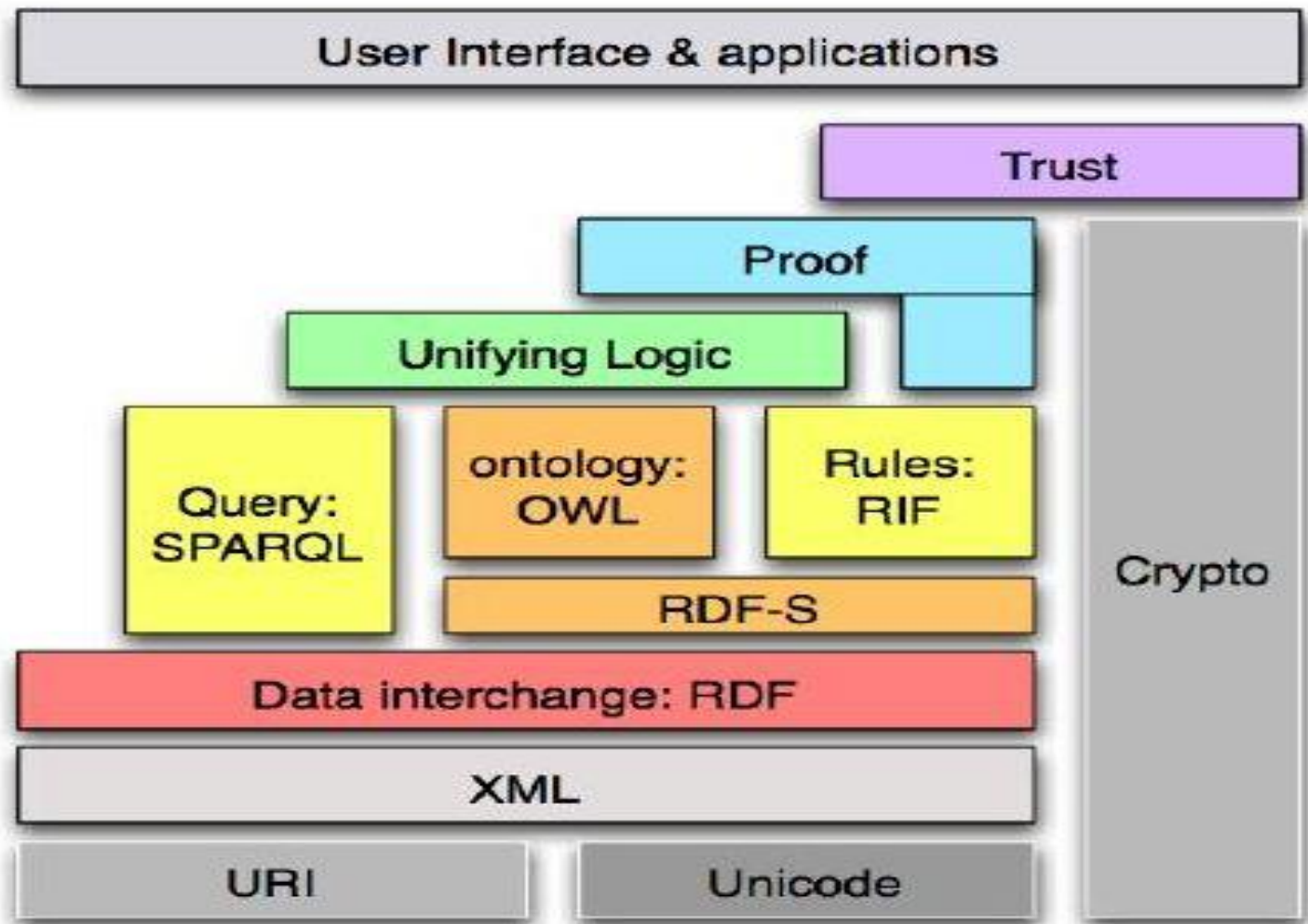


Tecnologías para la Web Semántica

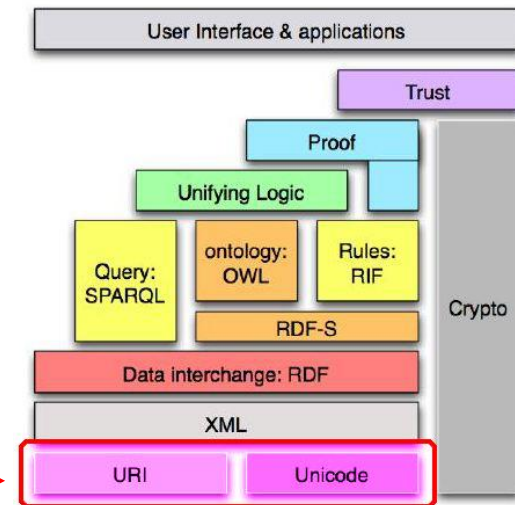
Arquitectura II
Lenguaje de etiquetas

Vamos por partes...



URI

Uniform Resource Identifiers,
identificadores de recursos únicos,
sin posibilidad de ambigüedad.



Puede ser una localización (URL), un nombre (URN) o ambos.

Ejemplos:

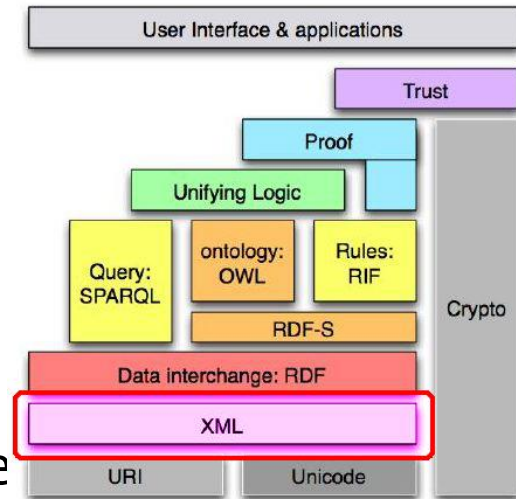
<http://www.utn.edu.ar>

urn:isbn:0-395-36341-1

XML

eXtensible Markup Language

- Metalenguaje de etiquetas extensibles, se puede “acomodar” a las necesidades de cada uno.



Lenguajes de etiquetas

- ▶ **Un lenguaje de etiquetas** debe especificar
 - Las etiquetas permitidas
 - Las etiquetas requeridas
 - Cómo se distinguen las etiquetas del texto
 - Qué significan las etiquetas
- ▶ XML sólo especifica las tres primeras, la cuarta es especificada por DTD

Lenguajes de etiquetas

GML

(Generalized Markup Language)

IBM, 1973



SGML

(Standard Generalized Markup Language)

ISO 8879, 1986



HTML Lenguaje para formato

(Hyper Text Markup Language)

Berners-Lee, 1991

SGML(ISO 8879)

- ▶ Standard Generalized Markup Language
- ▶ Estandar internacional para la definición de descripciones de estructura y contenido en documentos de texto
- ▶ Intercambiable: independiente del dominio, independiente del sistema
- ▶ Tags no están predefinidos
- ▶ Utiliza DTD para la validación de la estructura de un documento
- ▶ Extenso, potente, complejo
- ▶ Utilizado para industrias y comercios por décadas

HTML

HyperText Markup Language

- ▶ Hypertext
 - Documento que contiene vínculos a otros documentos
- ▶ Markup language
 - Notación para escribir texto con etiquetas
 - Utiliza un conjunto predefinido de etiquetas
 - Etiquetas indican la estructura del texto
 - Etiquetas pueden encerrar parte del texto
- ▶ Lenguaje estándar utilizado para dar formato a documentos utilizado en la Web

Ejemplo HTML

<h1>Agilitas Physiotherapy Centre</h1>

Welcome to the home page of the Agilitas Physiotherapy Centre. Do you feel pain? Have you had an injury? Let our staff Lisa Davenport, Kelly Townsend (our lovely secretary) and Steve Matthews take care of your body and soul.

