

Ontologías y Web Semántica

Universidad Nacional del Sur

Pablo R. Fillottrani

Depto. Ciencias e Ingeniería de la Computación
Universidad Nacional del Sur

marzo 2019



Objetivos

- ▶ en este curso nuestro objetivo es introducir las diversas [tecnologías para la gestión de información](#) en las que se basan los conceptos de Web Semántica y Linked Data
- ▶ en particular, revisaremos XML, RDF y OWL
- ▶ analizaremos aplicaciones, en particular servicios web, vocabularios y ontologías existentes y en desarrollo
- ▶ y presentaremos propuestas para estándares, como los [lenguajes de reglas](#) y los [servicios web semánticos](#)



Contenidos

Presentación del Curso

Introducción a la Web Semántica

Web Semántica: tecnologías

Ontologías y OWL



Cronograma

- ▶ Introducción (2 horas)
- ▶ XML (2 horas)
- ▶ RDF (3 horas)
- ▶ Linked Data (1 horas)
- ▶ Ontologías (3 horas)
- ▶ OWL (2 horas)
- ▶ Ingeniería Ontológica y Herramientas (3 horas)



Recursos

► página web del curso

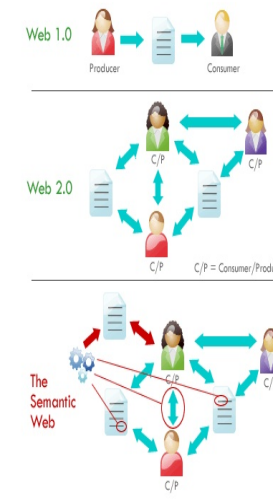
www.cs.uns.edu.ar/~prf/teaching/OWS

► Bibliografía:

- *Semantic Web Primer* 3rd edition, Grigoris Antoniou and Frank van Harmelen. MIT Press, 2012.
- *A Developers Guide to the Semantic Web* 2nd edition, Liyang Yu. Springer, 2014.
- *Linked Data: Structured Data on the Web*, David Wood, Marsha Zaidman, Luke Ruth. Mannig, 2014.
- *Linked Data: Evolving the Web into a Global Data Space*, Tom Heath, Christian Bizer. Morgan&Claypool, 2011.
- *Weaving the Web*, T. Berners-Lee. Harper, 1999.



Antecedentes: evolución de la web



Orígenes de la Web Semántica

- Tim Berners Lee, un físico británico y pionero de internet, propuso una visión extendida de la web
*"... si la interacción entre personas e hipertexto puede hacerse intuitiva tal que la **información procesada** represente las interacciones, patrones de trabajo y un estado de los pensamientos humanos, entonces el **análisis automático** sería una herramienta realmente poderosa, buscando patrones en nuestro trabajo y facilitando la solución problemas típicos."*
- él (y otros) han estado trabajando desde entonces para concretar esta visión, a la que se ha denominado como **Web Semántica**
- esto es probablemente muy difícil por el momento, pero han desarrollado un plan para lograr gradualmente este objetivo



La Web "sintáctica" actual

- es un grafo de recursos (páginas web, imágenes, sonido, video) conectados por enlaces (links)
- (Goble 2003) es un lugar dónde las computadoras hacen la presentación (fácil) y las personas hacen la interpretación (difícil)
 - **una biblioteca digital de hipermedios** documentos (páginas web) interconetados entre sí
 - **una base de datos y plataforma para aplicaciones** un portal común accesible mediante páginas web, y cuyos resultados también son páginas web
 - **un esquema de nombres** que otorga identificación única a esos documentos
 - ¿porqué no hacer que las computadoras hagan más del trabajo duro?



Hoy es casi imposible...

