

Vistas RDF

Carlos Hurtado L.

Trabajo conjunto con Claudio Gutiérrez.

Paper en trabajo: www.dcc.uchile.cl/~churtado/vistasrdf.ps

Contenido

- Web Semántica
- Problema
- Vistas RDF
- Buscador de RDF basado en Vistas
- Desafíos

Cultura de Bases de Datos vs. Cultura de la Web

Bases de Datos

- **Esquema**
 - Modelo abstracto de los datos, estructura lógica.
 - En bases de datos relacionales: nombres de relaciones y atributos.
- Consultas se expresan en términos del esquema

Web

- **Ausencia de esquema**, estructura lógica ignorada
 - **Hipertexto**
- Consultas textuales (ej., buscadores) y navegación

Referencia: [Database Techniques for the Web](#) (Florescu, Mendelzon, Levy, 98)

Ontología

Esquema universal, comprendido y utilizado por una comunidad amplia de individuos (o agentes)

Ejemplo: **Dublin Core**

“Breve ontología para describir recursos” (Baker, 2000)

Contenido	Propiedad Intelectual	Instanciación
Coverage	Contributor	Date
Description	Creator	Format
Type	Publisher	Identifier
Relation	Rights	Language
Source		
Subject		
Title		

Web Semántica y RDF

Web Semántica

- Web de datos con las siguientes características:
 - Hablan sobre otros datos (**metadatos**)
 - Sus esquemas son ontologías

Propuestas W3C:

- Metadatos: RDF

Creator(<http://www.w3.org/People/Lassila>, “Ora Lassila”)

- Ontologías: RDFS

Generación de RDF

Gran parte de los metadatos para la Web semántica ya existen en la Web.

- Ejemplo:

Creator(<http://www/w3.org/People/Lassila>, “Ora Lassila”)

Es difícil extraer RDF de páginas Web (Cream [Handschuh and Staab, WWW2002])

¿Qué sucede cuando los datos residen en **fuentes de datos**?

Fuentes de Datos en la Web

Fuente de Datos:

- Recurso Web con mecanismo de consulta
- Retorna datos estructurados en lugar de texto
- Tendencias:
 - implementación: [servicios Web](#)
 - descripción: [WSDL](#)

www.invisible-web.net: directorio con más de 10.000 fuentes:

- Librerías en línea, bases de datos de comercio electrónico, páginas blancas y amarillas, buscadores, etc.
- Ejemplos: Servicios Web de Amazon, Google, IMDb, etc.

Fuentes de Datos (cont.)

Raghavan, García-Molina. *Crawling the Hidden Web*. VLDB'01.

- 550.000 M de páginas dinámicas (Servicios Web potenciales)
- 500 veces la Web navegable
- Más del 50% de las visitas totales en la Web

