

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

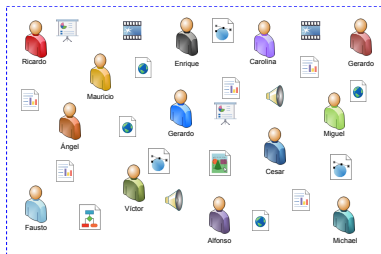
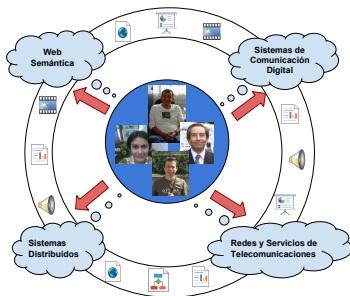
Dr. Héctor Pérez Urbina

Contenido

- 1 Contexto y motivación
- 2 Problema
- 3 Metodología
- 4 Resultados
- 5 Conclusiones
- 6 Referencias

Definición

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



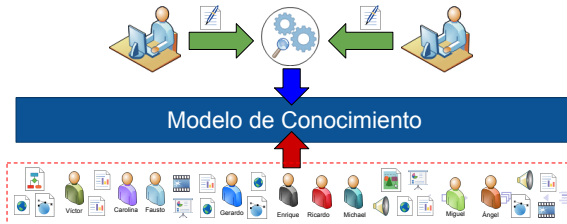
Integración de la información de los recursos de información

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 e información de los *recursos de información*.



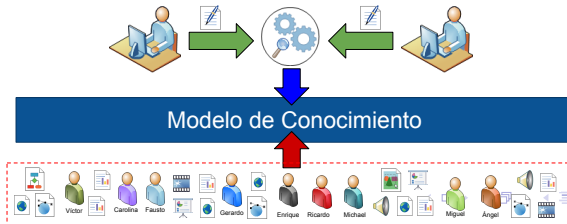
Integración de la información de los recursos de información

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 e información de los *recursos de información*.
- 2 Buscar y recuperar información, mediante la interrogación de la representación de conocimiento (modelo).



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]



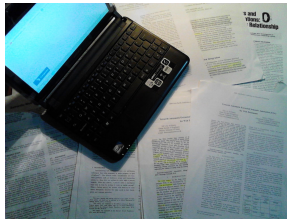
Integración semántica de recursos de información



Estado del Arte

Ejes claves

- 1 Integración de la información a partir del uso de tecnologías semánticas.
- 2 Búsqueda, recuperación y publicación de la información desde una ontología.
- 3 Gestión de una memoria corporativa.



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

Heterogeneidad y significado de la información

Diversidad en formato



pdf, doc, odp, html, txt, xsl, wav, png, mp3, mp4, mpeg, mov, ppt, mov

Diversidad en contenido



p2p, middleware, estado global, replicación, concurrencia, sincronización

Diversidad en estructura



estructurados

semi-estructurados

sin estructura

Homonimia

radio \in Química, Telecomunicaciones, Anatomía, Geometría

Sinonimia

herramienta \equiv aparato \equiv instrumento \equiv mecanismo \equiv artillugio

Pregunta Investigación

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



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

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 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*.

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 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*.
- 3 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.

Objetivos

Objetivo Principal

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

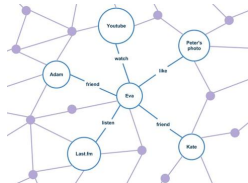
Objetivos Particulares

- 1 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*.
- 3 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.
- 4 La evaluación de la calidad de los **resultados recuperados** y los **tiempos de procesamiento** de la *integración semántica*.

Integración Semántica mediante tecnologías semánticas

Etapas

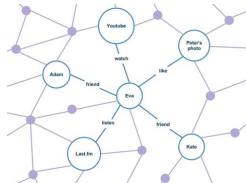
- 1 Representar las características y/o relaciones de los *recursos de información*, para construir un modelo semántico.