- ▶ realizar consultas complejas que involucren **conocimiento contextual**
 - ▶ encontrar información sobre “políticas de e-gov en países con población menor a 10 millones”
- ▶ ubicar información en los **repositorios de datos**
 - ▶ itinerarios de viajes
 - ▶ precios y condiciones de compra de productos y servicios
 - ▶ resultados de experimentos científicos en un área en común
- ▶ encontrar y usar “servicios web”
- ▶ delegar tareas complejas a “agentes”



El problema de la representación de los datos

- ▶ una página web típica consiste de
 - ▶ marcado para **mostrar información** (e.g., tipo de letra, color, tamaño)
 - ▶ enlace para **relacionar contenido**. El contenido semántico es accesible a los humanos pero no directamente a las computadoras...
- ▶ la información que una computadora puede leer

```
★++f12adf=?:@|5~pdf¬%l"%v*^(p2e]  
|#°vb2ñ]|fμ"æðđ"et6%)0~{-+çæ@»"ww1!  
»c°ð*■4gaf=¿dhkñpwwelþøeiföp←zcx3  
%þø|2#5¬a7[q|g-<&+ð■jkł~lkoty]
```



El problema de la representación de los datos

- ▶ una página web típica consiste de
 - ▶ marcado para **mostrar información** (e.g., tipo de letra, color, tamaño)
 - ▶ enlace para **relacionar contenido**. El contenido semántico es accesible a los humanos pero no directamente a las computadoras...
- ▶ (Rector and Horrocks 2004) la información que vemos

Universidad Nacional del Sur
Departamento de Ciencias de la Computación
Cursos Docentes
Investigación Contacto



El problema de la representación de los datos

- ▶ una página web típica consiste de
 - ▶ marcado para **mostrar información** (e.g., tipo de letra, color, tamaño)
 - ▶ enlace para **relacionar contenido**. El contenido semántico es accesible a los humanos pero no directamente a las computadoras...
- ▶ ¿solución? incluir marcado XML con etiquetas “significativas”

```
<uni>★++f12adf=?:@|5~pdf¬%l"%v*^(p2e]</uni>  
<facultad>|#°vb2ñ]|fμ"æðđ"et6%)0~{-+çæ@</facultad>»"ww1!  
<docentes>»c■4gaf=</docentes>¿dhkñp<cursos>øöp←zcx3</cursos>  
<investig>%þø|2#5¬a7[q|g-<&+ð■jkł~lkoty]
```



El problema de la representación de los datos

- ▶ una página web típica consiste de
 - ▶ marcado para **mostrar información** (e.g., tipo de letra, color, tamaño)
 - ▶ enlace para **relacionar contenido**. El contenido semántico es accesible a los humanos pero no directamente a las computadoras...
- ▶ de todas formas las computadoras sólo ven

```
<rb-/jð>*++f12adf=:@|5~pdf¬%1"%v*^(p½e|</rb-/jð>  
<½#rv>|°vb2ñ]|fμ"æðd"et6%)0~{-+çæ@</½#rv>»"ww1!  
<1@"fq>»c°ð*■4gaf=</1@"fq>¿dhkñp<u763ñ>øeiföp<-zc×3</u763ñ>  
<132x■o>%pø|2#5¬a7[</132x■o>q|g-<&yy87>+d■j</yy87>î~ty|
```



...pero es difícil

- ▶ los lenguajes de ontologías son **tramposos**
Todos los lenguajes tratables son de poco uso; todos los lenguajes útiles son intratables
- ▶ crear ontologías es **difícil**
 - ▶ las personas lo hacen fácilmente; pero las intuiciones son difíciles de formalizar
- ▶ el problema ha sido analizado por 3000 años: ontologías en filosofía, en lingüística
- ▶ la web semántica hace que la representación de conocimiento realmente interese



Necesidad de agregar semántica

- ▶ es necesario un **acuerdo externo** sobre el significado de las anotaciones
 - ▶ ejemplo, el Dublin Core para etiquetar información bibliográfica o de publicaciones: acuerdo en el significado de las etiquetas
 - ▶ hay problemas con esta solución: inflexible, sólo se puede expresar un limitado conjunto de cosas
- ▶ usar **ontologías** para especificar el significado de las anotaciones
 - ▶ las ontologías proveen un vocabulario de términos
 - ▶ nuevos términos pueden ser formados combinando los existentes
 - ▶ el significado (semántica) de los términos está formalmente definido
 - ▶ se pueden especificar también relaciones entre los términos de diferentes ontologías