Problema

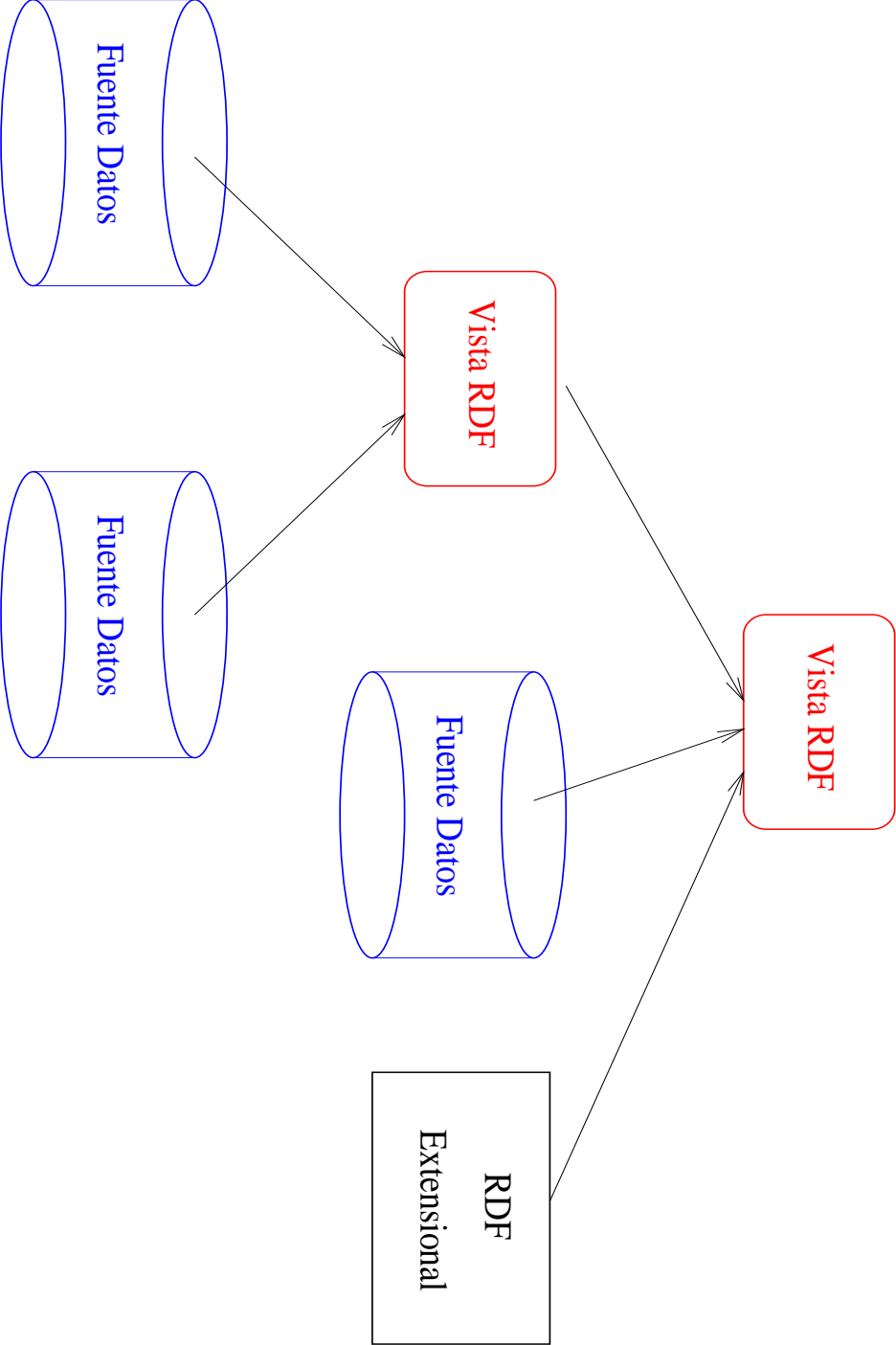
- RDF fue creado para describir propiedades de páginas Web.
 - Una descripción RDF es un conjunto de **proposiciones simples**. RDF Semantics, W3C working draft, P. Hayes.
Creator(<http://www.w3.org/People/Lassila>, “Ora Lassila”)
- Para extraer metadatos de fuentes, necesitamos incorporar noción de **vista** en RDF.
 - Metadatos definidos por **expresiones**

Propuesta

Vistas RDF

- Expresiones (consultas) que generan metadatos desde fuentes de datos, otras vistas o RDF común.
- También son metadatos (descripciones RDF). Por esto:
 - son universales (instancias de una ontología),
 - permiten generación descentralizada de metadatos.

Vistas RDF



Ejemplo 1: Descripción Dublin Core de un Libro

```
Namespace(source, http://...)
Namespace(dc, http://purl.org/metadata/dublin_core)
dc:Creator(urn:isbn:0-15-144647-4, "Umberto Eco")
dc>Title(urn:isbn:0-15-144647-4, "The Name of the Rose")
dc:Description(urn:isbn:0-15-144647-4, y) ←
    y = source:BookDatabaseService.bookAbstract.get("0-15-144647-4").value
```

Ejemplo 2: Extracción de Descripción Dublin Core desde Dos Fuentes de Datos

Namespace(SourceA, http://...)

Namespace(SourceB, http:// ...)

Namespace(dc, http://purl.org/metadata/dublin_core)

