

Integración semántica de los recursos de información en una memoria corporativa

Erik Alarcón Zamora

Enero 2014. México, D.F.

Asesores:

Dra. Reyna Carolina Medina Ramírez

Dr. Héctor Pérez Urbina

Contenido

- 1 Marco Introdutorio
 - Contexto y Motivación
 - Descripción del Problema

- 2 Integración Semántica de una Memoria Corporativa
 - Marco de Referencia
 - Arquitectura de la Integración Semántica
 - Casos de Uso
 - Representación el Conocimiento

- 3 Referencias

Memoria Corporativa

Definición

La representación explícita, tácita, consistente y persistente del conocimiento de una organización. [Gandon, 2002]

Finalidad

Una memoria corporativa conserva y mantiene el conocimiento de una organización [Dieng et al., 1998], para *facilitar el acceso, intercambio y difusión de éste*.

Caso de Estudio

El grupo de investigación del área de Redes y Telecomunicaciones (RyT) de la Universidad Autónoma Metropolitana Unidad Iztapalapa (UAM-I).

Memoria Corporativa

Naturaleza en los Recursos de Información

- Diversidad en el Formato:
pdf, doc, odp, html, txt, xsl, wav, png, mp3, mp4, mpeg, mov, ppt, flv, por mencionar algunas.
- Diversidad en el Contenido:
p2p, middleware, estado global, sistema operativo, replicación, concurrencia, sincronización, entre otras.
- Diversidad en la Estructura:
datos estructurados, semi-estructurados y sin estructura.
- Significado de la Información
 - Homonimia: **radio** tiene distintos significados que se asocian a la Química, Comunicación, Anatomía o Geometría.
 - Sinonimia: **resumen, sumario, síntesis y recapitulación.**

Tecnologías Semánticas

Definición

Un conjunto de metodologías, lenguajes, aplicaciones, herramientas y estándares para suministrar u obtener el significado de las palabras, información y las relaciones entre éstos. [Alfred et al., 2010]



Tecnologías Semánticas

Ontología

Una especificación formal y explícita de una conceptualización compartida.

[Gruber, 1993]

Objetivos de una ontología

- 1 Construir un vocabulario conceptual, formal y consensuado para un dominio dado.
- 2 Tener un conjunto de reglas para combinar los conceptos y relaciones, así como para componer expresiones complejas en el vocabulario.
- 3 Un vocabulario para construir descripciones y comunicar hechos.
- 4 Interpretar sin ambigüedad el conocimiento y vocabulario de un dominio dado.
- 5 Intercambiar y reutilizar el conocimiento para diferentes propósitos.
- 6 Inferir información a partir de un programa especializado (razonador) y los hechos en una ontología.

Tecnologías Semánticas

Componente Asertivo (ABox)

Representa el conocimiento e información explícita en los recursos del dominio. Este componente está constituido por las declaraciones (descripciones o hechos verdaderos) de los recursos que afirman que los individuos son instancias de una clase o propiedad.

Componente Terminológico (TBox)

Representa el conocimiento implícito en los recursos del dominio. Este componente describe las clases y propiedades relevantes, así como los axiomas que permiten aprovechar la manera en que las instancias se relacionan entre sí.

Tecnologías Semánticas

Resource Description Framework (RDF)

Marco de trabajo para representar el conocimiento e información de los recursos en un formato estándar. [Bouzid et al., 2012]

Recurso

Cualquier *persona, lugar, documento, objeto abstracto o físico* que tiene un identificador único de recursos (URI).

Propiedad

Un aspecto significativo, característica, o relación que se describe de un recurso (relación binarias).

Clase

Una colección de objetos que comparten características comunes.

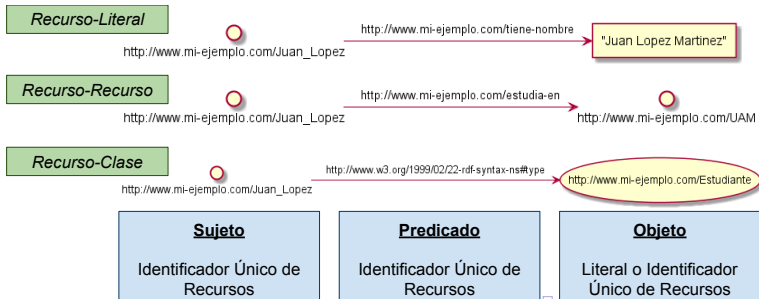
Tecnologías Semánticas

Declaración

Una afirmación de un hecho explícito de un recurso, en términos de una propiedad y el valor asignado a ésta.

Tripleta RDF

La forma básica para representa una declaración en un modelo semántico.

