Contexto y Motivación Descripción del Problem Estado del Arte Integración Semántica de una Memoria Corporativa Prototipo (Aplicación) Evaluación Conclusiones Referencias

# 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 I

- Contexto y Motivación
  - Memoria Corporativa
  - Integración Semántica
- Tecnologías Semánticas
- Descripción del Problema
  - Pregunta Investigación
    - Objetivos
    - Metodología
  - Hipótesis
  - Aportaciones
- Estado del Arte
  - Integración Semántica de Recursos de Información
  - Herramientas para la Integración Semántica de Recursos de Información
- 4 Integración Semántica de una Memoria Corporativa
  - Representación el Conocimiento
  - Enriquecer el conocimiento en el modelo semántico
  - Buscar y recuperar la información en el modelo semántico
- 5 Prototipo (Aplicación)
- 6 Evaluación
- Conclusiones
  - Referencias

# Memoria Corporativa

#### Definición

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

Referencias





(a) Conocimiento

(b) Memoria Corporativa

Memoria Corporativa Integración Semántica Tecnologías Semántica

# 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**

Representar el conocimiento de los recursos de información en un modelo semántico.

Buscar y recuperar información existente en la memoria corporativa mediante la interrogación del modelo semántico.

Memoria Corporativa Integración Semántica Tecnologías Semántica

# Heterogeneidad y Significado de la Información

Ξħ Diversidad en formato pdf, doc, odp, html, txt, xsl, wav, png, mp3, mp4, mpeg, mov, ppt, mov Ξū Diversidad en contenido p2p. middleware, estado global, replicación, concurrencia, sincronización Diversidad en estructurados semi-estructurados sin estructura estructura radio ---> Química, Comunicación, Anatomía o Geometría Homonimia herramienta = aparato = instrumento = mecanismo = artilugio Sinonimia

# 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]



Pregunta Investigación Objetivos Metodología Hipótesis Aportaciones

# Pregunta Investigación

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



Pregunta Investigació Objetivos Metodología Hipótesis Aportaciones

# 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

- Un marco de referencia para la integración semántica de los recursos de información.
- ② Un modelo semántico que representa el conocimiento explícito e implícito de los recursos de información.
- Un prototipo de interfaz gráfica de usuario que permita a los usuarios consultar y visualizar la información de los recursos de información, interrogando un modelo semántico.
- La evaluación de la calidad de los resultados recuperados y los tiempos de procesamiento de la integración semántica.

Pregunta Investigación Objetivos Metodología Hipótesis Aportaciones

# Metodología I

#### Marco de Referencia

- Identificar los casos de uso.
- Evaluar las herramientas semánticas.
- 3 Conformar los recurso de información de la memoria corporativa.

#### Modelo Semántico

- Representar el conocimiento explícito de los recursos de información en un modelo semántico (ontología).
- 5 Enriquecer el modelo semántico con reglas de inferencia.
- Escribir las principales consultas en la sintaxis correspondiente.
- © Emplear un razonador para hacer explícito el conocimiento implícito.
- Buscar y recuperar información en la memoria corporativa, interrogando el modelo semántico inferido.

Pregunta Investigación Objetivos Metodología Hipótesis Aportaciones

# Metodología II

### Prototipo de interfaz gráfica de usuario

Oconstruir el prototipo de interfaz de usuario para la (búsqueda y navegación) de los usuarios en un modelo semántico.

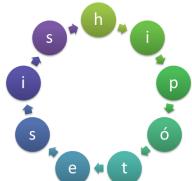
#### Evaluación

- Evaluar la calidad de los resultados con y sin inferencia.
- Evaluar los tiempos promedios de consulta sobre modelos con/sin inferencia.

Pregunta Investigació Objetivos Metodología Hipótesis Aportaciones

# Hipótesis

El uso de las tecnologías semánticas es adecuado para lograr la integración semántica de recursos de información en una memoria corporativa.



# **Aportaciones**

- Un marco de referencia para lograr la integración semántica de recursos de información.
- Un modelo semántico que representa el conocimiento de una memoria corporativa.
- Un prototipo (interfaz gráfica de usuario) para la interacción amigable (búsqueda y consulta de información) de los usuarios con el modelo semántico.
- 4 Los resultados de nuestra evaluación experimental.
- **1** Un par de scripts para la generación automática y controlada de descripciones (conocimiento explícito) de los *recursos de información*.