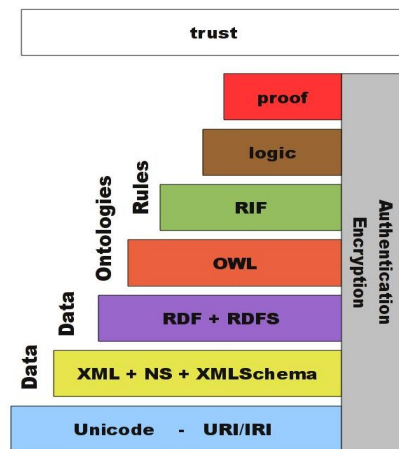


El camino

- ▶ la evolución puede lograrse no con una web separada, sino con una extensión de la web actual en la cual se le da a la información un significado preciso, que permita que las computadoras y las personas trabajen cooperativamente
- ▶ esta evolución se obtiene con un conjunto de tecnologías y lenguajes de marcado



La “torta” de la Web Semántica original

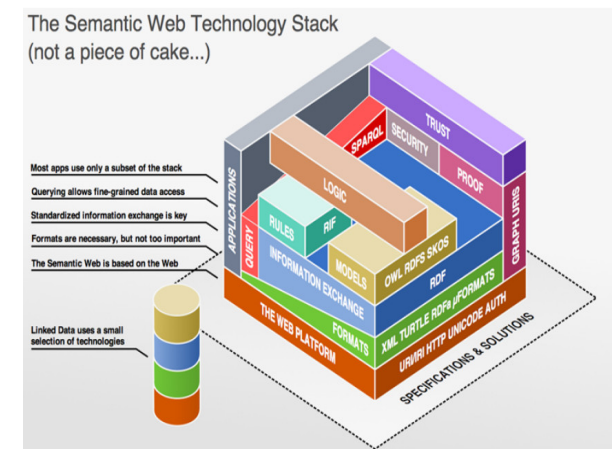


Los principios

- ▶ las siguientes características deben ser preservadas:
 - ▶ la Web es **distribuida** múltiples fuentes de información, inconsistencia
 - ▶ la Web es **dinámica** las necesidades de representación pueden cambiar
 - ▶ la Web es **enorme** los sistemas deben escalar bien
 - ▶ la Web es **incompleta**
- ▶ la Web Semántica propone
 - ▶ una **semántica clara** para la integración de información
 - ▶ mecanismos flexibles para la **actualización** de las representaciones
 - ▶ un **razonamiento eficiente**, que permita opearar con grandes cantidades de datos
 - ▶ tolerancia a la **información incompleta**; una transición gradual desde la web “sintáctica” actual



El “amontonamiento” de la Web Semántica actual



URI, URL, URN

- ▶ a **Uniform Resource Identifier** (URI) es una cadena de texto usada para identificar un nombre on un recurso en la Web
- ▶ la sintaxis de un URI consiste en un nombre de esquema (tales como http, ftp, mailto o file) seguido de “:”, y luego una parte específica del esquema
- ▶ un URI puede clasificarse como **ubicador** (URL), o **nombrador** (URN), o ambos
- ▶ en otras palabras: el URN define la identidad, mientras que URL provee un método para ubicar
- ▶ los URI's proveen el tipo de dato elemental de la Web



¿Qué es XML?

- ▶ XML representa **eXtensible Markup Language**, es una recomendación de la W3C desde 1998
- ▶ es un **metalenguaje de marcado**, pero es más flexible y adaptable
- ▶ su objetivo de diseño es **estructurar, almacenar y transporar datos**, no visualizarlos
- ▶ las etiquetas XML no están predefinidas, el usuario debe definir las
- ▶ XML es auto-descriptivo, pero también procesable por computadoras
- ▶ XML no es algo especial. Es simplemente texto. Cualquier software que puede procesar texto puede procesar XML



