

Universidad Politécnica de Valencia

Facultad de Informática

Licenciatura en Documentación

**SISTEMAS DE REPRESENTACIÓN Y PROCESAMIENTO
AUTOMÁTICO DEL CONOCIMIENTO
(SRP)**

**Estudio del buscador semántico
Swoogle**

Alumnos

Consuelo Barberá Montesinos
Mercé Millet Fuster
Emiliano Torres Martínez

Profesores

Carlos Carrascosa Casamayor
Vicente J. Julián Inglada

Curso 2005-2006

Sumario

1. Objetivos.....	3
2. Metodología.....	3
3. Swoogle	3
3.1. Introducción.....	3
3.2. Navegación y búsqueda en Swoogle	6
3.2.1. Panorámica sobre el servicio de búsqueda	6
3.2.2. Cómo buscar Documentos y Ontologías.	7
3.2.3. Cómo buscar términos (referencias URI).....	9
3.2.4. Características de navegación. Ejemplos.....	10
3.3. Otros servicios de Swoogle	12
3.3.1. Semantic Web Trackback Service	12
3.3.2. Suscribirse a una URL	12
3.4. Estadísticas de Swoogle (Swoogle Today).....	12
3.5. Gestión de usuario	13
3.5.1. Login/logout y gestión.....	13
3.6. News	13
4. Resultados.....	13
5. Conclusiones.....	15
6. Bibliografía.....	17
Anexo: Glosario.....	18

1. Objetivos

Con este trabajo se pretende relacionar los conocimientos adquiridos en la asignatura Sistemas de Representación y Procesamiento Automático del Conocimiento, dentro del último curso de la Licenciatura en Documentación, con los conceptos ligados a la Web Semántica, la diversidad de herramientas en desarrollo para impulsarla y sus aplicaciones.

Entre los proyectos para el desarrollo de la Web Semántica, este trabajo se enmarca dentro de la línea de investigación sobre buscadores semánticos. En concreto, el objetivo se centra en el estudio del buscador semántico Swoogle, y su valoración desde el punto de vista del profesional de la información.

2. Metodología

Para ello, dentro de unas limitaciones de tiempo y conocimientos, la metodología seguida ha consistido básicamente en familiarizarnos con los conceptos relacionados con la Web Semántica, con el buscador Swoogle y su evaluación como herramienta, explicando sus funcionalidades y características, en el apartado siguiente.

Para llevar a cabo el estudio de Swoogle se han realizado diversas búsquedas cuyos resultados se han tomado como base de la evaluación. Además, se han tenido en cuenta diversos trabajos y proyectos ya realizados en esta línea.

Por último, se presentan las impresiones derivadas de esta experiencia. Y como resultado de las conclusiones provisionales obtenidas se elaboran posibles líneas de continuidad.

3. Swoogle

3.1. Introducción

La herramienta que aquí se describe es una muestra del creciente interés que tanto organismos públicos como iniciativa privada están teniendo por el desarrollo de la Web Semántica y todo lo que contribuya a mejorar la actual Web en esa dirección. Prueba de ello es el elevado número de proyectos en marcha y la literatura científica que generan. En este contexto se sitúan los proyectos de desarrollo de buscadores semánticos¹.

Swoogle (<http://swoogle.umbc.edu/>) es un proyecto de investigación desarrollado por el grupo de trabajo "eBiquity", del Computer Science and Electrical Engineering Department, en la Universidad de Maryland, Baltimore County. Financiado parcialmente por el Departamento de Defensa (DARPA) y por la National Science Foundation (NSF), de Estados Unidos de América, se puso en marcha en febrero de 2004 y su duración está prevista que transcurra hasta diciembre² de 2006². Se enmarca en la línea de investigación sobre Web Semántica de este grupo, entre otras³ a las que se dedica dentro de la inteligencia Artificial.

Sus creadores partían de la idea de que la actual Web Semántica es, en realidad, una web de documentos de Web Semántica (SWDs), de la que poco se conoce sobre su estructura y crecimiento. Y de que además, los buscadores como Google han transformado la forma en la que las personas

¹ FININ, Tim; DING, Li: Search Engines for Semantic Web Knowledge. *Proceedings of XTech 2006: Building Web 2.0*, Amsterdam, 16-19 May 2006. [en prensa]. Disponible en: http://ebiquity.umbc.edu/_file_directory_/papers/268.pdf [consulta:19-05-2006].