## Estado del Arte

#### Ejes claves

 Integración de la información a partir del uso de tecnologías semánticas.

Conclusiones Referencias

- ② Búsqueda, recuperación y publicación de la información desde una ontología.
- Gestión de una memoria corporativa.



# Integración de la Información

### [Moner et al., 2006]

Una arquitectura dual que representa la información clínica de cualquier persona, para que los profesionales de la salud accedan al historial clínico de las mismas.

## [Zhai et al., 2008]

Una arquitectura basada en ontologías y el lenguaje de marcado extensible, para la integración semántica en sistemas de información de energía eléctrica (sistemas heterogéneos).

## [Yang and Steele, 2011]

Un *marco de integración semántica* para la integración de la información en el dominio del alojamiento en-línea (información en constante cambio), con el fin de reunir y compartir esta información.

# Búsqueda y Recuperación de la Información

#### [Cao et al., 2011]

Un sistema para la recomendación y el acceso a la información turística en una ontología que define conceptos y propiedades del dominio del turismo.

#### [Ha et al., 2011]

Una propuesta para representar y recuperar información en una ontología sobre el contenido de los *manuales de mantenimiento electrónicos*, así como un sistema para visualizar la información en una ontología.

#### [Suganyakala and Rajalaxmi, 2013]

Un *marco de trabajo* para recuperar información en una ontología (Película) y una *interfaz de consulta* para escribir consultas en lenguaje natural.

#### [Salam, 2013]

Un método basado en ontologías para recuperar la *información contenida en un documento* mediante el uso de conceptos de una ontología.

# Gestión de una Memoria Corporativa

#### [Xin and Guangleng, 2001]

Un enfoque basado en las **ontologías**, para describir los *objetos de justificación del diseño*, así como acceder de manera uniforme a los *recursos de información*.

Referencias

#### [Chakhmoune et al., 2011]

Un entorno de colaboración para la gestión de memorias corporativas, cuya función es construir una memoria corporativa a partir de lluvia de ideas y un mecanismo de toma de decisiones consensuadas

#### Evaluación Conclusiones Referencias

## Comparativa

Autor	Dominio	Modelo	Tecnologías Semánticas	Integración de la Información	Búsqueda y Recuperación de la Información	Motor de Búsqueda e Inferencia
Moner et al.	Salud	Orientado a objetos y Arquetipos	No	Sí	No	No
K. Yang y R. Steele	Alojamiento en-línea	Ontología	Sí	Sí	No	No
Jun Zhai et al.	Electricidad	Ontología	Sí	Sí	No	No
Tuan-Dung et al.	Turismo	Ontología	Sí	No	Sí	No
Ha Inay et al.	Mantenimiento de aeronaves	Ontología	Sí	No	Sí	No
Suganyakala y Rajalaxmi	Películas	Ontología	Sí	No	Sí	No
Salam	Urología	Ontología	Sí	No	Sí	No
Xin y Guangleng	Justificación del diseño	Ontología	Sí	Sí	Sí	No
Chakhmoune et al.	Memoria Documental	Ontología	Sí	Sí	No	No

# Herramientas para la Integración Semántica de Recursos de Información

#### Descriptor Semántico de Recursos

Herramienta para crear y almacenar tripletas RDF, en varias sintaxis de serialización, a partir de la información explícita de los recursos de información. *OntoMat Annotizer*, *MnM*, *GATE* y *Aktive Media*.

### Editor de Ontologías

Herramienta que proporciona una serie de interfaces amigables para la construcción y mantenimiento de ontologías. *Protégé*, *pOWL*, *TopBraid Composer* y *SWOOP*.

#### **Triplestore**

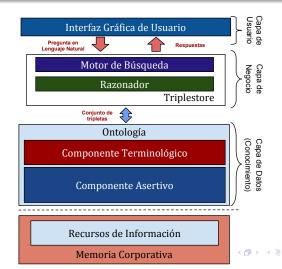
Programa para el almacenamiento e indexación de tripletas RDF, con el fin de permitir la consulta eficiente de información sobre estas tripletas. *Apache Jena*, *Stardog*, *4store* y *Sesame*.

