracteres

- Permiten incluir cualquier carácter dentro de un documento XML
 - Basado en el conjunto de caracteres ISO/IEC 10646
(<http://xml.coverpages.org/xml-ISOents.txt>)
 - Dos formatos:
 - `&#xvalor`; valor representado en decimal
 - `&#xvalor`; valor representado en hexadecimal
- ```
<?xml version="1.0" encoding="utf-8" standalone="yes" ?>
<Libros>
 <Libro Precio = "658 euros" Editorial = "Anaya">
 <Titulo>La vida es sueño</Titulo> (DECIMAL)
 <Autor>Calderón de la Barca</Autor> (DECIMAL)
 </Libro>
</Libros>
```



UNIVERSIDAD CARLOS III DE MADRID



## Entidades

Las entidades permiten:

- Dar modularidad al texto evitando tener que escribir algo de forma repetitiva (Re-uso)
- Incluir caracteres prohibidos `&`, `>`, `<`, `"`, `'`
- Incluir caracteres de otros idiomas: eñe, ...

Comienzan por `&` y terminan en `;`,  
Como por ejemplo: `&amp;`;

### Entidades predefinidas:

- Signo menor `lt` < `&lt;`;
- Signo mayor `gt` > `&gt;`;
- Ampersand `amp` & `&amp;`;
- Apóstrofe `apos` ' `&apos;`;
- Comilla doble `quot` " `&quot;`;

### Tipos de entidades:

- General y de Parámetro:
  - **General**: contiene texto XML u otros caracteres
  - **De Parámetro**: contiene texto XML que puede insertarse dentro de una DTD
- Interna y Externa:
  - **Interna**: contiene el texto dentro de una cadena entrecomillada
  - **Externa**: hace referencia a un archivo externo
- Analizada y no Analizada:
  - **Analizada**: texto XML que será *parseado* en su punto de inserción
  - **No Analizada**: no será *parseada*



UNIVERSIDAD CARLOS III DE MADRID



## Ejercicio de Atributos:

Hacer una DTD utilizando atributos:

```
<?xml version="1.0" encoding="utf-8" ?>
<Agenda>
 <Persona>
 <Nombre> Anabel </Nombre>
 <Apellido> Fraga </Apellido>
 <Sexo> Femenino </Sexo>
 <DNI> 44444444-O </DNI>
 <Nacionalidad> Española </Nacionalidad>
 <Email> anabel_fraga@mydomain.es </Email>
 <Email> anabel_fraga@fraga.es </Email>
 <Oficina>
 <Direccion CP="28911"> Av. Universidad 30 </Direccion>
 <Despacho> 2.1 B18 </Despacho>
 <Email_O> afraga@te.inf.uc3m.es </Email_O>
 <Telefono_01> 5555555 </Telefono_01>
 <Telefono_02> 5555556 </Telefono_02>
 </Oficina>
 <Telefono_1> 5555558 </Telefono_1>
 <Telefono_2> 5555559 </Telefono_2>
 <Movil> 5557777 </Movil>
 </Persona>
</Agenda>
```



UNIVERSIDAD CARLOS III DE MADRID



## XML Schemas

- Actualmente existe una nueva recomendación de W3C de Mayo 2001 para definiciones de XML: XML Schemas
- Uso de notación XML para definiciones
- Limitación de uso: Actualmente existe una gran cantidad de documentos definidos con DTDs.

```
<element name="Point">
 <complexType>
 <sequence>
 <element name="x"
 type="integer"/>
 <element name="y"
 type="integer"/>
 </sequence>
 </complexType>
</element>
```

point.xsd

```
<Point>
 <x>3</x>
 <y>4</y>
</Point>
```

point.xml



UNIVERSIDAD CARLOS III DE MADRID





## XML Schemas vs. DTDs (I)



### Desventajas de las DTDs

- No escritas en sintaxis XML
- Poco uso de namespaces
- Pocos tipos de datos (y lo que es peor, no se pueden definir nuevos tipos de datos)

### Ventajas de las XML Schemas:

- Muchas herramientas que lo soportan
- Existen muchos documentos: DTDs y XMLs basados en ellas
- Fácil de aprender
- Permite multitud de tipos de datos (pe xs:date, xs:int, xs:language, ...)
- Amplio uso de los namespaces
- Permite agrupar elementos para su reutilización, permite herencia (Ejemplo: Datos Personales en distintos Dominios de uso)



UNIVERSIDAD CARLOS III DE MADRID



## La Familia XML



- **XPointer/XLink**: permiten referenciar a diferentes recursos, dentro o fuera del documento XML
- **XPath**: lenguaje de consulta para recorrer ficheros XML
- **XQL** (XML Query Language): útil para localizar y extraer elementos de un documento XML
- **XIRQL**: Una extensión de XQL para Recuperación de Información
- **XSLT**: Lenguaje para transformación de documentos XML
- **XSL-FO**: Expresa semántica de formateado de documentos, provee los medios para producir impresiones de alta calidad.



