

LA WEB SEMÁNTICA

claudio gutierrez

Departamento de Ciencias de la Computación

Universidad de Chile

`cgutierrez@dcc.uchile.cl`

(Transparencias sobre originales de E. Krsulovic)

Agenda de la Presentación

- **Estandarizar para comunicarse**
 - Problemas de la Web actual
 - Posible solución: Web Semántica
- **HTML, XML**
 - ventajas
 - limitaciones
- **Metadatos**
 - ¿Para que?
 - El modelo RDF
- **Ontologías**
- **Conclusiones**

Ejemplo 1

Año 1936: el doctor Julio Medina posee una biblioteca personal con cientos de libros, que ocupan dos piezas grandes. El orden no es el fuerte del doctor. Así, cada vez que necesita cierta información, debe revisar cada libro, hasta encontrar lo que busca.

Problemas:

- no siempre encuentra lo que busca
- búsqueda toma mucho tiempo

Ejemplo 1 (continuación)

Solución: Catalogar los libros

- ¿Cómo **referenciar** cada libro de forma única?
 - Cuatro letras del autor + año (**como nombrarlo**)
 - Nro. de estante + correlativo (**donde encontrarlo**)
- ¿Qué **alfabeto** utilizar?
 - Latino
- ¿Cómo **anotar la información** de los libros?
 - En fichas almacenadas en tres cajas de madera según criterio. Cada ficha tiene la referencia a un libro.
- ¿Qué **criterios** utilizar para **catalogar** los libros?
 - autor, título y tema

Ejemplo 1 (continuación)

Año 1937: se muda a la casa contigua a la del doctor, un escritor ruso, Ivo Ronikoff, quien también tiene una gran biblioteca. Ivo es ordenado y su biblioteca está catalogada así:

- **Referencias:**
 - tres letras del título + año para nombrarlo
 - correlativo como ubicación
- **Alfabeto:** cirílico
- **Información** sobre los libros: anotada en gordos cuadernos tipo guía de teléfono.
- **Criterios de catalogación:** palabra clave, editorial y año.

Ejemplo 1 (continuación...)

Año 1938: Llega otro vecino, también un gran aficionado a los libros: el profesor Tsun Ming Fu, recién venido de Shangai.

...ni nos imaginemos como ordena su biblioteca...

Ejemplo 1: ¿Cómo compartir información?

El Dr. Medina, el Sr. Ronikoff y el profesor Ming Fu gustan compartir sus libros, pero nunca les fué fácil buscar información en las tres bibliotecas.

Problemas:

- Alfabetos diferentes
- Referencias diferentes
- Anotaciones diferentes
- Criterios de catalogación diferentes.

Ejemplo 1: ¿final feliz?

Año 2001: Ana Medina Ronikoff reta a sus abuelos por tercros y los insta a “subirse a las nuevas tecnologías”. Les propone usar:

- alfabeto: UNICODE
- referencias: URI
- Anotaciones: XML y RDF
- Criterios de catalogación: Ontologías

Ejemplo 1: Las dudas de los abuelos

Los abuelos se sienten insultados y replican:

- El Dr. Medina: ¡*tanta parafernalia!*
- El escritor Ivo, algo más tecnologizado:
¿Por que mejor no usar simplemente HTML?
- El profesor Ming Fu, diario lector de Infoweek:
¿Y por que XML no nos basta?

Ejemplo 2: la Web actual

Año 2001: Para un importante ejecutivo de una gran empresa es muy costoso buscar información en Internet por el tiempo que le toma. Por ello delega esta responsabilidad a un asistente.

El asistente busca la información así: utiliza un browser y en sitios como **Google** o **Altavista** encuentra gran cantidad de enlaces. Los revisa y si tiene suerte, extrae la información solicitada por su jefe. El ejecutivo recibe la información por parte de su asistente algunas horas después.

- Aquí el asistente es un **agente humano**.

Ejemplo 3: ¿la Web futura?

Año 2010(?): Un niño pregunta a su asistente personal que lleva puesto (con un monóculo como pantalla y algo como un teclado en la manga de su chaqueta): “¿Quién era el entrenador cuando la roja le puso dos goles al campeón del mundo en el 2001?”. El asistente personal levanta un proceso en background y busca sitios Web donde encontrar lo solicitado por el niño, quien en poco más de un minuto recibe la respuesta: Pedro García.

- Aquí el asistente es un **agente digital**.