Integración Semántica mediante tecnologías semánticas

Etapas

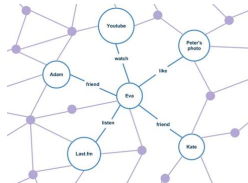
- 1 Representar las características y/o relaciones de los *recursos de información*, para construir un modelo semántico.
- 2 Introducir *reglas de inferencia* en el modelo, para enriquecer con *conocimiento implícito* de los *recursos de información* y del dominio de la memoria.



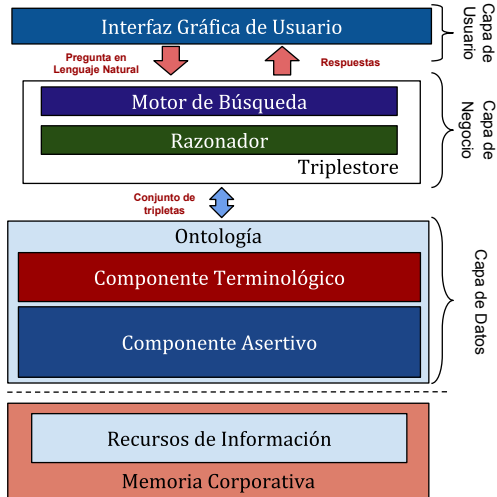
Integración Semántica mediante tecnologías semánticas

Etapas

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

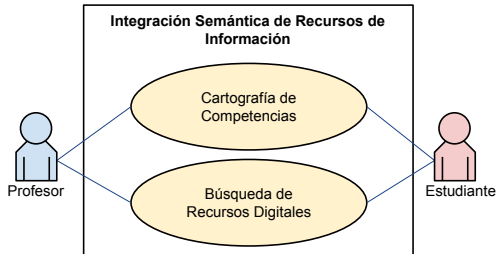


Arquitectura



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.



Representación del conocimiento

Definición

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



sirp:RicardoMarcelinJimenez

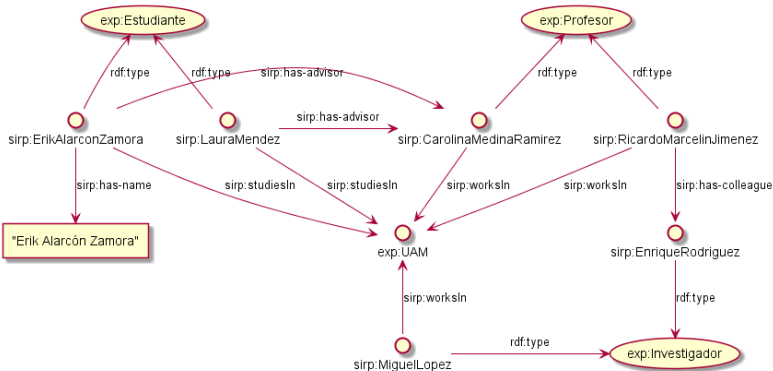
```

a      sirp:Teacher ;
sirp:has-name "Ricardo Marcelin Jiménez"^^xsd:string;
sirp:has-email "calu@xanum.uam.mx"^^xsd:anyURI;
sirp:has-webSite "http://cbi.izt.uam.mx/electrica/profs/ricardo_marcelin.html"^^xsd:anyURI;
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:expertiseln 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,
redes:N_and_ST;
sirp:competentIn 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.