UNIVERSIDAD CARLOS III DE MADRID





## Índice

1. Sistemas de Gestión de Contenidos
2. Representación de información estructurada
  1. XML
  2. METADATOS
  3. RDF y OWL



UNIVERSIDAD CARLOS III DE MADRID



## Metadatos

**Metadatos** son datos que describen otros datos. En general, un grupo de metadatos se refiere a un grupo de datos, llamado *recurso*.

Para campos de la informática, como la recuperación de información o la web semántica, los metadatos en etiquetas son un enfoque importante para construir un puente sobre el intervalo semántico.



UNIVERSIDAD CARLOS III DE MADRID



## Etiqueta META en HTML

La etiqueta META se utiliza dentro del encabezamiento HEAD de una página HTML, para identificar, indizar y catalogar documentos

Los atributos de esta etiqueta se encuentran indicados en el RFC 1866 bajo la siguiente DTD (*Document Type Definition*):

```
<!ELEMENT META>
<!ATTLIST META
 http-equiv NAME #IMPLIED
 name NAME #IMPLIED
 content CDATA #REQUIRED >
```



UNIVERSIDAD CARLOS III DE MADRID



```
Código fuente de:file:///C:/Documents%20and%20Settings/sonieta/Escritorio/prueba.html - Mozilla Firefox
Archivo Editar Ver Ayuda

<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="es" lang="es">
<head>
 <meta http-equiv="Content-Type" content="text/html; charset=ISO-8859-1"/>
 <meta name="description" content="Página Web desarrollada para la asignatura Sistemas avanzados de recuperación
información, contiene información sobre metadatos como XML, Dublin Core, RDF, etc."/>
 <meta name="Keywords" content="Sistemas avanzados de recuperación de información, SARI, XML, Metadatos, etiquet
Meta, RDF, recuperacion de informacion, Dublin Core, Dublincore, DTD, SGML, HTML, web semántica"/>

 <meta http-equiv="content-type" content="text/html"/>
 <meta name="DC.Title" content="Sistemas avanzados de recuperación de información"/>
 <meta name="DC.Creator" content="Juan Ramón Pérez González"/>
 <meta name="DC.Subject" content="Sistemas avanzados de recuperación de información, SARI, XML, Metadatos, etiquet
Meta, RDF, recuperacion de informacion, Dublin Core, Dublincore, DTD, SGML, HTML"/>
 <meta name="DC.Description" content="Página Web desarrollada para la asignatura Sistemas avanzados de recuperad
información, contiene información sobre metadatos como XML, Dublin Core, RDF, Web semántica, etc."/>
 <meta name="DC.Publisher" content="Marta Lázaro Fernández, María José Martínez Pérez, María Carabaño Cid, Antón
Eleazar Serrano López, Iespana"/>
 <meta name="DC.Date" content="29/3/2007"/>
 <meta name="DC.Type" content="text"/>
 <meta name="DC.Format" content="text/html"/>

 <meta name="DC.Identifier" content="http://sistemasavanzadosderecuperaciondeinformacion.iespana.es"/>
 <meta name="DC.Language" content="es"/>
 <meta name="DC.Relation" content="Página principal del wiki realizado para la asignatura Sistemas Avanzados de
Recuperación de Información"/>
 <meta name="robots" content="index,follow"/>
 <meta name="verify-v1" content="5KfxHosTujF4ednOqSkKROSQKenfht2UEOEfCVjASIA=" />
 <style type="text/css" media="all" @import url(CSS/estilo.css);</style>
</head>
<body>
</body>
</html>
```



## Ejemplo de un documento DC HTML

```

<META NAME="Title" CONTENT="FrontOffice selects Verity for
Microsoft Exchange basad document management system">

<META NAME="DC.Author" CONTENT="Padovani, Marguerite">

<META NAME="DC.Author" CONTENT="Siegel, Gail">

<META NAME=" DC.Publisher" CONTENT="Verity Inc.">

<META NAME=" DC.Date" CONTENT="1996">

<META NAME=" DC.Object" CONTENT="Press Release">

<META NAME=" DC.Form" CONTENT="1 ASCII file">

<META NAME=" DC.Language" CONTENT="English">

```




UNIVERSIDAD CARLOS III DE MADRID



Term Name: creator	
URI:	<a href="http://purl.org/dc/elements/1.1/creator">http://purl.org/dc/elements/1.1/creator</a>
Label:	Creator
Definition:	An entity primarily responsible for making the resource.
Comment:	Examples of a Creator include a person, an organization, or a service. Typically, the name of a Creator should be used to indicate the entity.
Type of Term:	Property
Version:	<a href="http://dublincore.org/usage/terms/history/#creator-006">http://dublincore.org/usage/terms/history/#creator-006</a>
Note:	A second property with the same name as this property has been declared in the dcterms: namespace ( <a href="http://purl.org/dc/terms/">http://purl.org/dc/terms/</a> ). See the Introduction to the document "DCMI Metadata Terms" ( <a href="http://dublincore.org/documents/dcmi-terms/">http://dublincore.org/documents/dcmi-terms/</a> ) for an explanation.

