

Uso de Tecnologías Semánticas para la Integración de Recursos de Información en una Memoria Corporativa

Erik Alarcón-Zamora
R. Carolina Medina-Ramírez
Héctor Pérez-Urbina

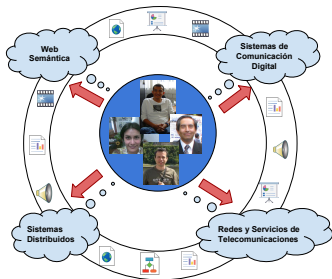
Universidad Autónoma Metropolitana
Departamento de Ingeniería Eléctrica
Iztapalapa, México

29 de octubre de 2013

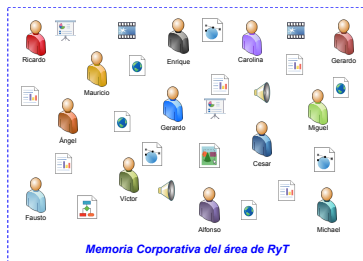
- Introducción.
- Integración semántica de recursos:
 - Casos de uso.
 - Representación del conocimiento.
 - Explotación del conocimiento.
 - Consulta del conocimiento.
- Prototipo de interfaz de usuario gráfica.
- Experimentación.
- Conclusiones.

Memoria Corporativa

*Una **memoria corporativa** (MC) es una representación explícita, consistente y persistente del conocimiento en una organización [1].*



(a) Conocimiento



(b) Memoria Corporativa

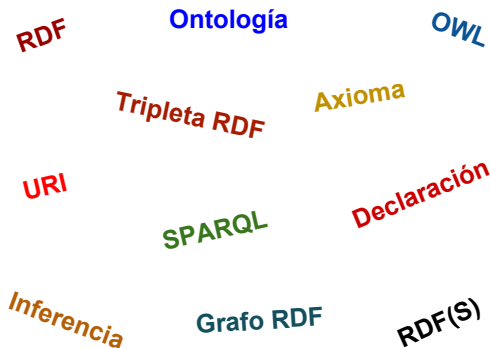
La *integración de los recursos* es el proceso de búsqueda y recuperación significativa de información existente en los recursos, para responder una consulta dada por un usuario.

Las actividades básicas para la integración de los recursos son:

- Representar el conocimiento de los recursos en un modelo.
- Consultar información en el modelo.

Tecnologías Semánticas

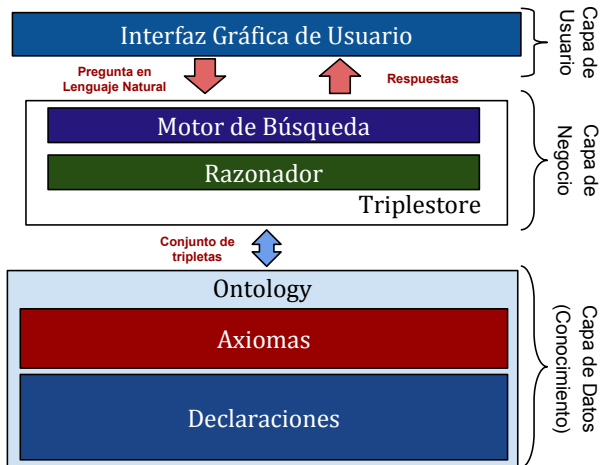
Las *tecnologías semánticas* son un conjunto de metodologías, lenguajes, aplicaciones, herramientas y estándares para suministrar u obtener el significado de las palabras, información y las relaciones entre éstos.



El uso de las tecnologías semánticas para realizar el proceso de integración de información para los recursos de una memoria corporativa.