<h2>Consultation hours</h2>

Mon 11am – 7pm

Tue 11am – 7pm

Wed 3pm – 7pm

Thu 11am – 7pm

Fri 11am – 3pm<p>

But note that we do not offer consultation during the weeks of the
State Of Origin games.

Personas no tienen problemas con la interpretacion
Maquinas (agentes de software):

¿Cómo distinguir terapeuta de la secretaria,
¿Cómo determinar horarios exactos de consulta
State Of Origin games...¿Cuándo?

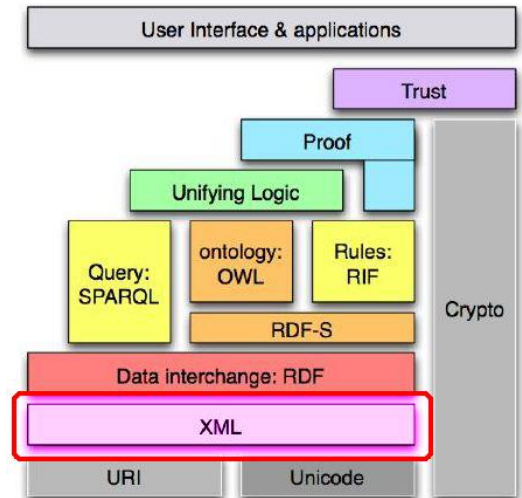
Ejemplo HTML. Mejor representacion

```
<company>  
  <treatmentOffered>Physiotherapy</treatmentOffered>  
  <companyName>Agilias Physiotherapy Centre</companyName>  
  <staff>  
    <therapist>Lisa Davenport</therapist>  
    <therapist>Steve Matthews</therapist>  
    <secretary>Kelly Townsend</secretary>  
  </staff>  
</company>
```

Documentos Web Estructurados: XML

eXtensible Markup Language

- Metalenguaje de etiquetas extensibles, se puede “acomodar” a las necesidades de cada uno.
- Esquemas que definen y restringen su estructura.



Establece relaciones básicas pero no una semántica.

Ejemplo HTML-XML

HTML	XML
<pre><h2>Nonmonotonic Reasoning: Context-Dependent Reasoning</h2> <i>by V. Marek and M. Truszczyński</i>
 Springer 1993
 ISBN 0387976892</pre>	<pre><book> <title>Nonmonotonic Reasoning: Context- Dependent Reasoning</title> <author>V. Marek</author> <author>M. Truszczyński</author> <publisher>Springer</publisher> <year>1993</year> <ISBN>0387976892</ISBN> </book></pre>

HTML

Nonmonotonic Reasoning: Context- Dependent Reasoning
by V. Marek and M. Truszczyński
Springer 1993
ISBN 0387976892

HTML vs. XML

- ▶ Ambos usan **tags** o **etiquetas** (ej. `<h2>` and `</year>`)
- ▶ Tags deben estar anidados (tags dentro de tags)
- ▶ Usuarios humanos pueden leer e interpretar ambas representaciones HTML y XML facilmente

... Pero que pasa con las máquinas?

Problemas con la interpretacion automática

Un agente inteligente trata de recuperar los nombres de los autores del libro:

- ▶ Podrían aparecer inmediatamente despues del titulo
- ▶ O inmediatamente luego de **by**
- ▶ Hay dos autores?
- ▶ O uno llamado “V. Marek and M. Truszczyński”?

HTML

Nonmonotonic Reasoning: Context- Dependent Reasoning

by V. Marek and M. Truszczyński

Springer 1993

ISBN 0387976892

HTML vs. XML

- ▶ Documentos HTML no contienen **información estructural**: contenidos del documento y sus relaciones.
- ▶ XML mas fácilmente accesible para máquinas:
 - Se describe cada elemento de información.
 - Se definen las Relaciones a traves de la estructura anidada.
 - Ej., <author> tags dentro de <book> tags, por lo tanto describen autores de un libro particular.