Term Name: date	
URI:	<a href="http://purl.org/dc/elements/1.1/date">http://purl.org/dc/elements/1.1/date</a>
Label:	Date
Definition:	A point or period of time associated with an event in the lifecycle of the resource.
Comment:	Date may be used to express temporal information at any level of granularity. Recommended best practice is to use an encoding scheme, such as the W3CDTF profile of ISO 8601 [W3CDTF].
References:	[W3CDTF] <a href="http://www.w3.org/TR/NOTE-datetime">http://www.w3.org/TR/NOTE-datetime</a>
Type of Term:	Property




UNIVERSIDAD CARLOS III DE MADRID

## Dublin Core: elementos



### Los 15 elementos:

Subject	ObjectType
Title	Form
Author	Identifier
Publisher	Relation
OtherAgent	Source
Date	Language
Type	Coverage
	Scheme

<http://dublincore.org/>

<http://dublincore.org/documents/dcmi-terms/>



UNIVERSIDAD CARLOS III DE MADRID



## Calificadores DC



- ❑ Propuestos en DC4 como Canberra Qualifiers.
  - Esquema.nombre\_de\_elemento.nombre\_de\_sub-elemento = "valor"
  - DC.Creator.personalName = "Scott Adams"
- ❑ Aprobación, el 17-04-00, de Dublin Core Qualifiers (qDC)
  - Lista no cerrada que formaliza el método de utilización



UNIVERSIDAD CARLOS III DE MADRID



## Calificadores DC

Elemento DCMES	Elemento refinado	Sistema de codificación
Title	Alternative	
Creator		
Subject		LCSH MeSH DDC LCC UDC
Description	Table Of Contents Abstract	
Publisher		
Contributor		
Date	Created Valid Available Issued Modified	DCMI Period W3C-DTF
Type		DCMI Type Vocabulary
Format	Extent Medium	IMT
Identifier		URI



UNIVERSIDAD CARLOS III DE MADRID



## Cualificadores DC

Source		URI
Language		ISO 639-2 RFC 1766
Relation	Is Version Of Has Version Is Replaced By Replaces Is Required By Requires Is Part Of Has Part Is Referenced By References Is Format Of Has Format	URI
Coverage	Spatial	DCMI Point ISO 3166 DCMI Box TGN
	Temporal	DCMI Period W3C-DTF
Rights		



UNIVERSIDAD CARLOS III DE MADRID



## Metadatos



- ▣ Registros: repositorios para gestionar, recuperar, referenciar y reutilizar vocabularios de metadatos existentes. Estos registros suelen facilitar información sobre la definición, origen y localización del recurso. Actualmente, estándar ISO 11179
  - Proyecto Schemas, para RDF(S) y namespaces relacionados con proyectos de la UE (<http://www.schemas-forum.org/>)
  - Open Metadata Registry y ULIS son otros proyectos que recopilan metadatos relacionados con la Dublin Core Metadata Initiative (<http://dublincore.org/dcregistry/navigateServlet>)
- ▣ Perfiles de Aplicación



UNIVERSIDAD CARLOS III DE MADRID



## Otros Metadatos famosos



- ▣ FOAF (*Friend of a Friend*) vocabulario sobre información personal y relaciones interpersonales. Sin vocabulario estable para realizar extensiones. Es generado automáticamente en websites que trabajan con blogs.
- ▣ RSS (*RDF Site Summary* una de las siglas hay para gustos) tiene un conjunto de metadatos multipropósito, suele ser utilizado principalmente para describir sitios web. *Sindicate pages*
- ▣ Text Encoding Initiative (TEI) (<http://www.tei-c.org/>) marcado de e-text.
- ▣ Encoded Archival Description (EAD) para descripción de archivos y colecciones especiales.

Según Swoogle en Junio de 2004 que los NS, asociados a vocabularios de metadatos, más utilizados eran: FOAF (1.126.002 documentos), DC (1.126.002), MCVB (8.838), RSS(7.560 Junio, 80.000 septiembre de 2004), vCard (6.229) y Bio (6.183 ).



UNIVERSIDAD CARLOS III DE MADRID



## Editores metadatos

- ❑ <http://www.ukoln.ac.uk/cgi-bin/dcdot.pl> Dublin Core metadata editor UKOLN
- ❑ <http://vancouver-webpages.com/META/mk-metas.html> META builder 2
- ❑ Reggie <http://metadata.net/dstc/>
- ❑ <http://www.ukoln.ac.uk/metadata/new-dcdot> Dc-Dot
- ❑ [http://rainbow.arch.scriptmania.com/tools/adv\\_metatag\\_generator.html](http://rainbow.arch.scriptmania.com/tools/adv_metatag_generator.html)