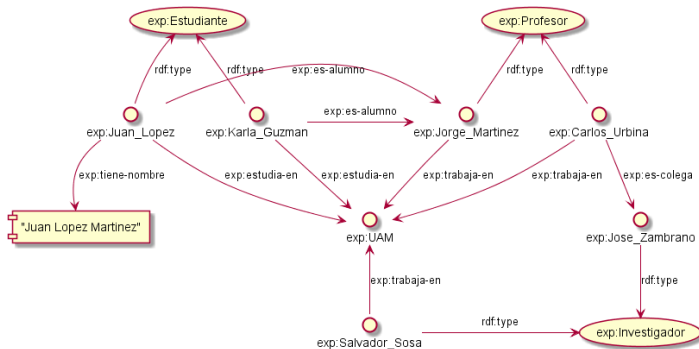


Grafo RDF

Un grafo estructurado y dirigido compuesto por nodos, aristas y etiquetas para representar las tripletas.

@prefix exp: <http://www.mi-ejemplo.com/>

@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>



Tecnologías Semánticas

SPARQL

Lenguaje de consulta y protocolo de acceso a RDF [?], para la búsqueda y recuperación de la información en un grafo RDF.

Motor de Búsqueda

Programa que interpreta una consulta SPARQL, compara la consulta con el grafo RDF y recupera los valores de la consulta.

Tecnologías Semánticas

Reglas de inferencia o Axiomas

Los *axiomas* o *reglas de inferencia* [Gruber, 1993] son expresiones para enriquecer el conocimiento explícito en un grafo RDF.

Funcionalidad Axiomas

Describir relaciones entre clases, definir propiedades en términos de otras, definir relaciones entre conceptos, definir restricciones de cómo las propiedades se relacionan, por mencionar algunos.

Razonador

Un programa que deduce declaraciones a partir de los axiomas y declaraciones explícitas en la ontología.

Integración Semántica

Definición

La búsqueda y recuperación significativa de información existente en los recursos de información para responder una consulta dada por un usuario.

Etapas

- 1 Representar el *conocimiento* de los *recursos de información* en un *modelo semántico*.
- 2 Buscar y recuperar información existente en la memoria corporativa mediante la interrogación del modelo semántico.

Objetivos

Objetivo Principal

Contribuir a la integración semántica de los recursos de información en una memoria corporativa, mediante el uso de las tecnologías semánticas.

Objetivos Particulares

- 1 Desarrollar una **marco de referencia** para la *integración semántica* de los *recursos de información* existentes en una *memoria corporativa*.
- 2 Implementar un **modelo semántico** que representa el *conocimiento explícito e implícito* de los *recursos de información*.
- 3 Implementar un **prototipo de interfaz gráfica de usuario** que permita a los usuarios una interacción amigable para la integración semántica de los recursos de información.
- 4 Evaluar los **resultados devueltos** y **tiempos de procesamiento** en la *integración semántica* para el dominio de redes y telecomunicaciones.

Metodología

Marco de Referencia

- 1 Identificar los *casos de uso* para encontrar los principales *recursos de información* existentes en la memoria, así como los criterios de búsqueda asociados a éstos.
- 2 Construir el diagrama de casos de uso.
- 3 Evaluar herramientas semánticas para: edición de descripciones semánticas, edición de reglas de inferencia, gestión de modelos semánticos.
- 4 Recopilar los recursos de información de acuerdo a los casos de uso.
- 5 Adquirir el conocimiento o información de los recursos de información con base en las características y relaciones de los mismos.
- 6 Construir el diagrama de clases.

Metodología

Marco de Referencia

Modelo Semántico

- 7 Describir el conocimiento explícito de los *recursos de información* recopilados en un modelo semántico.
- 8 Identificar las reglas de inferencia a introducir en el modelo, con base en el diagrama de clases.
- 9 Escribir las reglas de inferencia para enriquecer el modelo semántico con conocimiento implícito, mediante el uso del editor de reglas de inferencia.
- 10 Identificar las preguntas en lenguaje natural a partir de los casos de uso.
- 11 Diseñar las consultas en el *lenguaje estándar de búsqueda* que correspondan a las preguntas en lenguaje natural.

Metodología

Marco de Referencia

- 12 Emplear un proceso que permita hacer explícito el conocimiento implícito.
- 13 Buscar y recuperar información en la memoria corporativa, interrogando el modelo semántico.

Prototipo de interfaz gráfica de usuario

- 14 Diseñar un prototipo para interacción (búsqueda y navegación) amigable y transparente de los usuarios de la memoria con el modelo semántico.
- 15 Proponer funcionalidades básicas del prototipo.
- 16 Indicar cuáles son las interfaces para los usuarios (pantallas).
- 17 Describir las especificaciones de estas interfaces.
- 18 Implementar el prototipo y realizar pruebas del mismo.