Validación

- ▶ un **documento XML bien formado** es uno que tiene una sintaxis correcta, de acuerdo a las reglas anteriores
- ▶ este control es muy débil. Generalmente cada usuario tiene sus propias reglas adicionales para generar y procesar documentos XML
- ▶ existen dos lenguajes que permiten especificar estas reglas:
 - ▶ **DTD** es un lenguaje incluido en XML, pero que no sigue su sintaxis y es poco expresivo
 - ▶ **XML Schema** es una recomendación adicional a XML, provee tipo de datos básico y constructores de nuevos tipos
- ▶ las librerías XML pueden controlar que un documento, además de bien formado, sea **válido** con respecto a reglas DTD y/o XML Schema
- ▶ existen otras tecnologías útiles relacionadas con XML: XML Namespaces, XPath, XQuery, etc



Un pequeño ejemplo

```
<email> <from>prf@cs.uns.edu.ar</from>  
<to>ece@cs.uns.edu.ar</to>  
<date> <year>2010</year>  
<month>May</month><day>20</day></date>  
<subject>Reminder</subject>  
<body>Don't forget about course slides!</body>  
</email>
```

- ▶ las etiquetas XML son “creadas” por el autor, no tienen un significado especial
- ▶ las etiquetas deben estar parentizadas apropiadamente, y el documento completo debe ser encerrado en una sola etiqueta
- ▶ esto genera que los documentos XML tengan necesariamente una estructura de árbol



¡XML y XML Schema no son suficientes!

- ▶ XML Schema no es un lenguaje para definiciones semánticas (no puede definir `tieneHijo` como la inversa de `tienePadre`, o que `tienePadre` es lo mismo que `hijoDe`)
- ▶ XML + XML Schema funcionan bien en documentos de texto. Se puede referenciar URLs, pero no es inherente al modelo de datos
- ▶ la validación de un XML Schema es sólo un control sintáctico, no se genera conocimiento implícito de los datos
- ▶ la semántica es dejada para los programas que procesan los documentos, lo cual es malo porque esta semántica puede cambiar y evolucionar haciendo necesarios cambios en los programas
- ▶ se necesita un lenguaje para escribir definiciones semánticas, de forma que las mismas también sean procesables por las computadoras

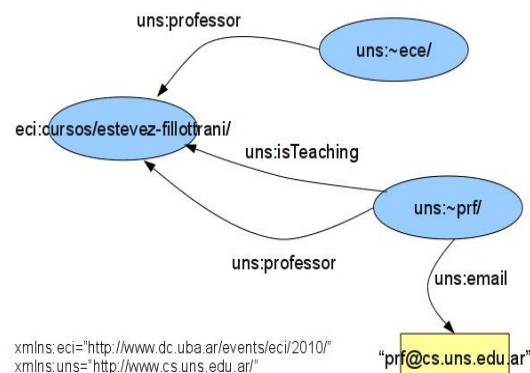


Introducción a RDF

- ▶ **Resource Description Framework** (RDF) es un mecanismo para la descripción de recursos en la web, junto con sus relaciones
- ▶ RDF está diseñado para ser procesable por computadoras, no para ser presentado a las personas
- ▶ básicamente describe el modelo de datos **grafo**, con una sintaxis en XML y una semántica formal
- ▶ es una recomendación W3C desde 2004, actualizada en 2014
- ▶ usa un vocabulario extensible basado en URIs para referenciar recursos



Ejemplo de base de datos RDF



El modelo de datos de RDF

- ▶ el modelo de datos de RDF describe pseudo-grafos dirigidos etiquetados, lo que significa:
 - ▶ **pseudo-grafo** porque hay nodos conectados por arcos, y puede haber varios arcos que conectan los mismos nodos
 - ▶ **dirigidos** todo arco tiene una dirección
 - ▶ **etiquetados** todo arco tiene una etiqueta
- ▶ tanto arcos como nodos pueden estar etiquetados con URIs; el nodo destino de un arco puede ser también un literal
- ▶ los URIs denotan **recursos**, cualquier cosa de la que se puede decir algo
- ▶ una base de datos RDF es solamente un conjunto sin orden de sentencias que describen cada arco
- ▶ estas sentencias se conocen también como triplas porque están compuestas por: un **sujeto** (nodo origen), un **predicado** (arco) y un **objeto** (nodo destino)