UNIVERSIDAD CARLOS III DE MADRID



## Interoperabilidad

- ❑ El término interoperabilidad ha sido definido (ALA, 2000) como la capacidad que tienen algunos sistemas para intercambiar y utilizar información procedente de otro sistema diferente.
- ❑ La forma más usual de que exista es alineando vocabularios de metadatos y esquemas XML.
- ❑ La alineación puede ser estructural o lingüística

**dc:creator**要素の内容は、タグも含めてそのまま文字列として扱われ、リテラルノード ... と1行にまとまります。3行目からの

**dc:creator** 要素もリソースプロパティ要素 ...



UNIVERSIDAD CARLOS III DE MADRID





## Namespaces

### Consideraciones

- XML permite crear etiquetas 'casi' sin ninguna limitación en sus nombres
  - Esto implica que, mezclar dos documentos, con diferentes etiquetas, podría resultar en una **duplicidad** de etiquetas
- ↓
- Mediante la definición de espacios de nombres (*namespaces*), se pueden evitar estas **colisiones**
  - Tecnologías como XSL y otras muchas hacen uso de *Namespaces*

```
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:taxo="http://purl.org/rss/1.0/modules/taxonomy/"
xmlns="http://purl.org/rss/1.0/"
```



UNIVERSIDAD CARLOS III DE MADRID



## Namespaces (Definición)

Un *namespace* se identifica por su prefijo.  
Por ejemplo:

```
<xsl:stylesheet xmlns:xsl="http://www.w3.org/XSL/Transform/1.0">
```

donde:

- **xsl** es el prefijo del *namespace*
  - **Stylesheet** es el nombre completo del *namespace*
  - **http://www...** es la URI donde se puede encontrar más información sobre el estándar
  - Puede incluir otros atributos como **version...**
  - Como todo elemento XML, ha de **cerrarse** `</xsl:stylesheet>`
- <http://www.w3.org/TR/REC-xml-names/> (1999)
- <http://www.w3.org/TR/xml-names11/> (2004)



UNIVERSIDAD CARLOS III DE MADRID



## Namespaces y qualified names (Qnames)

- Para escribir en XML metadatos hay que definir previamente la ubicación del vocabulario de metadatos (Namespaces) y un prefijo para hacer referencia al vocabulario empleado (Qname)

```
<?xml version="1.0"?>
<bo:libros
 xmlns:bo="http://loc.gov.org"
 xmlns:dc="http://purl.org/dc/elements/1.1/"
 xmlns:isbn="http://www.isbn.org">
 <bo:title>XML el futuro</bo:title>
 <dc:creator>yo mismo</dc:creator>
 <isbn:number>1568491379</isbn:number>
</bo:libros>
```

XML Namespaces FAQ  
Maintained by Ronald Bourret  
Copyright 2000-2005 by Ronald Bourret  
(Last updated November, 2005)

<http://www.rpbouret.com/xml/NamespacesFAQ.htm>



UNIVERSIDAD CARLOS III DE MADRID



## Ejemplo NameSpace

```
<rss version="2.0"
 xmlns="http://blogs.law.harvard.edu/tech/rss"
 xmlns:az="http://www.amazon.com/gp/aws/landing.html"
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 xsi:schemaLocation="http://www.amazon.com/gp/aws/landing.html ">
<channel>
 <title>Testing Amazon Namespace</title>
 <item>
 <title>Harry Potter and the Goblet of Fire (Book 4)</title>
 <az:ProductInfo>
 <az:Asin>0439139597</az:Asin>
 <az:ProductName>Harry Potter and the Goblet of Fire (Book
4)</az:ProductName>
 <az:Catalog>Book</az:Catalog>
 <az:Authors>
 <az:Author>J. K. Rowling</az:Author>
 <az:Author>Mary GrandPr</az:Author>
 </az:Authors>
 <az:ReleaseDate>08 July, 2000</az:ReleaseDate> ...
 </item>
 </channel>
</rss>
```

<http://www.decafbad.com/blog/2003/10/02/rss-data-versus-namespace/>



UNIVERSIDAD CARLOS III DE MADRID



## Y otro ejemplo NameSpace con DC



```
<?xml version="1.0"?>
<metadata xmlns="http://example.org/myapp/"
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 xsi:schemaLocation="http://example.org/myapp/
 http://example.org/myapp/schema.xsd"
 xmlns:dc="http://purl.org/dc/elements/1.1/">
 <dc:title> UKOLN </dc:title>
 <dc:description> UKOLN is a national focus of expertise in
 digital information management ...</dc:description>
 <dc:publisher> UKOLN, University of Bath </dc:publisher>
 <dc:identifier> http://www.ukoln.ac.uk/ </dc:identifier>
</metadata>
```



UNIVERSIDAD CARLOS III DE MADRID



## Índice



1. Sistemas de Gestión de Contenidos
2. Representación de información estructurada
  1. XML
  2. METADATOS
  3. RDF y OWL



UNIVERSIDAD CARLOS III DE MADRID



## RDF (Resource Description Framework)

Objetivos de RDF:

- ❑ Interoperabilidad de metadatos a través de diferentes descripciones de recursos Web e intercambio de Metadatos.
- ❑ RDF trata de hacer compatibles diferentes estándares.
- ❑ RDF trata de solucionar los problemas:
  - especificación de una sintaxis
  - especificación de un esquema.
- ❑ Un recurso es descrito a través de una colección de propiedades.
  - Colección de propiedades=RDF.
  - Cada propiedad tiene un tipo de propiedad y un valor

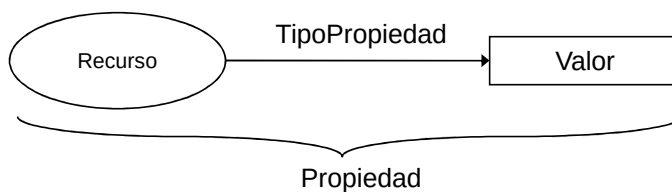
**Validadores:** <http://zoe.mathematik.uni-osnabrueck.de/RDF/parser.html>  
<http://www.w3.org/RDF/Validator/>



UNIVERSIDAD CARLOS III DE MADRID



## El modelo RDF



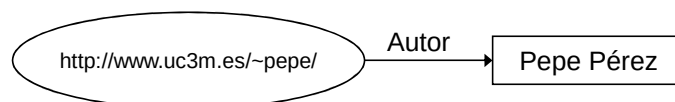
- ❑ Basado en un modelo matemático=triple
- ❑ Recursos Web representados por nodos URI
- ❑ Los conjuntos de propiedades se conocen como “descripciones”



UNIVERSIDAD CARLOS III DE MADRID



## RDF - ejemplo básico



“Pepe es el autor del recurso identificado por <http://www.uc3m.es/~pepe/>”

Sujeto/Resource: la URI

Verbo/Propiedad: autor

Predicado/Values: Pepe Pérez

```
<rdf:description rdf:about="www.uc3m.es/pepe" dc:creator="Pepe">
```



UNIVERSIDAD CARLOS III DE MADRID



## Ejemplo de RDF

```
<?xml version="1.0" ?>
<RDF xmlns = "http://w3.org/TR/1999/PR-rdf-syntax-19990105#"
 xmlns:DC = "http://purl.org/DC#" >
 <Description about = "http://www.amazon.com" >
 <DC:Title> Ontologia </DC:Title>
 <DC:Creator> Ruben Prieto-Diaz </DC:Creator>
 <DC:Date> 1999-12-31 </DC:Date>
 <DC:Subject> Metadata, RDF, Dublin Core </DC:Subject>
 </Description>
</RDF>
```



UNIVERSIDAD CARLOS III DE MADRID



## RDF - múltiples propiedades



```
...
<DC:Creator>
 <Bag>
 Maddie Azzurii
 Corky Brown
 Jacky Crystal
 </Bag>
</DC:Creator>.....
```

```
...
<DC:Creator>
 <Seq>
 Maddie Azzurii
 Corky Brown
 Jacky Crystal
 </Seq>
</DC:Creator>...
```

```
...
<SOFT:Location>
 <Alt>
 ftp://soft-sales.com.us/abc.exe
 ftp://soft-sales.com.au/abc.exe
 ftp://soft-sales.com.de/abc.exe
 ftp://soft-sales.com.uk/abc.exe
 </Alt>
</SOFT:Location>...
```



UNIVERSIDAD CARLOS III DE MADRID



## Elementos o etiquetas RDF



- ❑ <Description about = "html://dstc.com.au/report.html">
- ❑ <DC:Creator **parseType**="Resource"> para descripción de propiedades complejas.
- ❑ <DC:Creator **parseType**="Literal"> cuando el valor de la propiedad no debe ser interpretado.
- ❑ <bag> para expresar colecciones de la misma propiedad
- ❑ <seq> para expresar secuencias, donde el orden de los valores es importante
- ❑ <alt> para expresar alternativas, por ejemplo los idiomas o google.es si estás en españa y google.com
- ❑ <li> para poner los diferentes elementos dentro de la bolsa, no indica orden
- ❑ <bagID> id para un recurso y a través de ese id, <Description about = "http://www.com/report.html" bagID = "ID001" >
- ❑ se pueden identificar propiedades en bloque <Description **aboutEach** = "#ID001" >
- ❑ Resource puede ser propiedad valor <li **resource** = "http://ieee.org/papers/xml98-1.xml" />



UNIVERSIDAD CARLOS III DE MADRID



## RDF - namespaces

- ▣ Utilizados en XML para representar atributos
- ▣ Identifican Tipos de Propiedades
- ▣ Deben especificarse previamente
- ▣ Precedidos de dos puntos
  - `<DC:Title>Título del recurso</DC:Title>`
- ▣ Tienen asociados un URI