² Home Page de eBiquity: <http://ebiquity.umbc.edu/> y del proyecto: <http://ebiquity.umbc.edu/project/html/id/53/> [consulta: 16-5-2005].

³ <http://ebiquity.umbc.edu/research/area/id/9/Semantic-Web>

acceden y usan la web, convirtiéndose en una tecnología clave para la búsqueda y distribución de información.

Sin embargo, la mayoría de buscadores existentes ofrecen unos servicios pobres para acceder a los SWDs y no explotan plenamente toda la información semántica y estructural que esté codificada en ellos. Como respuesta, Swoogle ha sido desarrollado para indexar SWDs y para servir tanto a los desarrolladores de la Web Semántica como a los agentes software.

Swoogle está diseñado como herramienta de apoyo a las investigaciones sobre Web Semántica, y especialmente en las siguientes tareas:

- Estudio de la magnitud, crecimiento y evolución de la web semántica mediante un sistema de consulta eficiente sobre una extensa base de datos de metadatos de SWDs;
- Recolección, indexación y búsqueda tanto de la definición y utilización de términos de Web Semántica (SWTs) (i.e. Clases y Propiedades) como de sus correspondientes Ontologías (SWOs).
- Posibilitar la compartición de conocimiento distribuido, haciéndolo visible y de fácil acceso.
- Apoyo a herramientas orientadas a la web semántica, como los editores de ontologías, ayudando a encontrar ontologías relevantes, etc.

Podemos dividir la evolución del proyecto en tres etapas: la primera, la inicial en 2004, la segunda, hasta 2005 incluido, y la tercera, desde febrero de 2006, cuando se presentó la actual versión 3.1, hasta la actualidad⁴. En esta evolución cabe destacar el cambio que ha supuesto la migración a la actual versión, pasando de un conjunto de servicios muy extenso a la simplificación del interfaz y la reducción de los servicios.



Figura 1. Interfaz de usuario de Swoogle.

⁴ History of Swoogle. Véase: http://swoogle.umbc.edu/index.php?option=com_swoogle_manual&manual=history
The current version of Swoogle is version 3.1

[Jan, 2006] Swoogle 3.1 gets online and has replaced version 2.2.

[July, 2005] Swoogle 2.2 has been demoed in AAI 2005.

[Dec, 2004] Swoogle 2.1 has been demoed in ISWC 2004.

[Aug, 2004] Swoogle 2.0 gets online and has replaced version 1.0.

[May, 2004] Swoogle 1.0 was released for demonstration in WWW 2004 and DAML PI meeting.

[Mar, 2004] Swoogle born, version 1.0.

Swoogle es un buscador que detecta, que analiza e indexa el conocimiento codificado como documentos de Web Semántica.

Swoogle entiende por documentos de Web Semántica (en los sucesivo SWD) aquéllos que están escritos con algunos de los lenguajes orientados a la construcción de ontologías (RDF, OWL, DAML, N3, etc). Recupera tanto documentos escritos completamente en estos lenguajes (que para Swoogle son documentos de web semántica estrictos), como documentos escritos parcialmente con alguno de ellos.

Asímismo, proporciona un algoritmo inspirado también en el algoritmo *Page Rank* de Google, que para Swoogle se ha denominado *Ontology Rank*. El *Ontology Rank* se ha adaptado a la semántica y a los modelos y patrones de uso que se encuentran en los documentos de Web Semántica⁵.

Actualmente, Swoogle tiene indexados alrededor de 1.5 M de documentos de web semántica. Dicha información está disponible mediante un enlace interno a los datos estadísticos relativos a su estado⁶.

Swoogle 2005 está casi superado por la nueva versión 3.1 de Swoogle 2006, en la que se ha añadido^{7,8}:

- **Nuevo hardware.** Se ejecuta sobre tres máquinas: EB2 es un Sun v20z con dos procesadores y 4G de memoria y ejecuta el rastreador, el DBMS y las interfaces de desarrollo; LOGOS es un servidor IBM y ejecuta las interfaces de producción web, y NATRAJ es el servidor de ficheros para la caché de la SW y el archivo.
- **Más datos.** Tiene unos 950K de documentos en sus índices, comparados con los 340K en 2005.
- **Mejor ranking.** Utiliza los algoritmos de ranking mejorados documentados en la comunicación presentada en la **International Semantic Web Conference 2005**⁹.
- **Mejor rastreador.** Ahora realiza mejor el trabajo de búsqueda rastreando nuevas URLs, incluyendo las que han sido suscritas por personas.
- **Servicio Web.** Dispone de más servicios Web, actualmente con interfaces GCI sencillas que recuperan los resultados como grafos RDF. La utilización de los servicios web requiere la utilización de una clave, con finalidad de control de su utilización (perfiles de uso) y posibles abusos.
- **RDF output.** Todos los resultados de consultas, bien via servicio web o a través de la interface de navegación, están disponibles en RDF.
- **Interface sencilla.** La interfaz gráfica sistema-persona es mejor y más clara.
- **Caché y archivo.** Mantiene una caché de los SWDs y también guarda las copias de versiones más antiguas en un archivo.
- **Un servicio de usuarios registrados.** Los usuarios no registrados solamente ven un número limitado de resultados. Se necesita estar registrado para ver todos los resultados encontrados.