```

Grafo RDF

@prefix sirp: <http://arte.izt.uam.mx/ontologies/personRyT.owl#>
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

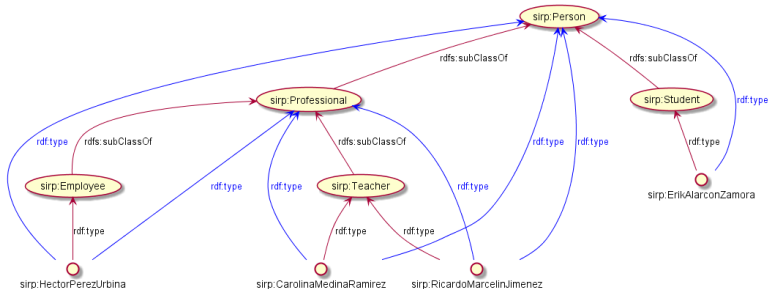


Existen distintas sintaxis de serialización: N3, turtle, RDF/XML, N-triples.

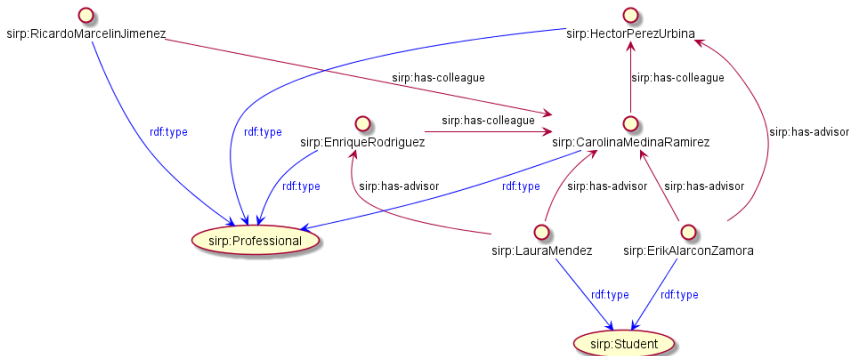
Emplear un programa para inferir conocimiento

Razonador

Un programa que deduce conocimiento a partir de los axiomas y declaraciones explícitas en un modelos semántico.



Emplear un programa para inferir conocimiento



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 Ramírez?



```
PREFIX sirp: <http://arte.izt.uam.mx/ontologies/personRyT.owl#>

SELECT ?name ?ws
WHERE
{
    {?x sirp:has-colleague sirp:CarolinaMedinaRamirez.} UNION
    {sirp:CarolinaMedinaRamirez sirp:has-colleague ?x.} UNION
    {?x sirp:knows sirp:CarolinaMedinaRamirez.} UNION
    {sirp:CarolinaMedinaRamirez sirp:knows ?x.} UNION
    {?x sirp:has-advisor sirp:CarolinaMedinaRamirez.}
    ?x sirp:has-name ?name;
        sirp:has-webSite ?ws.
}
```

Herramientas para la Integración Semántica

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

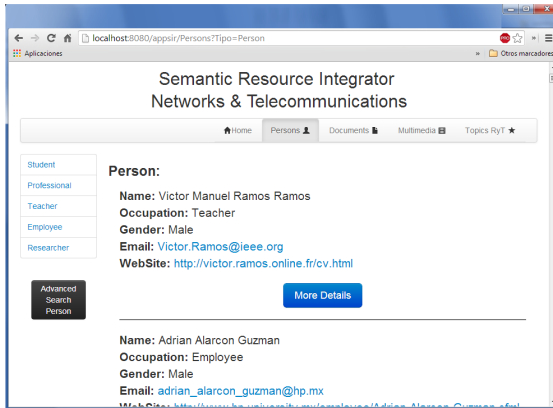
Herramienta que proporciona una serie de interfaces amigables para la construcción y mantenimiento de ontologías. **Protégé**, **pOWL**, **TopBraid Composer** y **SWOOP**.

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**.

Construir un Prototipo (Aplicación)

Una aplicación Web para que los usuarios puedan consultar y visualizar la información de los *recursos de información*, proveniente de un modelo semántico.