UNIVERSIDAD CARLOS III DE MADRID



## Ejemplo RDF

```
<?xml version="1.0"
<RDF xmlns="http://w3.org/TR/1999/PR-rdf-syntax-
19990222#"
 xmlns:DC="http://purl.org/DC#">

 <Description about="http://www.ugr.es">
 <DC:Title> Web de la Universidad de Granada
 </DC:Title>
 <DC:Creator>Servicio de informática
 </DC:Creator>
 <DC:Date> 1998-02-08 </DC:Date>
 <DC:Description> Resumen del contenido del
 sitio</DC:Description>
 </Description>
</RDF>
```



UNIVERSIDAD CARLOS III DE MADRID



## RDF con varios vocabularios de metadatos



```
<?xml version="1.0"
<RDF xmlns:="http://w3.org/TR/1999/PR-rdf-syntax-19990222#"
 xmlns:DC="http://purl.org/DC#"
 xmlns:AGLS="http://na.gov.au/AGLS#"

 <Description about="http://www.ugr.es">
 <DC:Title> Web de la Universidad de Granada
 </DC:Title>
 <DC:Creator>Servicio de informática </DC:Creator>
 <DC:Date> 1998-02-08 </DC:Date>
 <AGLS:Function> Information management - Internet
 </AGLS:FunctionDescription>
 </Description>
</RDF>
```



UNIVERSIDAD CARLOS III DE MADRID



## Áreas de aplicación de RDF



- Ejemplos de áreas de aplicación para RDF:
  - descripción de recursos,
  - mapas Web o mapas del sitio,
  - clasificación de contenido,
  - comercio electrónico,
  - servicios colaborativos
  - ontologías
  - sindicación



UNIVERSIDAD CARLOS III DE MADRID







## ¿Qué es RSS?

- ❑ RSS = “Really Simple Syndication”
- ❑ RSS = “Rich Site Summary”
- ❑ RSS = “RDF Site Summary”.

- <http://www.rss-specifications.com/what-is-rss.htm>
- [http://en.wikipedia.org/wiki/RSS\\_%28file\\_format%29](http://en.wikipedia.org/wiki/RSS_%28file_format%29)

♦ Básicamente: RSS es un lenguaje XML para **publicar** simultaneamente (*sindicate*) noticias en Internet.

- Noticias (prensa, anuncios)
- Eventos
- Información de proyectos
- Bibliografías
- Información de contacto
- ...

♦ La principal ventaja: La fuente informa cuando se producen cambios.

<http://www.rss-specifications.com/>

 UNIVERSIDAD CARLOS III DE MADRID



## Un poco más técnico ...

- ❑ Los archivos RSS se actualizan de forma regular y contienen **metadatos** sobre una fuente de noticias determinada y su contenido.
- ❑ Consta fundamentalmente de:
  - **Channel:** que representa la fuente de las noticias.
  - **Title:** título del canal.
  - **Link:** vínculo del canal.
  - **Description:** descripción del canal.
- ❑ Además, consta de uno o varios elementos **item** que representan elementos de noticias individuales, cada uno de los cuales debe disponer de un campo **title**, **link** o **description**.

 UNIVERSIDAD CARLOS III DE MADRID



## Ejemplo

```
<?xml version="1.0" encoding="UTF-8" ?>
<rss version="2.0">
 <channel>
 <title> BCR: The Third Indicator</title>
 <link>http://www.bcr.org/publications/thirdind</link>
 <description>The Third Indicator, published monthly, is a technical memo focusing on
 OCLC products and services. It includes general OCLC news as well as detailed
 technical information on cataloging, reference and resource sharing. Announcements
 of new OCLC developments are also included.</description>
 <lastBuildDate>Tue, 21 Sep 2004 21:37:39 GMT</lastBuildDate>
 <generator>ListGarden Program 1.01</generator>
 <docs>http://blogs.law.harvard.edu/tech/rss</docs>
 <item>
 <title>WorldCat Resource Sharing Training</title>
 <link>http://www.bcr.org/publications/thirdind/2004/august/augsharetrain04.html</link>
 <description>If you'd like to see what WorldCat Resource Sharing looks like and learn
 more about it, visit the OCLC Web site at www.oclc.org/ill/migration/ or view the
 WorldCat Resource Sharing tutorial at
 www5.oclc.org/downloads/tutorials/firstsearch/sv/rsbasics/intro/index.html/.</descri
 ption>
 <pubDate>Tue, 21 Sep 2004 19:29:47 GMT</pubDate>
 <guid isPermaLink="false">thirdind-2004-08-21-19-29-47</guid>
 </item>
 </channel>
</rss>
```




UNIVERSIDAD CARLOS III DE MADRID



## RSS en Versión 0.91 (XML)

<http://blogs.law.harvard.edu/tech/rss>

```
- <rss version="0.91">
- <channel>
 <title>DOCUMENTAMANIA</title>
 <link>
 http://www.documentamania.freesevers.com/index.html
 </link>
 <description>Pagina sobre DOCUMENTAMANIA y RECUPERACION WEB</description>
 <language>es-es</language>
 <webMaster>100029793@alumnos.uc3m.es</webMaster>
- <item>
 <title>Documentamania</title>
 <link>
 http://www.documentamania.freesevers.com/index.html
 </link>
 <description>
 Página principal con la definicion de Documentamania y otras preguntas sobre esta.
 </description>
</item>
+ <item></item>
+ <item></item>
+ <item></item>
+ <item></item>
+ <item></item>
</channel>
</rss>
```