Agentes - Hoy y Mañana

Los asistentes de los ejemplos 2 y 3 son **agentes** que tienen las siguientes características:

- **Entienden** lo que se les pide buscar
- **Comprenden** el contenido de los sitios Web que visitan
- **Validan** si lo encontrado corresponde a lo que se le pidió buscar
- **Deducen** nueva información de la ya obtenida

Web Semántica

¿Que se necesita para automatizar el proceso de:

- **entender** una consulta?
- **Comprender** sitios Web?
- **Validar** información?
- **Deducir** nueva información?

Estandarizar alfabeto, referencias, lenguaje, formato, anotaciones sobre significados, conceptos generales, reglas y sistemas de deducción, etc.

Web Semántica: HTML es limitado

- Orientado a **presentación** de datos
- No permite describir datos
- No es extensible: congela pequeño universo de etiquetas

Web Semántica: XML es necesario

- Definición de datos que se auto-describen en formato no propietario (estándares)
- Formato común para intercambio de documentos (certificados, facturas, referencias de libros, currículum)
- Plantilla para elaborar documentos estándar (esquemas XML) (ej. W3C tiene uno para sus reportes)
- Integración de información de diferentes fuentes en documentos uniformes

Ejemplo 4: ¿Por que entendemos?

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<curso>
  <profesor>Eduardo Godoy</profesor>
  <nombre>Bases de Datos</nombre>
  <semestre ano="2001">Primavera</semestre>
  <horario>
    <modulo tipo="catedra">2.1</modulo>
    <modulo tipo="catedra">3.1</modulo>
    <modulo tipo="auxiliar">1.5</modulo>
  </horario>
</curso>
```


Ejemplo 4: ¿Por que **no** entendemos?

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<rukangma>
  <chaw> Aukan </chaw>
  <nguke> Peyeché </nguke>
  <yall>
    <pichi s="wentru"> mari </pichi>
    <pichi s="wentru"> epu </pichi>
    <pichi s="domo"> regle </pichi>
  </yall>
</rukangma>
```

Ejemplo 4: Moraleja

Necesitamos descripción de datos

- Significado del documento es **intuitivamente** claro a un humano debido a:
 - marcado “semántico”
 - las etiquetas son términos de un dominio
 - el contexto del documento es manejado por el lector
- Pero, los computadores no tienen intuición
 - el nombre de etiquetas **per se** no provee semántica

Web Semántica: limitaciones de XML

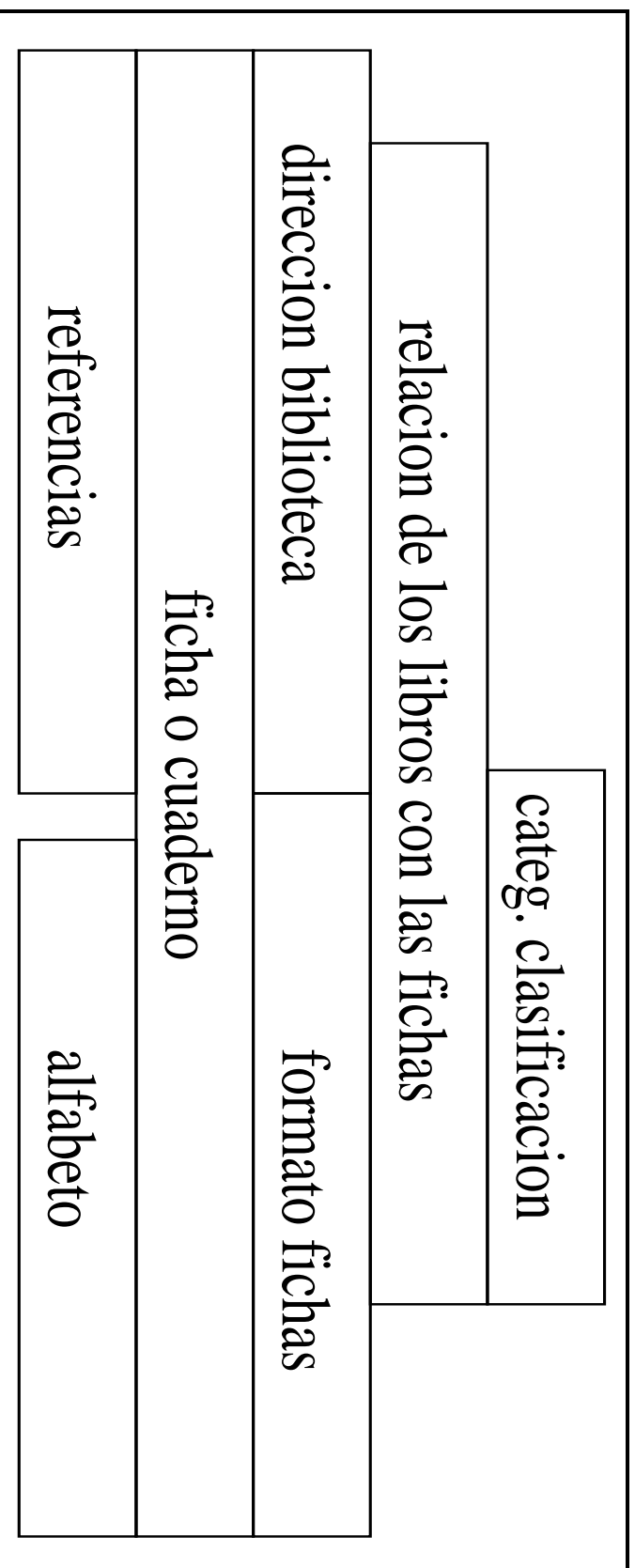
- ¿Cómo sé que determinado documento **es** una factura y no una boleta?
- ¿Cómo determino si el documento que es la firma digital de Ana Medina?
- ¿Qué hago para validar un certificado de notas de un estudiante de Namibia que quiere estudiar en Chile?
- ¿Cómo determino si el cheque que recibí está “bueno”?