## Marco de Referencia

#### Etapas

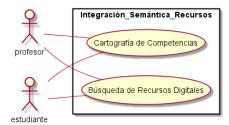
- Representar las características y/o relaciones de los recursos de información mediante el estándar RDF, para construir un modelo semántico.
- 2 Introducir *reglas de inferencia* en el modelo semántico, para enriquecer con *conocimiento implícito* de los *recursos de información* y del dominio de la memoria.
- Buscar y recuperar información en el modelo semántico para responder un conjunto consultas SPARQL.

# Arquitectura de la Integración Semántica



## Casos de Uso

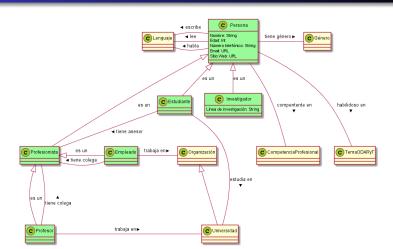
- Cartografía de Competencias es la búsqueda y recuperación de información significativa de las personas a partir de las características personales y profesionales de las mismas.
- Búsqueda de Recursos Digitales es la búsqueda y recuperación de información significativa de los documentos y archivos multimedia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



# Identificar los principales recursos de información



# Adquirir y expresar el conocimiento de los recursos de información



nriquecer el conocimiento en el modelo semantico Buscar y recuperar la información en el modelo semántico

# Representar el conocimiento e información mediante el estándar RDF

#### Definición

Marco genérico para describir el conocimiento e información explícita de los recursos mediante sus características y relaciones. [Bouzid et al., 2012]

#### Actividades en la representación del conocimiento

- Asignar un identificador único de recursos (URI) a cada recurso de información en la memoria corporativa.
- ② Asignar un URI a cada característica y/o relación (propiedad) de de los recursos de información.
- Generar las tripletas RDF asociadas a las descripciones de los recursos de información.

# Representar el conocimiento e información mediante el estándar RDF

```
@prefix sirp: <a href="http://arte.izt.uam.mx/ontologies/personRyT.owl">http://arte.izt.uam.mx/ontologies/personRyT.owl</a>.
@prefix xsd: <a href="http://www.w3.org/2001/XMLSchema#">http://www.w3.org/2001/XMLSchema#>.
```

@prefix redes: <a href="http://mcvti.izt.uam.mx/arios/odarvt.owl#">http://mcvti.izt.uam.mx/arios/odarvt.owl#>.

#### sirp:RicardoMarcelinJimenez sirp:Teacher:

sirp:has-name "Ricardo Marcelin Jiménez"^^xsd:string;

redes:N and ST:

sirp:has-email "calu@xanum.uam.mx"^^xsd:anvURI:

sirp:has-webSite "http://cbi.izt.uam.mx/electrica/profs/ricardo\_marcelin.html"^^xsd;anvURI:

sirp:has-gender sirp:Male: sirp:worksIn sirp:UAM:

sirp:researchesOn "El almacenamiento distribuido, las redes inalámbricas de sensores y la simulación de eventos discretos."^^xsd:string;

sirp:expertiseIn redes:Distributed Systems, redes:Distributed Storage, redes:MDS Codes,

redes:Performance evaluation, redes:Semantic Annotations, redes:Image compression. redes:Routing Protocols, redes:Distributed Algorithms, redes:Wireless Sensor Networks,

sirp:competentln sirp:Article Reviewing Skills, sirp:Thesis Supervision Skills,

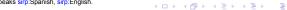
sirp:Oral And Written Communication Skills, sirp:Area Expert, sirp:Analysis Skills,

sirp:Decision Making Skills, sirp:Research Skills, sirp:Problem Solving Skills, sirp:Synthesis Skills, sirp:Abstraction Skills, sirp:Counseling Skills for Social Service,

sirp:IT And Communication Skills: sirp:has-colleague sirp:MiguelLopez, sirp:CarolinaMedinaRamirez;