Construir un Prototipo (Aplicación)

Metropolitan Autonomous University

Semantic Resource Integrator
Networks & Telecommunications

Home Persons Documents Multimedia Topics RyT

Advanced Search Multimedia

* Topics And
* Exactly these issues * Associated with these issues

Resource Type
☐ Multimedia ☐ Audio ☐ Image ☐ Presentation * ☐ Video

* Language

☐ Author

* File Extensions

* Year

Order results by ☐ Number of results

Evaluar la integración semántica

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	¿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

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)	
	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

Aportaciones

- 1 Un *marco de referencia* para lograr la *integración semántica* de *recursos de información*.

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.

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.
- 3 Un prototipo (interfaz gráfica de usuario) para la interacción amigable (búsqueda y consulta de información) de los usuarios con el modelo semántico.

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.
- 3 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.

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.
- 3 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.
- 5 Un par de scripts para la generación automática y controlada de descripciones (conocimiento explícito) de los *recursos de información*.

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

- Modelos flexibles, extensibles y reutilizables.

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

- Modelos flexibles, extensibles y reutilizables.
- Uso de Lenguajes estándar (World Wide Web Consortium).

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

- Modelos flexibles, extensibles y reutilizables.
- Uso de Lenguajes estándar (World Wide Web Consortium).
- Modelos con conocimiento explícito e implícito.

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

- Modelos flexibles, extensibles y reutilizables.
- Uso de Lenguajes estándar (World Wide Web Consortium).
- Modelos con conocimiento explícito e implícito.
- **Inferencia para materializar tripletas RDF.**

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

- Modelos flexibles, extensibles y reutilizables.
- Uso de Lenguajes estándar (World Wide Web Consortium).
- Modelos con conocimiento explícito e implícito.
- Inferencia para materializar tripletas RDF.
- **Aplicaciones genéricas.**

Recomendaciones

- Introducir nuevos *casos de uso* para modelar mayor conocimiento del área de *Redes y Telecomunicaciones*.

Recomendaciones

- Introducir nuevos *casos de uso* para modelar mayor conocimiento del área de *Redes y Telecomunicaciones*.
- Mejorar la seguridad del prototipo y agregar un recuadro para búsquedas por *palabras clave*.

Recomendaciones

- Introducir nuevos *casos de uso* para modelar mayor conocimiento del área de *Redes y Telecomunicaciones*.
- Mejorar la seguridad del prototipo y agregar un recuadro para búsquedas por *palabras clave*.
- Construir un modulo (aplicación) para generar *tripletas RDF* a partir de las descripciones de los *recursos de información*.

Recomendaciones

- Introducir nuevos *casos de uso* para modelar mayor conocimiento del área de *Redes y Telecomunicaciones*.
- Mejorar la seguridad del prototipo y agregar un recuadro para búsquedas por *palabras clave*.
- Construir un modulo (aplicación) para generar *tripletas RDF* a partir de las descripciones de los *recursos de información*.
 - Generación guiada por los usuarios.

Recomendaciones

- Introducir nuevos *casos de uso* para modelar mayor conocimiento del área de *Redes y Telecomunicaciones*.
- Mejorar la seguridad del prototipo y agregar un recuadro para búsquedas por *palabras clave*.
- Construir un modulo (aplicación) para generar *tripletas RDF* a partir de las descripciones de los *recursos de información*.
 - Generación guiada por los usuarios.
 - **Generación automatizada.**

Recomendaciones

- Introducir nuevos *casos de uso* para modelar mayor conocimiento del área de *Redes y Telecomunicaciones*.
- Mejorar la seguridad del prototipo y agregar un recuadro para búsquedas por *palabras clave*.
- Construir un modulo (aplicación) para generar *tripletas RDF* a partir de las descripciones de los *recursos de información*.
 - Generación guiada por los usuarios.
 - Generación automatizada.
- Comparar los tiempos de procesamiento y recursos relevantes con otros triplestores: *Stardog* y *Sesame*.

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.
- [Gandon, 2002] Gandon, Fabien, L. (2002). Ontology Engineering: a Survey and a Return on Experience. Technical Report RR-4396, INRIA.