No obstante, la versión anterior posibilitaba la obtención de información más elaborada. Algunos de los aspectos de Swoogle 2005 se han perdido en la versión actual, como el “shopping cart” y el “triple shop”; el “Ontology Dictionary”; la información de gráficos de “swoogle statistics”, y el “Swoogle's top ten”. Se prevé volver a ofrecerlos próximamente.

⁵ DING, Li et al.: “Swoogle: A search and Metadata Engine for the Semantic Web”. En *Proceedings of the Thirteenth ACM Conference on Information and Knowledge Management*, November 2004.

⁶ En: http://swoogle.umbc.edu/index.php?option=com_swoogle_stats [consulta: 19-05-2005]

⁷ Swoogle 2006 released. 1-02-2006. En: <http://ebiquity.umbc.edu/getnews/html/id/35/Swoogle-2006-released> [consulta: 19-05-2005]

⁸ Swoogle 2005: swoogle.umbc.edu/2005 [consulta: 19-05-2005]

⁹ DING, Li et al.: “Finding and ranking knowledge on the Semantic Web”. *Proceedings of the 4th International Semantic Web Conference*, Galway IE, November 2005, Springer-Verlag. Disponible en: http://ebiquity.umbc.edu/_file_directory_/papers/197.pdf [consulta: 18-05-2005].

3.2. Navegación y búsqueda en Swoogle

3.2.1. Panorámica sobre el servicio de búsqueda

El Modelo sobre el que se basa la recuperación de información en Swoogle es el modelo de navegación en Web Semántica.

La arquitectura de Swoogle consta de cuatro componentes^{10,11}: SWD Discovery, Index o metadata creation, Data Analysis, y la Interface. Los cuatro componentes interactúan y trabajan independientemente sobre una base de datos.

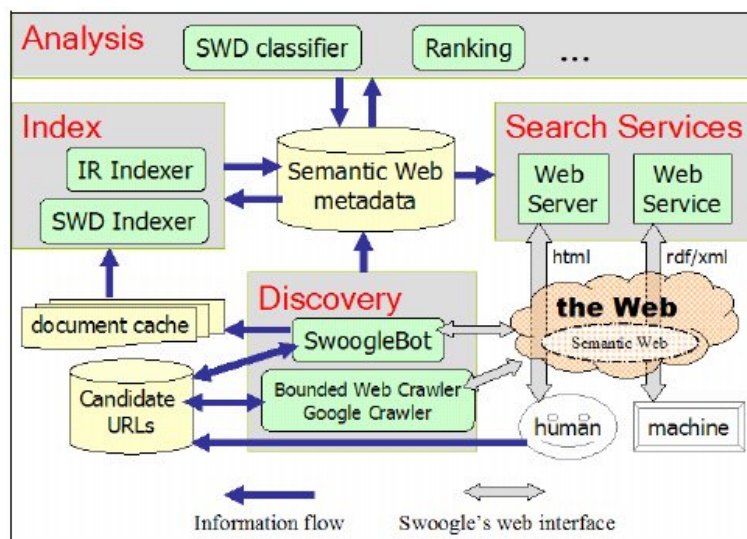


Figura 2. Arquitectura de Swoogle. Fuente: Tim Finin and Li Ding, Search Engines for Semantic Web Knowledge, *Proceedings of XTech 2006: Building Web 2.0*, Amsterdam, 16-19 May 2006.

El componente **Discovery** recoge SWDs potenciales en la Web, y almacena los URLs candidatos. Utiliza cuatro procedimientos: (i) las URLs suscritas, (ii) un rastreador Web que se limita al rastreador de Google y explora sitios candidatos, (iii) un rastreador personalizado que descubre URL utilizando los motores de búsqueda convencionales, por ejemplo, el de Google; y (iv) el rastreador de Web Semántica SwoogleBot, el cual analiza y valida los SWDs recogidos, manteniendo una copia en caché.