Web Semántica: XML no es suficiente

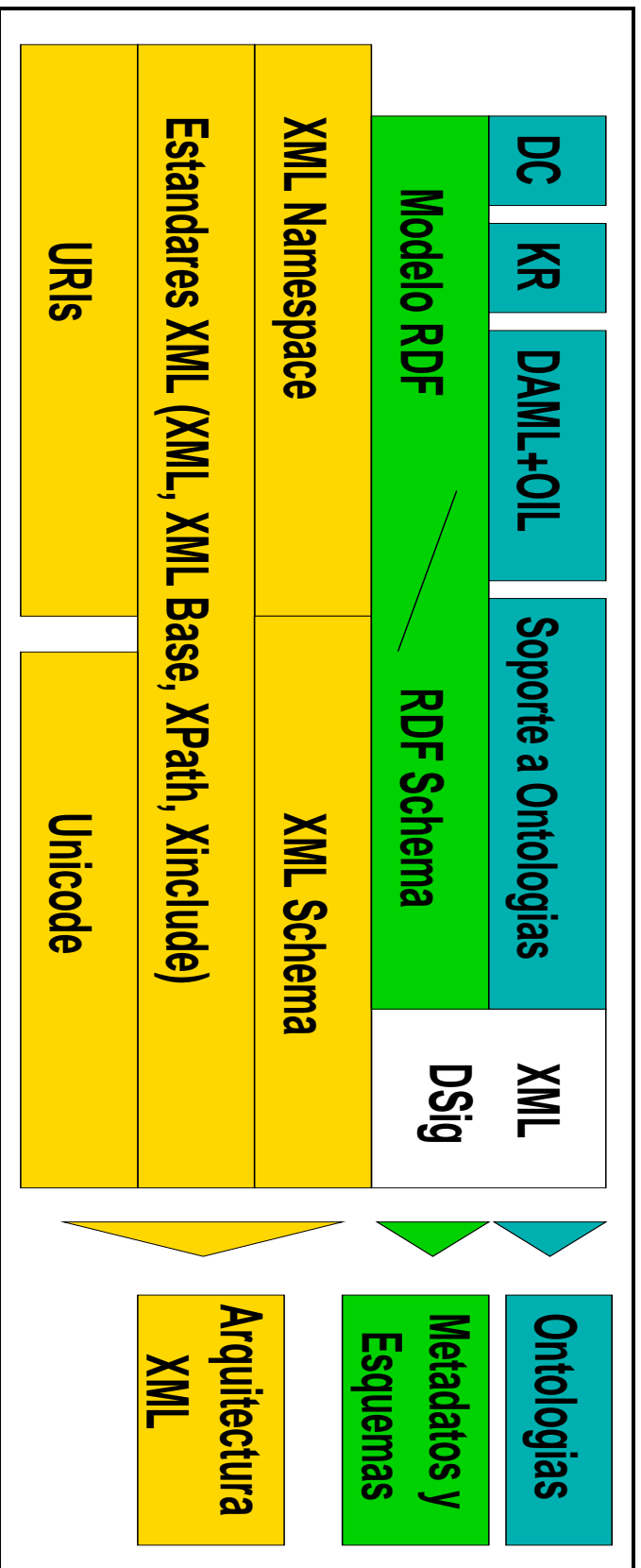
XML estandariza formato, no significado

Ejemplo 1 (continuación)

La Arquitectura de la solución de Ana Ronikoff



Arquitectura de la Web Semántica



¿Que es una URI?