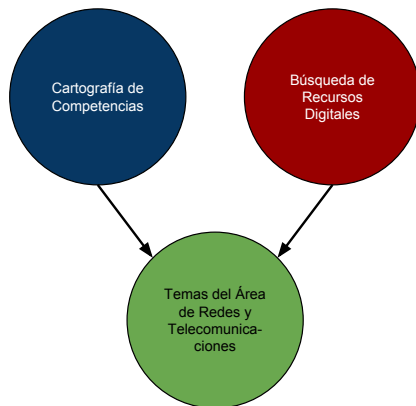
Las etapas de esta integración semántica son:

- 1 representar el conocimiento explícito de los recursos en un formato estándar.
- 2 enriquecer el conocimiento implícito mediante la introducción de axiomas.
- 3 buscar y recuperar la información en este modelo para responder una pregunta dada.



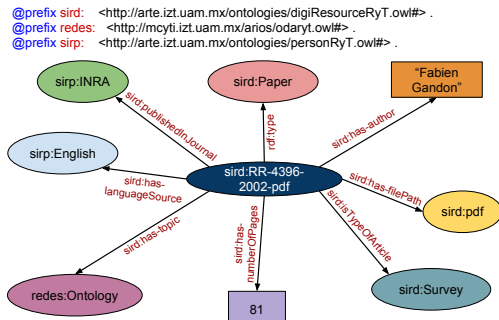
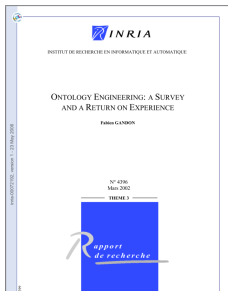
- ***Cartografía de Competencias*** consiste en la búsqueda y recuperación de información significativa de las personas, a partir de las características personales y profesionales de las mismas.
- ***Búsqueda de Recursos Digitales*** consiste en la búsqueda y recuperación de información significativa de los documentos y archivos multimedia a partir del contenido de los mismos

*Una **ontología** es una especificación formal y explícita de una conceptualización compartida [2].*



Representación del Conocimiento

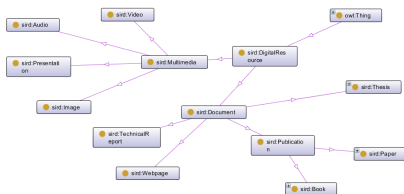
Representación del conocimiento explícito a partir de las características significativas y relaciones de los mismos.



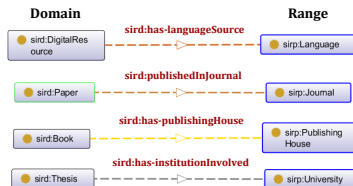
Las herramientas para la transformación de información a triples es una combinación de Formularios y Scripts.

Enriquecimiento del Conocimiento

Introducción de axiomas para enriquecer el conocimiento explícito en el grafo RDF.



(e) Jerarquía de Clases

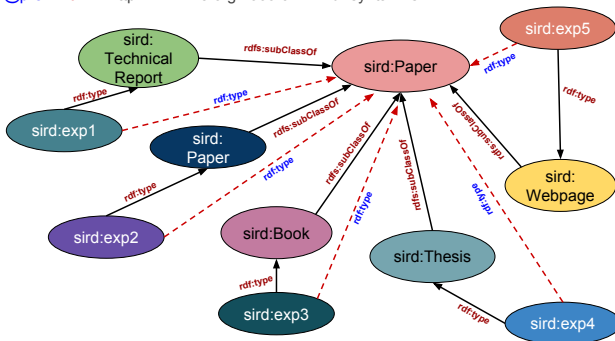


(f) Dominio y Rango de las propiedades

Protégé es la herramienta que se emplea para representar los axiomas.

Un razonador es un programa para inferir hechos o asociaciones a partir del conocimiento existente (axiomas y propiedades).

@prefix sird: <http://arte.izt.uam.mx/ontologies/digiResourceRyT.owl#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .



Lenguaje SPARQL es la especificación para consultar, recuperar y modificar la información de ternas RDF.

```
SELECT DISTINCT ?title ?path
WHERE
{
  {?x rdf:type sird:TechnicalReport.} UNION
  {?x rdf:type sird:Book.} UNION
  {?x rdf:type sird:Paper.} UNION
  {?x rdf:type sird:Thesis} UNION
  {?x rdf:type sird:Webpage.}
  ?x sird:has-topic redes:Ontology;
    sird:has-title ?title;
    sird:has-filePath ?path.
}
```

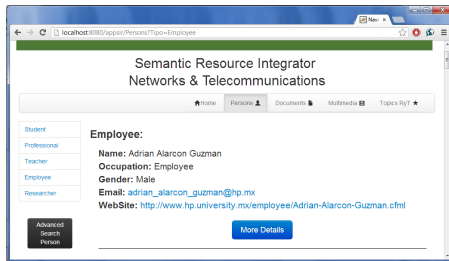
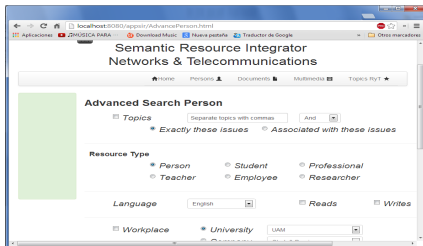
(g) Consulta sin razonamiento

```
SELECT DISTINCT ?title ?path
WHERE
{
  ?x rdf:type sird:Document;
    sird:has-topic redes:Ontology;
    sird:has-title ?title;
    sird:has-filePath ?path.
}
```

(h) Consulta con razonamiento

Apache Jena es una librería que permite el almacenamiento, carga, indexación, inferencia y consulta de/en tripletas RDF.

La interfaz gráfica de usuario permite la interacción transparente y amigable del usuario con el modelo semántico.



En la integración semántica de recursos, se tienen dos criterios de evaluación para el triplestore Jena:

- Calidad de los resultados (sin inferencia vs con inferencia).
- Tiempo medio de procesamiento (sin inferencia vs con inferencia).

Id. Consulta	Pregunta	No. de Recursos
Q1	¿Cuáles son los títulos, rutas, extensión, idioma de todos los recursos digitales de RyT?	1330
Q2	¿Cuáles libros tratan sobre algunos temas de Sistemas Distribuidos?	103
Q3	¿Qué recursos fueron publicados por la UAM?	18
Q4	¿Qué documentos son para dar un curso de Sistemas P2P?	31
Q5	¿Qué recursos multimedia son mayores al año 2009?	119
Q6	¿Cuáles documentos tratan sobre Ontologías?	30
Q7	¿Qué recursos fueron publicados en una Revista científica?	156
Q8	¿Qué recursos tienen en su contenido las palabras "linked data"?	159
Q9	¿Cuáles documentos en inglés y mayores al año 2000 son de autoría de Erik Alarcón Zamora?	2
Q10	¿Cuáles la tesis de Samuel Hernández Maza?	4

Resultados

Id. Consulta	Modelo (ABox)		Modelo (Razonador+Ontología)	
	Tiempo promedio (ms)	No. Recursos	Tiempo promedio (ms)	No. Recursos
Q1	12	1330/1330	138	1330/1330
Q2	10	0/103	194	103/103
Q3	8	18/18	406	18/18
Q4	28	15/31	129	31/31
Q5	7	66/119	157	119/119
Q6	9	15/30	4016	30/30
Q7	12	156/156	3520	156/156
Q8	16	159/159	3472	159/159
Q9	42	0/2	3451	2/2
Q10	13	3/4	3312	4/4

- Beneficios en el uso de las tecnologías semánticas para la integración de los recursos de información en una memoria corporativa.
- Modelo con inferencia se tienen mejores resultados, pero el tiempo de procesamiento se incrementa.

- [1] Rose Dieng, Olivier Corby, Alain Giboin, and Myriam Ribi  re.
Methods and Tools for Corporate Knowledge Management.
Technical Report RR-3485, INRIA, September 1998.
- [2] Thomas R. Gruber.
A translation approach to portable ontology specifications.
Knowl. Acquis., 5(2):199–220, June 1993.