HTML vs. XML

HTML	XML
<pre><h2>Relationship force- mass</h2> <i> F = M × a </i></pre>	<pre><equation> <meaning>Relationship force-mass</meaning> <leftside> F </leftside> <rightside> M × a </rightside> </equation></pre>

HTML`

Relationship force-mass

$$F = M \times a$$

HTML vs. XML

- ▶ En ambos ejemplos HTML mismos tags
- ▶ En XML completamente diferentes
- ▶ HTML tags definen display: color, negrita ...
- ▶ XML tags no son fijos: tags definibles por el usuario

Que es XML ?

- ✓ XML es un lenguaje de marcas tal como HTML
- ✓ XML fue diseñado para describir datos
- ✓ XML tags no están predefinidos.
- ✓ XML usa un Document Type Definition (DTD) o un XML Schema para describir los datos
- ✓ XML con un DTD o XML Schema fue diseñado para ser auto-descriptivo.

XML es Recomendación de la W3C

HTML vs. XML

HTML	XML
<pre><html><head> <title>Weather Report</title> </head> <body> <h1>Weather Report</h1> Date: July 1, 1999
 Time: 9:00 a.m.
 City: Frankfurt
 Country: Germany <p> Temperature: 18° Celsius </body> </html></pre>	<pre><?xml version="1.0"?> <weather-report> <date>July 1, 1999</date> <time>9:00 a.m.</time> <area> <city>Frankfurt</city> <country>Germany</country> </area> <measurements> <skies>rain</skies> <temp scale="C">18</temp> </measurements> </weather-report></pre>

Diferencias entre XML and HTML

- ▶ XML no es un reemplazo de HTML.
- ▶ XML es un **metalenguaje** que no tiene un conjunto fijo de etiquetas pero permite al usuario definir sus propias etiquetas.

XML fue diseñado para describir datos.
HTML fue diseñado para mostrar datos.

Estructura de un documento XML

<?xml version="1.0" encoding="ISO-8859-1"?>

<nota Id = "1">

<de> Juan </de>

<dirigidaA> María </dirigidaA>

<titulo> Recordatorio </titulo>

<cuerpo> s de mi! </cuerpo>

</nota>

tag de inicio

atributo

Prólogo
<?xml

tag de cierre

contenido

Identificando marcas

- ▶ Las marcas *tags* están encerrados entre *delimitadores*
 - Tags de comienzo están delimitados por `< y >`
 - Tags de cierre están delimitados por `</ >`
 - valores de atributos encerrados entre comillas
 - Instrucciones de procesamiento están delimitadas por `<? y ?>`
 - comentarios están delimitados por `<!-- y -->`
 - declaraciones `<! y >`
- ▶ Todo el texto que no es marca, constituye los datos del documento

Examples

```
<nota>
    </nota>
<nota Id ="1">

<?xml ... ?>
<!-- Comment -
->

<!ELEMENT ... >
```

XML es case sensitive

XML Building blocks

- ▶ PI (Processing Instruction)
- ▶ Etiquetas
- ▶ Elementos
- ▶ Contenido
- ▶ Atributos
- ▶ Entidades
- ▶ Comentarios

Etiquetas

- ▶ Se utilizan para especificar un nombre para una porción de información dada.
- ▶ Una etiqueta consiste en abrir y cerrar *angular brackets* (<>) que encierran el nombre de la etiqueta.
- ▶ Ejemplo

```
<EMP_NAME>Nick Shaw</EMP_NAME>
```


Elementos

- ▶ Se representan utilizando etiquetas.
- ▶ Un documento XML debe siempre tener un elemento raíz.
- ▶ Formato General :

`<element> ... </element>`

- ▶ Elemento vacío:

`<empty-Element />`

- ▶ Ejemplo

`<Authorname>John Smith</Authorname>`

Contenido

- ▶ Información representada por los elementos de un documento XML.
- ▶ Ejemplo

```
<BOOKNAME>The Painted house</BOOKNAME>
```

Atributos

- ▶ Localizados en la primer etiqueta de un elemento
- ▶ Provee información adicional acerca de los elementos
- ▶ Van entre comillas
- ▶ Debo utilizar un elemento o un atributo?