- URL: descripción de ubicación
- URN: descripción de nombre
- URI = URL + URN + ...
- **Todo** es una URI

Cómo anotar significados

Metadatos = datos sobre datos

- Relaciones entre entidades (URI)
Ej: Nicanor – autor de **Obra Gruesa**
- Afirmaciones: Sujeto - Predicado - Objeto
Ej: autor(Nicanor, ObraGruesa) .
- Formalización para la web: Modelo RDF

¿Qué es el modelo RDF?

Resource: recursos que pueden ser nombrados con URIs

Description: afirmaciones sobre las propiedades de los recursos

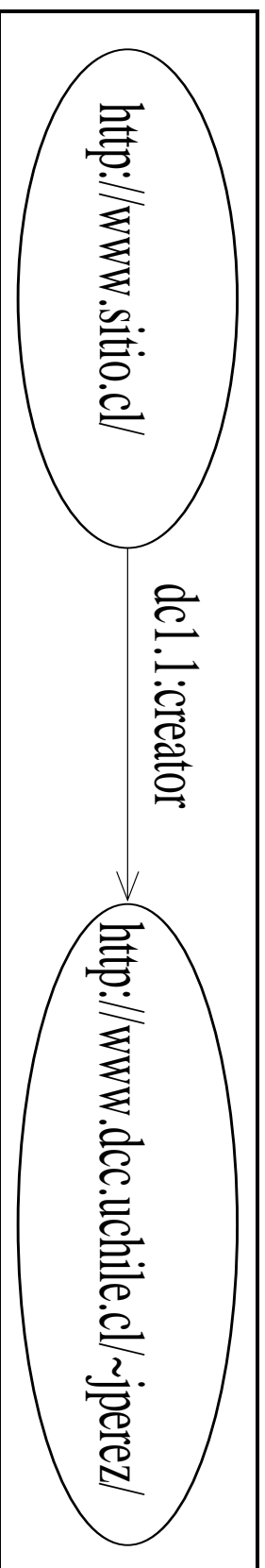
Framework: un modelo común

RDF es un modelo. Hay varias sintaxis: RDF XML, Notation3

RDF: Orígenes y Objetivos

- Orígenes
 - W3C necesita metadatos: PICS, Dsig
 - Influencias de diversas comunidades: KR, OOP, DC, W3C
- Objetivos
 - Convención para aplicaciones que intercambian metadatos
 - XML usado como sintaxis
 - Vocabularios semánticos definidos por comunidades de expertos
 - Mezcla fina de diversos metadatos