dc:Creator(x,y) ← y=SourceA:Service.Book.get(x).author

dc:Title(x,u) ← u=SourceA:Service.Book.get(x).title

dc:Description(x,y) ← y=SourceB:Service.bookAbstract.get(x).value

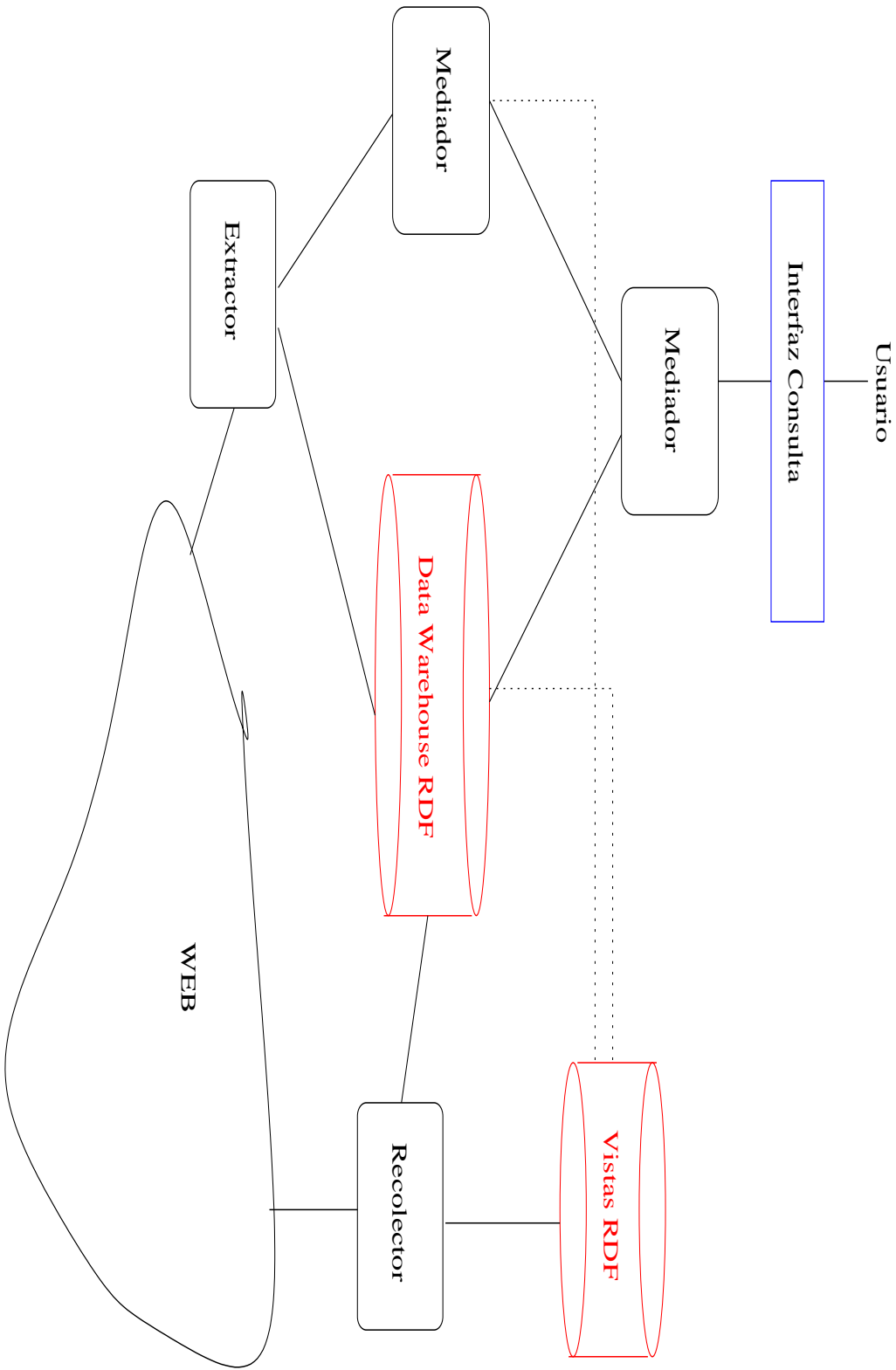
Ejemplo 3: Recomendaciones de URLs en la Web

Los primeros 30 URLs retornados por Google a la consulta “Neruda” y los URLs referenciados por éstos por enlaces intra-sitio, contienen información confiable sobre Pablo Neruda

```
NameSpace(authOnto: http://...)
NameSpace(Google: http://...)
```

```
GoogleAuthority(x, “Pablo Neruda”) ←
    x=Google:Service.pageSearch.get(“Neruda”, y).URI, y ≤ 30
GoogleAuthority(x, “Pablo Neruda”) ← GoogleAuthority(x, “PabloNeruda”)
GoogleAuthority(y, “Pablo Neruda”) ← GoogleAuthority(x, “PabloNeruda”), x ~ y
```

Buscador de RDF Basado en Vistas



Trabajo Relacionado

APIs para extracción de RDF.

- Ontobroker
- KAON
- InfoHarness
- Observer

Problemas:

- Codificados en lenguajes ad-hoc (no universales)
- Envoltorios no son metadatos
 - Generación centralizada de metadatos
- No permiten integración de datos (o metadatos)

Desafíos

- RDF (V) : ontología para vistas RDF
 - Sintaxis y semántica
 - Reificación de vistas
 - Permiten reificar “partes” de fuentes de datos
- Implementación
 - Recolector (crawler)
 - Mantenedor/Extractor
 - Data warehouse
 - Mediadores