XML Reglas para nombres de elementos

- Nombres pueden contener letras, números y otros caracteres.
- Nombres no pueden tener el carácter “:”
- Nombres no deben comenzar con un número o caracter de puntuación
- Nombres no deben empezar con las letras xml (o XML o Xml ..)
- Nombres no pueden contener espacios en blanco.
- Nombres no tienen restricciones de longitud.

No hay palabras reservadas.
Invente nombres descriptivos

Estructura de un documento XML

ATRIBUTOS

- ▶ Los atributos brindan información adicional sobre los elementos.
- ▶ No es parte del dato.
- ▶ Por ejemplo:
`<file type="gif">computer.gif</file>`
- ▶ Los valores de atributos deben aparecer entre comillas simples o dobles
`<gangster name='George "Shotgun" Ziegler'>`

Cuando usar atributos?

Estructura de un documento XML

Atributos	Elementos
<pre><persona sexo="femenino"> <nombre>Ana</nombre> <apellido>Gomez</apellido> </persona></pre>	<pre><persona> <sexo>femenino</sexo> <nombre>Ana</nombre> <apellido>Gomez</apellido> </persona></pre>

- No hay reglas definidas en la especificación de cuando usar atributos y cuando elementos.

“Use atributos

cuando la información que contiene ES sobre el DATO”

“Use elementos hijos

cuando la información que contiene ES un DATO”

Estructura de un documento XML

XML

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<nota Id = "1">
  <titulo> mensaje </titulo>
  <remitente> de: Juan </remitente>
  <dirigidaA> a: María </dirigidaA>
  <titulo> Recordatorio </titulo>
  <cuerpo>No te olvides de mi! </cuerpo>
</nota>
```

XML

mensaje
de: Juan
a: María
Recordatorio
No te olvides de mi!

Ejemplo de XML

XML

```
<?xml version="1.0"?>
<products>
  <product id="PRO001">
    <name>Coca Cola</name>
    <price>4</price>
  </product>
  <product id="PRO002">
    <name>Pepsi</name>
    <price>3</price>
  </product>
  <product id="PRO003">
    <name>Number One</name>
    <price>2</price>
  </product>
</products>
```


Modelo de arbol de documentos

XML: ejemplo

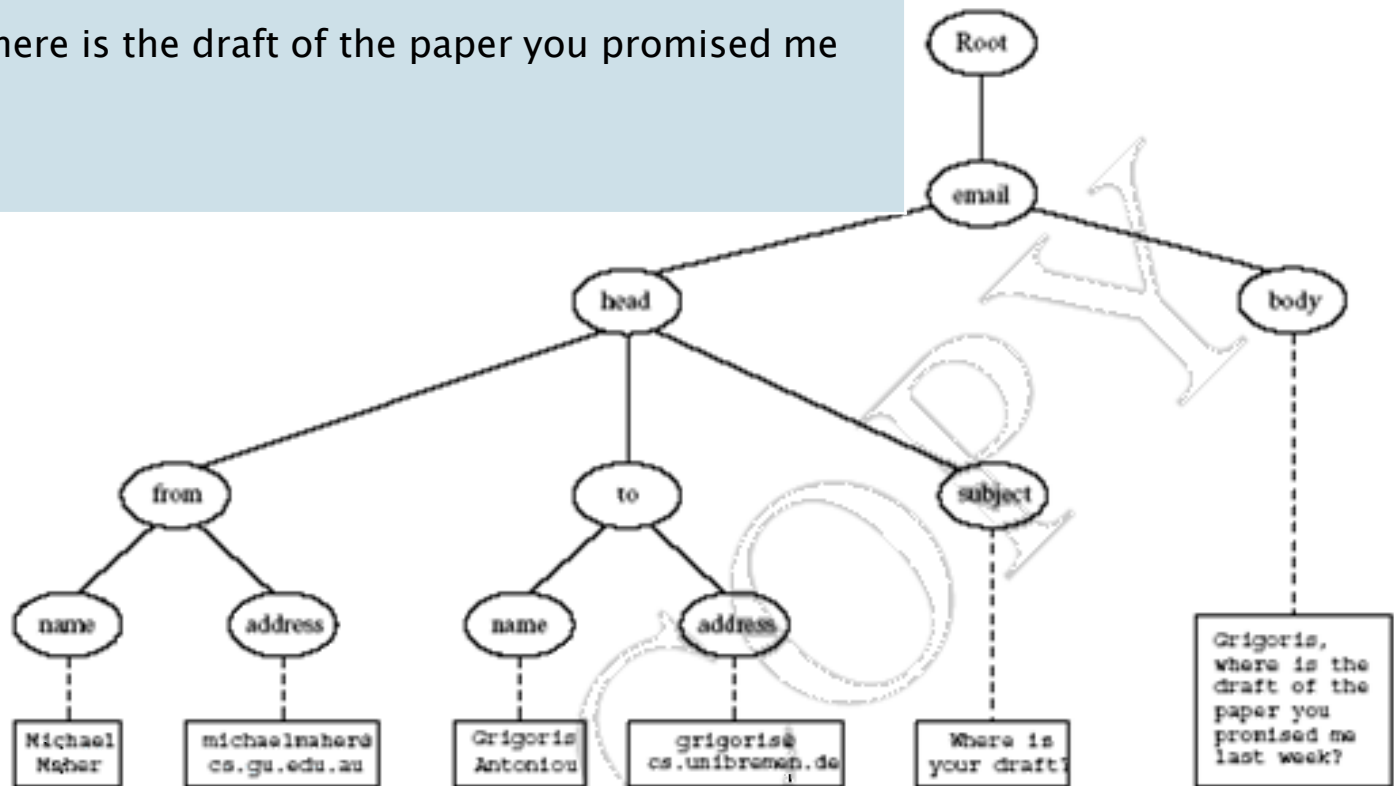
XML