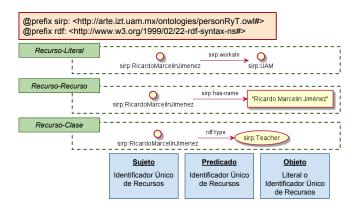
sirp:reads sirp:Spanish, sirp:English; sirp:writes sirp:Spanish, sirp:English:

sirp:speaks sirp:Spanish, sirp:English,



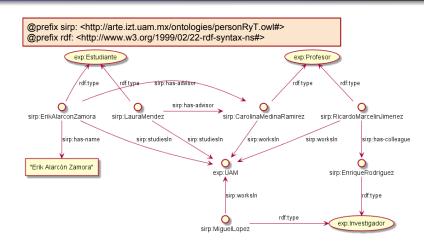
Enriquecer el conocimiento en el modelo semantico Buscar y recuperar la información en el modelo semántico

# Representar el conocimiento e información mediante el estándar RDF



Buscar y recuperar la información en el modelo semántico

# Representar el conocimiento e información mediante el estándar RDF



## Ontología

#### Definición

Una definición formal, explícita y compartida de los conceptos, así como las relaciones de un determinado dominio. [Gruber, 1993]

#### Componentes

- Componente Asertivo (ABox) está constituido por descripciones que afirman que los individuos son instancias de una clase o propiedad.
- Componente Terminológico (TBox) describe las clases y propiedades relevantes, así como las reglas de inferencia que permiten aprovechar la manera en que las instancias se relacionan entre sí.

## Axiomatización

#### Reglas de inferencia o Axiomas

Expresiones para enriquecer un grafo RDF con conocimiento implícito.

### Lenguajes

Especificaciones para describir clases, propiedades e individuos.

- RDF Schema RDF(S)
- Web Ontology Language OWL

@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>

@prefix owl: <http://www.w3.org/2002/07/owl#>

Lo que es obvio para un humano, no lo es para una maquina.

## Herencia de Clases I

#### Subclase (rdfs:subClassOf)

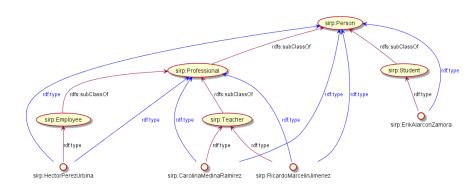
Afirma que una *clase A* se subsume por una *clase B*, es decir, la clase A es un caso particular de la *clase B*. En este caso, las instancias de la clase A son instancias de la clase B.



## Herencia de Clases II



## Herencia de Clases III



# Herencia de Propiedades I

## Subpropiedad (rdfs:subPropertyOf)

Afirma que todos los recursos que se relacionan por la *propiedad X*, también se relacionan por la *propiedad Y*.



# Herencia de Propiedades II



# Dominio y Rango en las Propiedades I

#### Dominio (rdfs:domain)

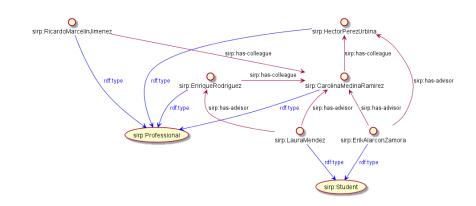
Especifica qué clase se aplica a una propiedad.

### Rango (rdfs:range)

Especifica los valores (clase o tipo de literal) que puede asumir una propiedad.



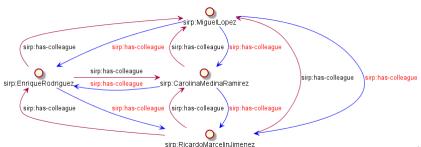
# Dominio y Rango en las Propiedades II



# Características en las propiedades

#### Propiedad simétrica (owl:SymmetricProperty)

Afirma que la propiedad X es su propia propiedad inversa, es decir, si la propiedad X relaciona al individuo A con el individuo B, entonces, esta propiedad debe relacionar al individuo B con el individuo A.



## Buscar y recuperar la información en el modelo semántico

#### SPARQL

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

¿Cuáles son los nombres y sitios Web de las personas que conocen a Carolina Medina



### Uso de inferencia I

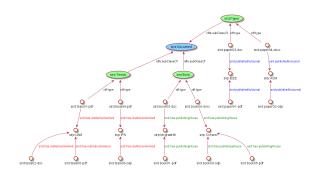


Figura: Grafo RDF sin inferencia

### Uso de inferencia II

```
?x
sird:tesis01-pdf
sird:tesis04-pdf
sird:book02-doc
sird:book04-docx
sird:paper01-pdf
sird:paper02-odp
```