UNIVERSIDAD CARLOS III DE MADRID

E. Obligatorios:

- Rss Channel:
  - Title
  - Link
  - Item:
    - Title o Description
    - Link opcional pero recomendable



## RSS en Versión 2.0 (XML)

---

```

<?xml version="1.0" encoding="ISO-8859-1" ?>
<rss version="2.0" xmlns:dc="http://purl.org/dc/elements/1.1/">
 <channel>
 <title>DOCUMENTAMANIA</title>
 <link>
 http://www.documentamania.freesservers.com/index.html
 </link>
 <description>Pagina sobre DOCUMENTAMANIA y RECUPERACION WEB</description>
 <language>es-es</language>
 <webMaster>100029793@abamos uc3m.es</webMaster>
 </channel>
 <item>
 <title>Documentamania</title>
 <link>
 http://www.documentamania.freesservers.com/index.html
 </link>
 <description>
 Página principal con la definición de Documentamania y otras preguntas sobre esta.
 </description>
 <dc:creator>Mariano Ruiz</dc:creator>
 <pubDate>Wed, 16 Mar 2005 09:00:00 GMT</pubDate>
 <dc:date>2005-03-16</dc:date>
 </item>
 <item></item>
 <item></item>
 <item></item>
 <item></item>
 <item></item>
</rss>

```




UNIVERSIDAD CARLOS III DE MADRID





<http://www.documentamania.freesservers.com/rss10.xml>

---

```

<?xml version="1.0" encoding="ISO-8859-1" ?>
<rdf:RDF
 xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
 xmlns="http://purl.org/rss/1.0/"
 xmlns:dc="http://purl.org/dc/elements/1.1/"
 xmlns:content="http://purl.org/rss/1.0/modules/content/">
 <rdf:RDF>
 <channel rdf:about="http://www.documentamania.freesservers.com/">
 <title>DOCUMENTAMANIA</title>
 <link></link>
 <description>Pagina sobre DOCUMENTAMANIA y RECUPERACION WEB</description>
 <language>es-es</language>
 </channel>
 <rdf:Seq>
 <rdf:li rdf:resource="http://www.documentamania.freesservers.com/index.html"/>
 <rdf:li rdf:resource="http://www.documentamania.freesservers.com/documentamania/enlaces.html"/>
 <rdf:li rdf:resource="http://www.documentamania.freesservers.com/documentamania/autor.html"/>
 <rdf:li rdf:resource="http://www.documentamania.freesservers.com/documentamania/foro.html"/>
 <rdf:li rdf:resource="http://www.documentamania.freesservers.com/documentamania/historico.html"/>
 <rdf:li rdf:resource="http://www.documentamania.freesservers.com/documentamania/recuperacion_web.html"/>
 </rdf:Seq>
 </rdf:RDF>
 <item rdf:about="http://www.documentamania.freesservers.com/index.html">
 <title>Documentamania</title>
 <link></link>
 <description></description>
 <dc:creator>Mariano Ruiz</dc:creator>
 <dc:date>2005-03-26</dc:date>
 </item>
 <item rdf:about="http://www.documentamania.freesservers.com/documentamania/enlaces.html"></item>
 <item rdf:about="http://www.documentamania.freesservers.com/documentamania/autor.html"></item>
 <item rdf:about="http://www.documentamania.freesservers.com/documentamania/foro.html"></item>
 <item rdf:about="http://www.documentamania.freesservers.com/documentamania/historico.html"></item>
 <item rdf:about="http://www.documentamania.freesservers.com/documentamania/recuperacion_web.html"></item>
</rdf:RDF>

```

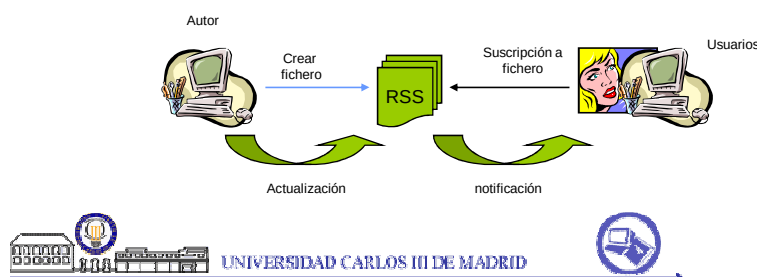



UNIVERSIDAD CARLOS III DE MADRID



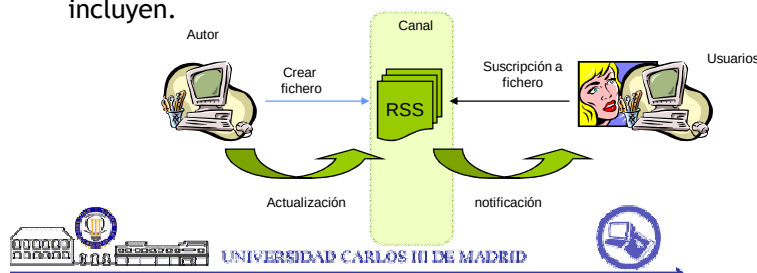
## ¿Cómo funciona RSS?