Ejemplo de base de datos RDF (cont)

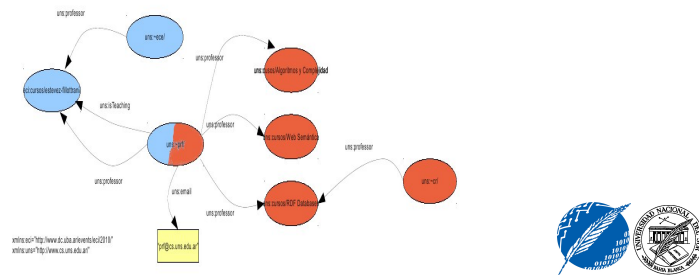
- ▶ la base de datos anterior es equivalente al conjunto de las siguientes sentencias

```
uns:~prf--uns:professor ->eci:courses/estevez-fillottrani/
uns:~ece--uns:professor-->eci:courses/estevez-fillottrani/
uns:~prf--uns:isTeaching->eci:courses/estevez-fillottrani/
uns:~prf--uns:email----->"prf@cs.uns.edu.ar"
```
- ▶ los predicados son URIs, por lo que pueden referenciar recursos como cualquier otro URI
- ▶ existen distintas sintaxis para representar las triplas, la oficial para intercambio de información está basada en XML



Integración de bases de datos RDF

- ▶ RDF está diseñado con una idea fuertemente descentralizada. Todo dato especificado en RDF puede reusarse, en el mismo o en distinto modelo
- ▶ la propiedad fundamental del modelo de datos RDF es que es posible combinar cualesquiera base de datos en RDF.
- ▶ la semántica y los URIs cumplen un papel fundamental para ello



RDF Schema

- ▶ RDF describe recursos y sus propiedades
- ▶ pero RDF necesita también una forma de definir **clases** específicas de aplicaciones, es decir recursos que comparten características en común
- ▶ esto se hace con extensiones a RDF, como **RDF Schema (RDFS)**
- ▶ RDF Schema es básicamente un conjunto de sentencias RDF para definir clases y propiedades de clases
- ▶ además de RDFS, existen otras tecnologías relacionadas con RDF, como SPARQL



Usos actuales de RDF

- ▶ **metadatos en la Web metadata** describe información sobre recursos en la Web
- ▶ **aplicaciones** que requieren modelos de información abiertos mas bien que restringidos
- ▶ **interoperabilidad de datos entre aplicaciones** combinando datos desde distintas fuentes para producir nueva información
- ▶ **procesamiento automático de información en la Web mediante agentes de software**
- ▶ ejemplos concretos:
 - ▶ propiedades de items en venta
 - ▶ descripción de información sobre páginas web (contenido, autor)
 - ▶ contenido y catalogo de imágenes
 - ▶ contenido para motores de búsqueda
- ▶ de esta forma, el valor de la información se incrementa a medida que se hace accesible a más aplicaciones a lo largo de la web



Vocabulario RDF Schema

- ▶ el tipo de cosas que se puede decir con RDF y RDFS son por ejemplo:
 - ▶ este URI debe ser considerado de `rdf:type` de esta `rdfs:Class` o esta propiedad `rdf:Property`
 - ▶ este URI está definido por `rdfs:isDefinedBy` (metadatos)
 - ▶ esta clase es subclase de esta otra clase `rdfs:subClassOf`
 - ▶ esta propiedad es subpropiedad de esta otra `rdfs:subPropertyOf`
 - ▶ esta propiedad conecta recursos de `rdfs:domain` con recursos de `rdfs:range`



Desde los vocabularios hasta las ontologías

- ▶ la integración de información necesita acuerdos:
 - ▶ en los términos, ej. “traductor”, “autor”
 - ▶ en las categorías, ej. “persona”, “literatura”
 - ▶ en las relaciones entre ellos, ej. “un autor es una persona”, “la ficción histórica es una clase de ficción”
 - ▶ en la forma en que nuevas relaciones pueden ser deducidas
- ▶ existe entonces una necesidad de **lenguajes** para definir vocabularios y asignar significado claro a sus relaciones explícitas e implícitas: esto se denomina **ontologías**



¡RDFS no es suficiente!