(a) Consulta sin inferencia

(b) Resultados de la consulta

### Uso de inferencia III

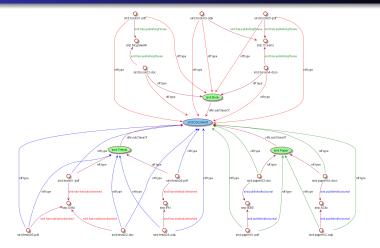


Figura : Grafo RDF con inferencia

## Uso de inferencia IV

PREFIX sird: <a href="http://arte.izt.uam.mx/ontologies/digiResourceRyT.owl#">PREFIX rdf: <a href="http://www.w3.org/1999/02/22-rdf-syntax-ns#">http://www.w3.org/1999/02/22-rdf-syntax-ns#</a>

SELECT ?x

WHERE

{ ?x rdf:type sird:Document.}

(a) Consulta con inferencia

?x
sird:tesis01-pdf
sird:tesis02-doc
sird:tesis02-dop
sird:tesis04-pdf
sird:tesis04-pdf
sird:tesis05-pdf
sird:book01-pdf
sird:book02-doc
sird:book03-odp
sird:book05-pdf
sird:paper01-pdf
sird:paper02-odp
sird:paper03-doc
sird:paper04-docx

(b) Resultados de la consulta Contexto y Motivación
Descripción del Problema
Estado del Arte
Integración Semántica de una Memoria Corporativa
Prototipo (Aplicación)
Evaluación
Conclusiones

# Prototipo (Aplicación)

#### Dificultad

La búsqueda y recuperación en un modelo semántico no son actividades sencillas, porque se requiere que un usuario tenga conocimientos en el uso de las tecnologías semánticas y los vocabularios en las ontologías.

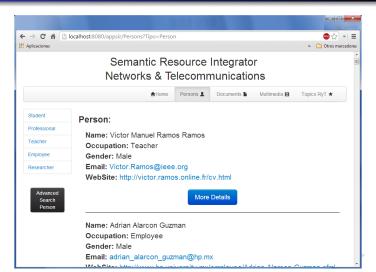
### Prototipo

Una aplicación Web para que los usuarios puedan consultar y visualizar la información de los *recursos de información*, mediante una ontología (Redes y Telecomunicaciones).

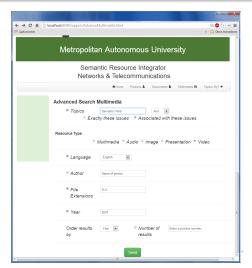
#### Prototipo

- Navegación entre información de los recursos de información.
- Búsqueda Avanzada de los recursos de información.
- Detalles de un recurso de información.

# Navegación entre información de los recursos de información



# Búsqueda Avanzada de los recursos de información



## Detalles de un recurso de información



### Evaluación

#### Evaluar la calidad de los resultados

Esta evaluación consiste en comparar los *recursos relevantes recuperados* por Jena (con/sin inferencia) para una consulta dada, con los resultados que de antemano se sabe responden a esta consulta (total de recursos relevantes).

#### Medir los tiempos promedio de procesamiento de Jena

Esta evaluación consiste en comparar los tiempos de consulta para un modelo con inferencia y otro que no emplea ésta; estos tiempos se toman desde la ejecución de la consulta hasta la presentación de los resultados.

# Preguntas en lenguaje natural

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	$\ccup\cup\cup\cup\cup\cup\cup\cup\cup\cu$	31
Q5	0.00000000000000000000000000000000000	119
Q6	¿Cuáles documentos tratan sobre Ontologías?	30
Q7	$ \stackrel{\cdot}{_{\cdot}}$ Qué recursos fueron publicados en una Revista científica?	156
Q8	$\ensuremath{\dot{\iota}}$ Qué recursos tienen en su contenido las palabras "linked data $\ensuremath{\dot{\iota}}$	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

## Calidad en los Resultados

Id. Consulta	Recursos relevantes recuperados sin inferencia	Recursos relevantes recuperados con inferencia	Total recursos relevantes
Q2.1	1330	1330	1330
Q2.2	0	103	103
Q2.3	18	18	18
Q2.4	15	31	31
Q2.5	66	119	119
Q2.6	15	30	30
Q2.7	156	156	156
Q2.8	159	159	159
Q2.9	0	2	2
Q2.10	3	4	4

# Tiempos de Procesamiento

Id. Consulta	Tiempo promedio (milisegundos)		
Id. Consulta	Modelo sin inferencia	Modelo con inferencia	
Q2.1	24	3520	
Q2.2	9	4016	
Q2.3	12	3520	
Q2.4	16	3472	
Q2.5	42	3451	
Q2.6	14	3392	
Q2.7	13	3431	
Q2.8	32	3312	
Q2.9	34	3570	
Q2.10	11	3398	

### Conclusiones

Un marco de referencia para la integración semántica de los recursos.