- ❑ El autor crea un fichero RSS.
- ❑ Los usuarios se suscriben al fichero a través de un lector de noticias o agregador.
- ❑ Cuando el autor actualiza el fichero RSS, los nuevos elementos se notifican automáticamente a los usuarios quedando a su disposición para su lectura.



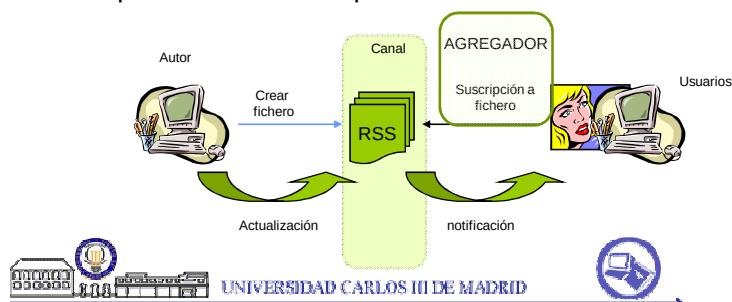
## ¿Qué es un canal?

- ❑ Un canal, **bitácora** o **blog** (abreviatura de *weblog*, pronunciado “we blog”), es un conjunto de noticias *online*, representadas en orden cronológico inverso e incluidas en un fichero RSS.
- ❑ También se utiliza para denominar el sistema que aloja y sirve un conjunto de canales.
- ❑ Típicamente, este tipo de sistemas incluyen enlaces entre ellos, proporcionando información adicional sobre los temas que incluyen.



## ¿Qué es un agregador?

- ❑ Una aplicación o un servicio remoto que, periódicamente, lee un conjunto de fuentes o **canales** en formato XML.
- ❑ Cuando detecta nuevos elementos, muestra un resumen de los mismos en un listado ordenado cronológicamente, comenzando por el más moderno.
- ➔ La aplicación necesaria para leer ficheros RSS.



## Tipos de agregadores

- ❑ Clientes/agentes independientes
  - **FeedReader**, Radio UserLand
- ❑ Complementos PIM (*Personal Information Manager*)
  - **Pluck**, NewsGator, intraVnews
- ❑ Complementos de Navegador
  - **Firefox 1.0**, Sage
- ❑ Sitios Web
  - **Bloglines**, NewsIsFree
- ➔ Listado de agregadores

<http://www.lights.com/weblogs/rss.html>





## Aplicaciones

<i><b>Tipos de kos</b></i>	<i><b>Aplicaciones posibles</b></i>	<i><b>Ejemplos</b></i>
Ontologías (OWL, RDF), metadatos	Rss, foat	<a href="http://rss.elmundo.es/rss/">http://rss.elmundo.es/rss/</a> (noticias)   <a href="http://www.elpais.com/rss/index.html">http://www.elpais.com/rss/index.html</a> (noticias)   <a href="http://www.edreams.es/edreams/espanol/rss/eDreams-feeds.jhtml">http://www.edreams.es/edreams/espanol/rss/eDreams-feeds.jhtml</a> (Agencia de viajes)   Alertas (p.e. calendario para uso público o privado )   <a href="http://www.rsscalendar.com/rss/">http://www.rsscalendar.com/rss/</a>   <b>Clasificados</b> (Casas/Empleos)   <a href="http://www.masternewmedia.org/es/2006/04/11/rss_usos_y_aplicaciones_no.htm">http://www.masternewmedia.org/es/2006/04/11/rss_usos_y_aplicaciones_no.htm</a>   <b>Seguimiento de Envíos</b> notificaciones vía alimentadores RSS cuando los paquetes son enviados   RSS o Atom para ser escuchados en dispositivos móviles y computadoras personales <a href="http://www.apple.com/itunes/">http://www.apple.com/itunes/</a>   Edición de videos colaborativo pronto puede ser una realidad para muchos   <a href="http://www.masternewmedia.org/es/2004/12/22/la_segunda_capa_de_la.htm">http://www.masternewmedia.org/es/2004/12/22/la_segunda_capa_de_la.htm</a>   FOAF es un proyecto de Web Semántica, que permite crear páginas Web para describir personas, vínculos entre ellos, y cosas que hacen y crean.   <a href="http://www.hipertexto.info/documentos/web2.htm">http://www.hipertexto.info/documentos/web2.htm</a>   Para la bolsa







UNIVERSIDAD CARLOS III DE MADRID