- ▶ RDFS es un lenguaje de definición de ontologías, existen las clases, subclases, propiedades, subpropiedades y tipos de datos
- ▶ RDFS es suficiente para muchos vocabularios útiles, pero no para todos:
 - ▶ sólo permite relaciones binarias
 - ▶ no permite definir características de las propiedades (ej. inversa, transitiva, simétrica)
 - ▶ no permite restricciones de rango locales (ej para la clase `Persona` la propiedad `tieneNombre` tiene dominio `xsd:string`)
 - ▶ no permite descripción compleja de conceptos (ej. `Persona` se define como la unión total y exclusiva de `Hombre` y `Mujer`)
 - ▶ no permite restricciones de cardinalidad (ej. toda `Persona` debe tener por lo menos un `nombre`)



Definición de ontologías

- ▶ formalmente, hay dos definiciones para el concepto de **ontología** en el área representación del conocimiento y sistemas de información:
 - ▶ *especificación explícita y forma de una conceptualización compartida de un dominio*
 - ▶ *especificación consistente de definiciones de Clases, Relaciones entre clases, Individuos y Axiomas*



Las ontologías y la Web Semántica

- ▶ las ontologías establecen el esqueleto de la Web Semántica: definen el vocabulario básico para las anotaciones y permite el razonamiento con el conocimiento contextual
- ▶ están basadas en lenguajes formales, y permiten cubrir la brecha entre el significado de los humanos y el de las máquinas
- ▶ en base a las limitaciones de RDFS, tres tecnologías han surgido
 - ▶ definición de vocabularios más complejos: **OWL**
 - ▶ reuso de tesaurus, glosarios, etc: **SKOS**
 - ▶ framework genérico para definir reglas que se aplican a términos y datos: **RIF**



OWL

- ▶ con esta motivación, y teniendo en cuenta el compromiso entre expresividad del lenguaje y tratabilidad en el razonamiento, se introdujo el lenguaje **Web Ontology Language (OWL)** en base a experiencias previas como DAML+OIL
- ▶ es recomendación de la W3C desde febrero de 2004. En octubre 2009 se aprobó una revisión que se conoce como **OWL2**
- ▶ la especificación reconoce tres variantes de OWL, con distinto grado de expresividad
 - ▶ **OWL Lite** jerarquías de clasificaciones y restricciones limitadas
 - ▶ **OWL DL** máxima expresividad manteniendo razonamiento eficiente
 - ▶ **OWL full** preserva compatibilidad con la semántica de RDFS, pero con razonamiento intratable
- ▶ con sintaxis formal basada en RDF/XML, y varias sintaxis abstractas más “human-oriented”



Ejemplo

- ▶ las propiedades, cuando se aplican una después de otra, pueden ser subsumidas por una nueva propiedad:
si la persona «P» nació en la ciudad «A» y «A» está en el país «B», entonces «P» nació en el país «B»

```
ex:nació_en_país owl:propertyChainAxiom
    (ex:nació_en_ciudad ex:ciudad_en_país)
```
- ▶ se pueden usar más de dos constituyentes para la nueva propiedad, pero hay restricciones para evitar especificaciones circulares



Componentes ontológicos de OWL

- ▶ **clases**: grupos de individuos con propiedades en común
 - ▶ **clases intensionales**: Persona, Universidad
 - ▶ **clases extensionales**:
ColorSemáforo = {rojo, amarillo, verde}
- ▶ **relaciones**: conexiones entre individuos que pueden ser asignadas a las clases: tieneNombre, dicta, esDeColor
- ▶ **individuos**: objetos en el dominio, que pueden ser instancias de clases
- ▶ **axiomas**: sentencias adicionales sobre el dominio, especificadas en un lenguaje lógico: tieneNombre debe tener exactamente un valor