Un modelo semántico que representa el conocimiento de los recursos de información.



Un prototipo de interfaz gráfica de usuario para la integración semántica.



Los resultados de nuestra evaluación de recursos recuperados con Jena y de tiempos de consulta.



Nuestro trabajo es una contribución a la integración semántica de los recursos de información existentes en una memoria corporativa.

## Conclusiones

#### Hipótesis

El uso de las tecnologías semánticas es adecuado para lograr la integración semántica de recursos de información en una memoria corporativa.

#### Ventajas de las Tecnologías Semánticas

- Formato estándar.
- Modelos flexibles, extensibles y reutilizables.
- Lenguajes estándar.
- Modelos con conocimiento explícito e implícito.
- Inferencia para materializar tripletas RDF.
- Aplicaciones genéricas.

Contexto y Motivación Descripción del Problema Estado del Arte Integración Semántica de una Memoria Corporativa Prototipo (Aplicación) Evaluación Conclusiones Referencias

## Recomendaciones

## 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.
- [Cao et al., 2011] Cao, T.-D., Phan, T.-H., and Nguyen, A.-D. (2011).
  - An ontology based approach to data representation and information search in smart tourist guide system.
  - In Knowledge and Systems Engineering (KSE), 2011 Third International Conference on, pages 171–175.
- [Chakhmoune et al., 2011] Chakhmoune, R., Behja, H., and Marzak, A. (2011). Building corporate memories in collaborative way using ontologies: Case study of a ssii. In Next Generation Networks and Services (NGNS), 2011 3rd International Conference on,

In Next Generation Networks and Services (NGNS), 2011 3rd International Conference on, pages 23–28.

## Referencias II

```
[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.

[Ha et al., 2011] Ha, I., Oh, K.-J., and Jo, G.-S. (2011).

Ontology-driven visualization system for semantic search.

In Information Science and Applications (ICISA), 2011 International Conference on, pages 1–6.

[Moner et al., 2006] Moner, D., Maldonado, J., Bosca, D., Fernandez, J., Angulo, C., Crespo, P., Vivancos, P., and Robles, M. (2006).

Archetype-based semantic integration and standardization of clinical data.

In Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE, pages 5141–5144.

## Referencias III

[Salam, 2013] Salam, F. (2013).

New semantic indexing and search system based on ontology.

In Emerging Intelligent Data and Web Technologies (EIDWT), 2013 Fourth International Conference on, pages 313–318.

[Suganyakala and Rajalaxmi, 2013] Suganyakala, R. and Rajalaxmi, R. (2013).

Movie related information retrieval using ontology based semantic search.

In Information Communication and Embedded Systems (ICICES), 2013 International

Conference on, pages 421–424.

[Xin and Guangleng, 2001] Xin, W. and Guangleng, X. (2001).

Design rationale as part of corporate technical memory.

In Systems, Man, and Cybernetics, 2001 IEEE International Conference on, volume 3, pages 1904–1908 vol.3.

[Yang and Steele, 2011] Yang, K. and Steele, R. (2011).

A semantic integration solution for online accommodation information integration. In *Industrial Electronics and Applications (ICIEA), 2011 6th IEEE Conference on*, pages 1105–1110.

Contexto y Motivación Descripción del Problema Estado del Arte Integración Semántica de una Memoria Corporativa Prototipo (Aplicación) Evaluación Conclusiones **Referencias** 

## Referencias IV

[Zhai et al., 2008] Zhai, J., Li, J., and Wang, Q. (2008).
Using ontology and xml for semantic integration of electricity information systems.
In Electric Utility Deregulation and Restructuring and Power Technologies, 2008. DRPT 2008. Third International Conference on, pages 2197–2201.