```
<email>  
  <head>  
    <from name="Michael Maher"  
      address="michaelmaher@cs.gu.edu.au"/>  
    <to name="Grigoris Antoniou"  
      address="grigoris@cs.unibremen.de"/>  
    <subject>Where is your draft?</subject>  
  </head>  
  <body>  
    Grigoris, where is the draft of the paper you promised me  
    last week?  
  </body>  
</email>
```

Modelo de arbol XML

XML

```
<email>
  <head>
    <from name="Michael Maher"
address="michaelmaher@cs.gu.edu.au"/>
    <to name="Grigoris Antoniou"
address="grigoris@cs.unibremen.de"/>
    <subject>Where is your draft?</subject>
  </head>
  <body> Grigoris, where is the draft of the paper you promised me
last week?
  </body>
</email>
```



Modelo de arbol de documentos

XML: ejemplo

- ▶ La representacion jerarquica de un documento XML es un arbol etiquetado:
 - Una raiz
 - Sin ciclos
 - Cada nodo no-raíz tiene exactamente un padre
 - Cada nodo tiene una etiqueta
 - El orden de los elementos es importante
 - ... pero el orden de los atributos no es importante

XML no es un reemplazo para HTML

- ▶ XML es un complemento para HTML.
- ▶ En la mayoría de las aplicaciones web, XML es utilizado para transportar datos, mientras HTML se utiliza para formatear y mostrar datos.
- ▶ XML es una herramienta independiente del software y hardware para transportar información.
- ▶ XML es tan importante para la Web como HTML lo fue para la fundación de la Web.
- ▶ XML está en todos lados. Es la herramienta más común para la transmisión de datos entre todo tipo de aplicaciones.

XML — Intercambio de datos

XML se utiliza para intercambiar datos

- ▶ Formato de texto
- ▶ Software-independiente, hardware-independiente
- ▶ Intercambio de datos entre sistemas incompatibles, **si se corresponden con la misma definición de etiquetas.**
- ▶ Puede ser leído por diferentes aplicaciones

Beneficios:

- ▶ Reduce la complejidad de la interpretación de los datos
- ▶ Mayor facilidad para expandir y actualizar un sistema