Metodología

Evaluar los resultados devueltos

- 19 Evaluar la calidad de los resultados (recursos relevantes recuperados) con y sin inferencia, mediante el uso de métricas que se emplean en la recuperación de la información: exhaustividad y precisión.
- 20 Identificar aquellos recursos (total de recursos relevantes) que responden las preguntas del paso 10 de este listado.
- 21 Consultar al modelo semántico y comparar los recursos relevantes recuperados con los recursos relevantes que se identificaron en el paso 20 de este listado.
- 22 Calcular la exhaustividad y precisión.

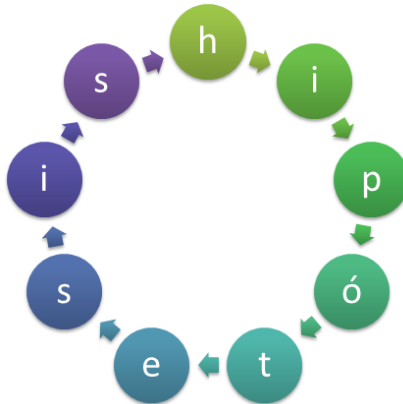
Metodología

Evaluar los tiempos de procesamiento

- 23 Evaluar los *tiempos promedios* que toma la herramienta electa de gestión de los modelos semánticos, para consultar los modelos con/sin inferencia.
- 24 Elaborar un script que calcule '*n*' veces el *tiempo de procesamiento* al consultar un modelo semántico (con o sin inferencia). Las consultas se hacen a las preguntas identificadas del paso 10 de este listado.

Hipótesis

¿Las **tecnologías semánticas** son viables para solucionar la **integración** de los **recursos de información** de una **memoria corporativa**?



Aportaciones

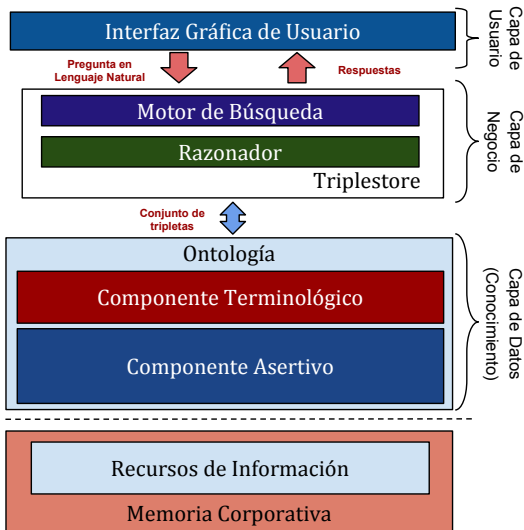
- 1 Un *marco de referencia* para lograr la integración semántica de recursos de información.
- 2 Un modelo semántico que representa el conocimiento de una memoria corporativa, el cual tiene tres ramas principales (Personas, Recursos Digitales y Conceptos del Redes y Telecomunicaciones).
- 3 Un prototipo (interfaz gráfica de usuario) para la interacción amigable (búsqueda y consulta de información) de los usuarios al modelo semántico.
- 4 Los resultados de nuestra evaluación experimental.
- 5 Un par de scripts para la generación automática y controlada de descripciones (conocimiento explícito) de los *recursos de información*, con el fin de poblar la base de conocimiento.

Marco de Referencia

Etapas

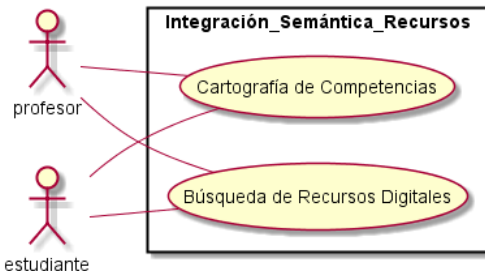
- 1 ***Representación del conocimiento explícito de los recursos*** consiste en identificar los recursos de información de la memoria corporativa, así como representar las características y relaciones (conocimiento explícito) de estos recursos en un modelo semántico.
- 2 ***Enriquecimiento del conocimiento en el modelo*** consiste en introducir reglas de inferencia (axiomas) para completar y enriquecer el modelo semántico con conocimiento implícito del dominio de la memoria corporativa.
- 3 ***Búsqueda y recuperación de la información en el modelo*** consisten en identificar las principales consultas de los usuarios, así como interrogar el modelo semántico para recuperar la información que responda a estas consultas.

Arquitectura de la Integración Semántica



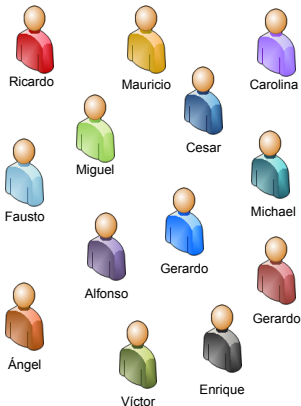
Casos de Uso

- ***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.