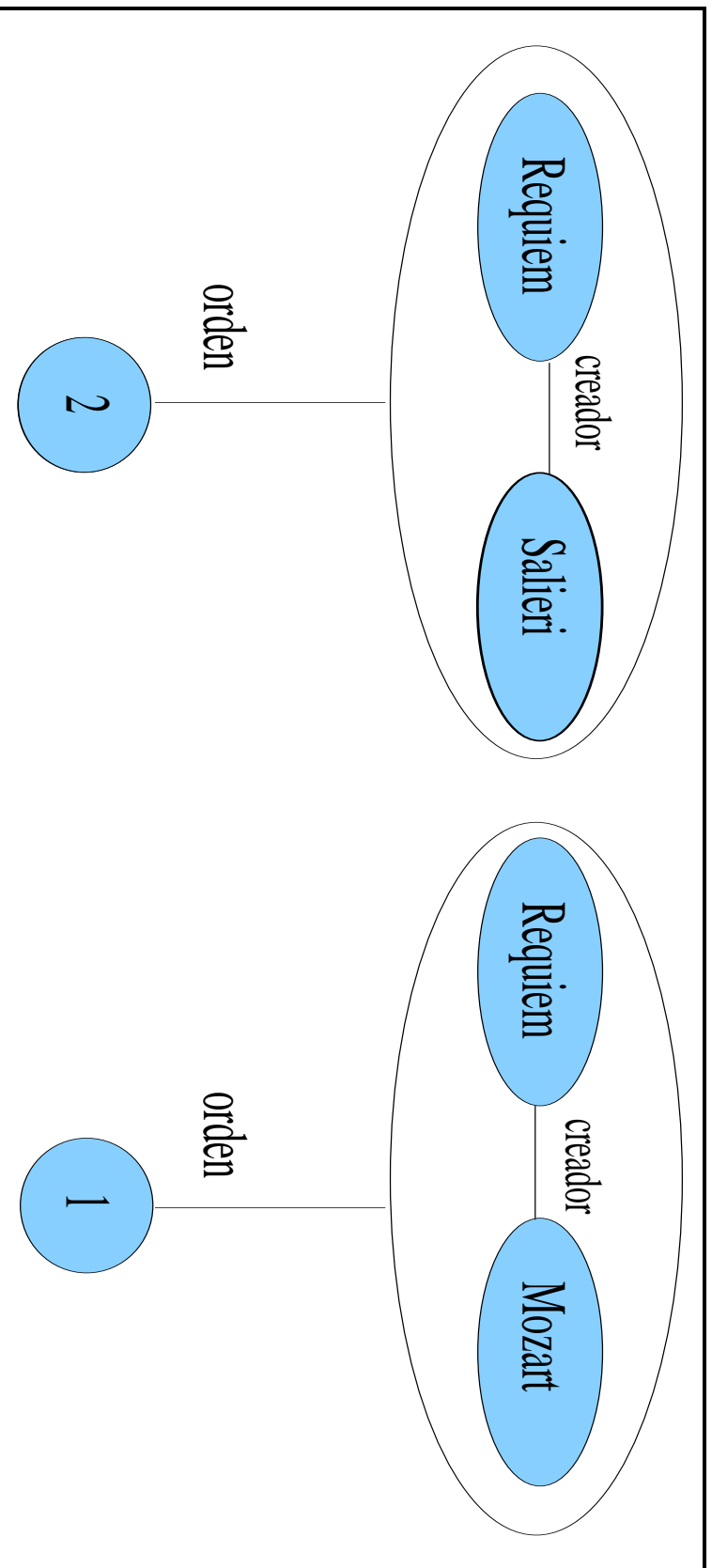
Ejemplo 5: una afirmación RDF



Ejemplo 5 (continuación)

```
1 <?xml version="1.0" encoding="ISO-8859-1"??>
2 <rdf:RDF
3   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
4   xmlns:dc="http://purl.org/dc/elements/1.1/">
5   <rdf:Description about="http://www.sitio.cl/">
6     <dc:creator
7       rdf:resource="http://www.dcc.uchile.cl/~jperez/">
8   </rdf:Description>
9 </rdf:RDF>
```

Ejemplo 6: afirmaciones sobre afirmaciones



RDF Schema

- Un esquema puede definir el significado, características y relaciones de un conjunto de propiedades.
- Puede incluir restricciones en los valores potenciales y herencia de propiedades en otros esquemas.
- RDFS provee un vocabulario definido sobre RDF que permite el modelo de objetos con una semántica claramente definida.

¿Qué es Ontología?

- Ciencia del ser
- Se ocupa de las características y organización de la realidad.
- Intenta responder las preguntas:
 - ¿Qué es el ser?
 - ¿Cuales son las características comunes de todos los seres?
 - ¿Como se clasifican los seres?

Ontología(s) en computación

- Especificación explícita de una conceptualización.
- Entendimiento compartido de algún dominio de interés.
- Especificación formal
 - ... de una conceptualización de un dominio
 - ... de alguna parte de la realidad que es de interés

El uso de ontologías provee una herramienta muy poderosa para describir objetos y sus relaciones con otros objetos.

Ejemplo 7: Dublin Core

Vocabulario de metadatos para facilitar el descubrimiento de recursos.

<http://purl.org/dc/elements/1.1/>

Title, Contributor, Author/Creator, Publisher, Subject, Description, Date, Resource, Format, Resource Identifier, Language, Source, Coverage, Rights.

Ejemplo 7: un fragmento de DC

```
<rdf:Property rdf:about = "&dcns;creator">
  <rdfs:label>Author/Creator</rdfs:label>
  <rdfs:comment>An entity primarily responsible for
    making the content of the resource.
</rdfs:comment>
  <eor:comment> Examples of a Creator include a person,
    an organisation, or a service. Typically, the name
    of a Creator should be used to indicate the entity.
</eor:comment>
  <rdfs:isDefinedBy rdf:resource = "&dcns;" />
</rdf:Property>
```

¿Que nos falta?

- Firmas Digitales
- Lógica, Deducción, Confianza
- Lenguajes de Consulta
- Aplicaciones **matadoras**
- Varias otras cosas...

Conclusiones

- La infraestructura de web semántica está evolucionando (tiene problemas no resueltos aún).
- RDF (S) es un estandar para describir la semántica de información, lo que provee interoperabilidad para compartir conocimiento.
- Piense en usar RDF para sus próximas aplicaciones.
- Reutilizar ontologías (¡no re-invente la rueda!)

Conclusiones (final)

Regla de tres sobre las tendencias en la Web:

$$\frac{\text{Web Semántica}}{2001} = \frac{\text{Web actual}}{1994}$$