Ejemplo

- ▶ en notación abstracta

```
Class( firstYearCourse partial
    restriction (isTaughtBy allValuesFrom ( Professor )))
Class(mathCourse partial
    restriction (isTaughtBy hasValue (949352)))
Class(academicStaffMember partial
    restriction (teaches someValuesFrom (undergraduateCourse)))
Class(course partial restriction (isTaughtBy
    minCardinality (1)))
Class(department partial restriction (hasMember
    minCardinality(10))
    restriction (hasMember maxCardinality(30)))
```



Representación del Conocimiento

- ▶ para que la Web Semántica funcione, las computadoras tienen que tener acceso a **fuentes de información estructurada** y **conjuntos de reglas de inferencia**
- ▶ en conjunto estos elementos requerirá **razonamiento automatizado**
- ▶ investigadores de la IA han analizado el tema mucho antes de que WWW sea desarrollada



- ▶ los sistemas tradicionales basados en Representación del Conocimiento han sido generalmente **centralizados**
- ▶ requiriendo que todos los usuarios compartan exactamente la misma definición de conceptos comunes
- ▶ el problema es que al incrementar el tamaño y el contexto de estos sistemas, el control central se hace inmanejable
- ▶ los sistemas tradicionales usualmente limitan cuidadosamente las preguntas que pueden ser respondidas de forma confiable por la computadora
- ▶ la Web Semántica, por el contrario, acepta paradojas y preguntas sin respuesta al precio de lograr **versatilidad**



- ▶ el área de **Representación del Conocimiento** está actualmente en un estado comparable con el del hipertexto antes de la WWW
- ▶ es claramente una **buena idea**, pero no ha cambiado el mundo todavía
- ▶ para realizar todo su potencial debe estar vinculada a un nico sistema global de información: la Web Semántica.



- ▶ esta filosofía es similar a la de la WWW convencional
- ▶ al principio, los detractores de la WWW señalaban que nunca podría convertirse en una biblioteca organizada de información
- ▶ decían con la ausencia de una base de datos centralizada y una estructura arbórea, uno nunca puede estar seguro de encontrar todo lo que se busca
- ▶ tenían razón al respecto



- ▶ el poder expresivo de la WWW hace posible que vastas cantidades de información estén disponibles
- ▶ los motores de búsqueda pueden producir índices completos del material disponible
- ▶ el desafío de la Web Semántica es, por lo tanto, proveer un lenguaje que exprese tanto datos como reglas para razonamiento sobre los datos
- ▶ de forma que las reglas que ya existan en sistemas de representación del conocimiento puedan ser fácilmente exportadas



- ▶ dos tecnologías importantes para el desarrollo de la Web Semántica ya están disponibles
 - ▶ eXtensible Markup Language (XML)
 - ▶ the Resource Description Framework (RDF)
- ▶ XML permite crear tags propios (etiquetas ocultas para describir secciones de texto en una página)
- ▶ de forma que pequeños programas puedan usar los tags en manera sofisticada
- ▶ el problema es que el programador de estos programas debe conocer **qué quiso decir** con cada tag el diseñador de la página web
- ▶ luego, XML permite agregar estructura arbitraria a los documentos pero no dice nada sobre su significado



- ▶ incorporar **lógica** a la Web significa usar las reglas para
 - ▶ hacer inferencias
 - ▶ elegir cursos de acción
 - ▶ responder preguntas
- ▶ esta es la tarea en la que se encuentra actualmente la comunidad de la Web Semántica
- ▶ una mezcla de restricciones matemáticas, informáticas e ingenieriles complican la tarea
- ▶ la lógica debe ser **suficientemente expresiva** como para expresar propiedades complejas de los objetos
- ▶ pero no tanto como para que los agentes se bloqueen analizando paradojas



- ▶ el significado está expresado por RDF
- ▶ que codifica un conjunto de triplas (sujeto, verbo, objeto) para describir un grafo
- ▶ estas triplas pueden ser expresadas en XML
- ▶ un documento RDF establece aserciones sobre
 - ▶ objetos
 - ▶ propiedades entre objetos
 - ▶ valores específicos