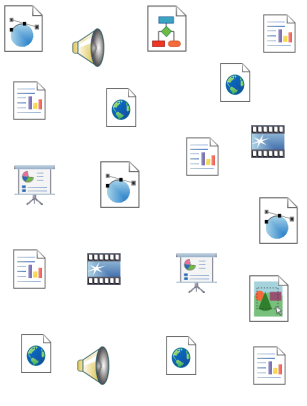


Identificar los principales recursos de información

Cartografía de Competencias



Búsqueda Recursos Digitales



Memoria Corporativa del área de RyT

[illegible]

Representar el conocimiento e información mediante el estándar RDF

Resource Description Framework (RDF)

RDF es un estándar para representar el conocimiento e información de los recursos en un formato estándar. [Bouزيد et al., 2012]

Actividades en la representación del conocimiento

- 1 Asignar un identificador único de recursos para cada recurso de información en la memoria corporativa.
- 2 Asignar los identificadores únicos de recursos a las propiedades.
- 3 Reconocer los valores de las propiedades: otro recurso o literal.
- 4 Generar las tripletas RDF asociadas a las descripciones de los recursos de información.

Representar el conocimiento e información mediante el estándar RDF

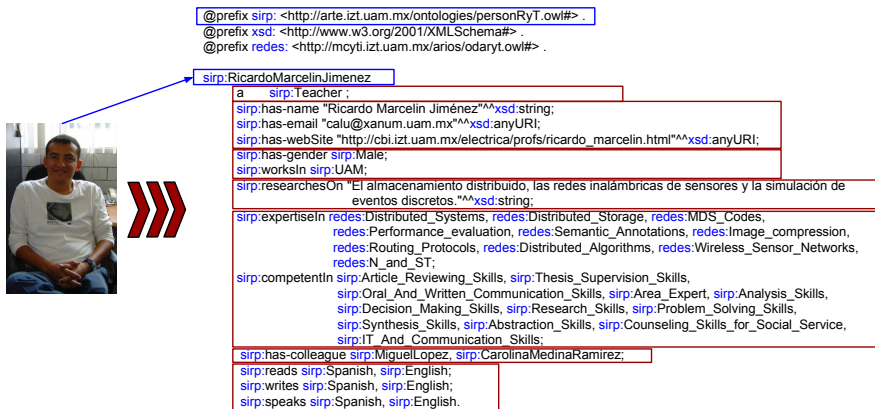


Figura : Ejemplo de declaraciones del Dr. Ricardo Marcelin Jiménez en forma de tripletas RDF

Representar el conocimiento e información mediante el estándar RDF



```
@prefix sird: <http://arte.izt.uam.mx/ontologies/digiResourceRyT.owl#> .  
@prefix sirdp: <http://arte.izt.uam.mx/ontologies/personRyT.owl#> .  
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .  
@prefix redes: <http://mcyti.izt.uam.mx/arios/odaryt.owl#> .
```

sird:What-isLinkedData-2012-flv

```
a sird:Video ;  
sird:has-title "What is Linked Data?"^^xsd:string ;  
sird:has-author "Manu Sporny"^^xsd:string ;  
sird:has-filePath "http://www.youtube.com/watch?v=4x_xzT5eF5Q"^^xsd:anyURI ;  
sird:has-fileExtension sird:flv ;  
sird:has-languageSource sirdp:English ;  
sird:has-topic redes:Linked_data, redes:Data, redes:Web, redes:Semantic,  
redes:Knowledge_graph ;  
sird:has-synopsis "A short non-technical introduction to Linked Data, Google's  
Knowledge Graph, and Facebook's Open Graph Protocol"  
^^xsd:string ;  
sird:has-yearOfCreation "2012"^^xsd:int .
```

Figura : Ejemplo de declaraciones del vídeo "What is Linked Data?" en forma de tripletas RDF

Referencias I

- [Alfred et al., 2010] Alfred, S., Arpah, A., Lim, L. H. S., and Sarinder, K. K. S. (2010). Semantic technology: An efficient approach to monogenean information retrieval. In *Computer and Network Technology (ICCNT), 2010 Second International Conference on*, pages 591–594.
- [Bouzid et al., 2012] Bouzid, S., Cauvet, C., and Pinaton, J. (2012). A survey of semantic web standards to representing knowledge in problem solving situations. In *Information Retrieval Knowledge Management (CAMP), 2012 International Conference on*, pages 121–125.
- [Dieng et al., 1998] Dieng, R., Corby, O., Giboin, A., and Ribière, M. (1998). Methods and Tools for Corporate Knowledge Management. Technical Report RR-3485, INRIA.
- [Gandon, 2002] Gandon, Fabien, L. (2002). Ontology Engineering: a Survey and a Return on Experience. Technical Report RR-4396, INRIA.
- [Gruber, 1993] Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowl. Acquis.*, 5(2):199–220.