Ontologías

- ▶ por ejemplo, supongamos que tenemos acceso a una variedad de bases de datos con información sobre personas, incluyendo sus direcciones
- ▶ si queremos encontrar todas las personas viviendo en un determinado zona de código postal, se necesita saber cuáles campos en cada bases de datos representan “nombre” y “codigo postal”
- ▶ RDF puede especificar “(field5 in database1) (isa field of type) (zip code)” usando URI en lugar de palabras



- ▶ una solución a este problema la da el tercer componente básico de la Web Semántica: las **ontologías**
- ▶ en filosofía, una ontología es una teoría sobre la naturaleza de la existencia, o sobre el tipo de cosas que existen
- ▶ en IA una ontología es un documento que define formalmente relaciones entre términos
- ▶ las ontologías típicas en la Web tienen una taxonomía y un conjunto de reglas de inferencia



- ▶ pero el problema es que dos bases de datos distintas pueden usar diferentes identificadores para el mismo concepto
- ▶ un programa que combina información entre estas dos bases de datos debe conocer que dos términos están siendo usados con el mismo significado
- ▶ idealmente, el programa debe descubrir estas semejanzas para cualquier base de datos que encuentre



- ▶ una **taxonomía** define clases de objetos y relaciones entre ellos
- ▶ por ejemplo, una **dirección** puede ser una subclase de **ubicación**, y un **código postal** puede ser una **componente de dirección**
- ▶ clases, subclases y relaciones entre entidades son una herramienta muy poderosa para usar en WWW



- ▶ podemos expresar un gran número de relaciones entre entidades asignando propiedades a clases, y permitiendo que las subclases hereden estas propiedades
- ▶ si un estudiante es una persona, y cada persona tiene una dirección, entonces podemos hablar de la dirección de un estudiante aun cuando ninguna base de datos vincule explícitamente estudiantes con direcciones



- ▶ las reglas de inferencia en una ontología proveen poder expresivo extra
- ▶ una ontología puede expresar la regla “si un código postal está asociado con un código de provincia, y una dirección usa ese código postal, entonces la dirección está en esa provincia”
- ▶ la máquina verdaderamente no “entiende” nada de esta información, pero puede manipular los términos más fácilmente en forma que sean útiles y significativas para el ser humano



- ▶ con las ontologías disponibles en la Web, las soluciones al uso de diferentes terminologías comienzan a aparecer
- ▶ el significado de términos, o tags XML, usados en la WWW puede ser definido con punteros al concepto definido en una ontología
- ▶ pero problemas similares aparecen si se hace referencia a conceptos de distintas ontologías
- ▶ esto puede resolverse si las ontologías proveen relaciones de equivalencia entre sí



- ▶ las ontologías pueden aumentar el poder de la WWW en varias formas
- ▶ por ejemplo, pueden ser usadas en forma simple para mejorar la exactitud de los buscadores WWW
- ▶ el programa de búsqueda solo considerará válidas aquellas páginas que se refieren a un concepto preciso, en lugar de todas aquellas que usan keywords ambiguos
- ▶ aplicaciones más avanzadas usarán ontologías para relacionar información en una página con las estructuras de conocimiento asociadas, y sus reglas de inferencia



- ▶ este lenguaje de marcado hará mucho más fácil desarrollar programas ([agentes](#)) que puedan resolver problemas complicados cuyas respuestas no pertenecen a una única página web
- ▶ el potencial de la Web Semántica sólo se realizará cuando los usuarios puedan crear fácilmente programas que junten información de la WWW de diversas fuentes, la procesen e intercambien sus resultados con otros programas
- ▶ la efectividad de tal software se incrementará exponencialmente a medida que más contenido “machine friendly” sea disponible en la WWW, y que sean disponibles [servicios automatizados](#) (incluyendo otros agentes)



Conclusiones

- ▶ la Web Semántica está basada en tres capas: sintaxis, semántica y ontologías
- ▶ no está bien caracterizado aun los alcances de cada una de estas capas, pero existen estándares que pretenden definir las
- ▶ a medida que se generalice el uso de estas capas, se podrá definir agentes y servicios web automáticos que exploten esta información