El componente **Index** analiza los SWDs descubiertos y genera sus metadatos sobre la Web Semántica; no solamente los de los SWDs y los SWTs, sino también las relaciones entre éstos.

El componente **Analysis** analiza los metadatos generados y los almacena en los mecanismos de ranking modular establecidos: clasificaciones, el Ontoranking, fechas o informes estadísticos.

El componente **Search Services** proporciona un conjunto de servicios de búsqueda a personas y agentes software, permitiéndoles acceder a los metadatos, hacer restricciones en las consultas y navegar en la Web Semántica.

¹⁰ FININ, Tim; DING, Li: Search Engines for Semantic Web Knowledge. *Proceedings of XTech 2006: Building Web 2.0*, Amsterdam, 16-19 May 2006. [en prensa]. Disponible en: http://ebiquity.umbc.edu/_file_directory_/papers/268.pdf [consulta: 19-05-2006].

¹¹ DING, Li, et al.: "Finding and Ranking Knowledge on the Semantic Web". En: *Proceedings of the 4th International Semantic Web Conference*, Galway IE, November 2005, Springer-Verlag. [en prensa]. Disponible en: http://ebiquity.umbc.edu/_file_directory_/papers/197.pdf [consulta: 20-05-2006].

Swoogle utiliza MySQL para gestionar su base de datos.

3.2.2. Cómo buscar Documentos y Ontologías.

Existen tres tipos de consultas disponibles actualmente:

- **Ontology:** este tipo de consulta busca ontologías de la Web Semántica, i.e. SWDs que tienen definida al menos una clase o propiedad.
- **Document:** Recupera no solo ontologías sino otros tipos de documentos, como instancias, siempre y cuando estén introducidas en la SWDB
- **Term:** se refiere a categorías, que dentro de SWDs, representan conceptos, ya sean de forma referencial (p.e. URLs) o bien literales, por ejemplo, términos de un vocabulario o tesauro, o un número de teléfono. Siempre serán instanciados en Swoogle con sus URIs. Anteriormente esta funcionalidad se ofrecía a través de un Diccionario, el Ontology Dictionary.

Componer una cadena de búsqueda (search string)

Por defecto, Swoogle permite al usuario especificar una cadena de búsqueda arbitraria para encontrar documentos relevantes de la Web Semántica.

En vez de construir un índice a partir del texto completo, Swoogle parsea el documento, es decir, analiza si está bien formado y presenta estructuras semánticas de relevancia para la recuperación semántica según modelos establecidos u ontologías, y solamente indexa la parte que más aporta a la semántica del documento, es decir, URL del documento, los términos que están siendo usados por el documento, y/o siendo definidos en él, descripciones explícitas sobre el documento, y los *namespaces* o espacios de nombres usados en el código fuente del documento.

Ejemplo de búsqueda de una cadena de texto:

food	Busca documentos que son relevantes para "food", comida.
------	--

Búsqueda Avanzada utilizando restricciones de campo

(Advanced Search using field constraints)

Swoogle también permite a los agentes/usuarios especificar restricciones sobre los siguientes campos de un documento de Web Semántica, que llamaremos abreviadamente D:

- **Metadatos básicos**

url: el URL de D

desc: la descripción de D, obtenida normalmente de la anotación de la ontología de D (i.e., ternas que contienen el URL de D como sujeto y un literal como objeto)

def: términos que son definidos por D (para cada uno de estos términos, Swoogle indexa su local-name o nombre tal como se encuentra en D y su tipo, es decir si es una clase o propiedad)

ref: términos que son referenciados por D (para cada uno de estos términos Swoogle indexa solo su local-name)

pop: términos que son usados por primera vez en ese sentido concreto por un D

ns: los namespaces utilizados por D

Ejemplos de búsqueda avanzada de metadatos:

url:foaf	Busca documentos que tienen "foaf" como parte de su URL
url:"http://www.w3.org/2000/01/rdf-schema"	Busca un SWD concreto que utilice el URL "http://www.w3.org/2000/01/rdf-schema"
hasGrammar:n3	Busca documentos que son determinados utilizando "N3"
desc:timbl	Busca documentos que tienen "timbl" en su anotación de ontología
def:food	Busca documentos que contienen términos definidos relacionados con "food"
ref:ObjectProperty	Busca documentos que contienen términos que hacen referencia a "ObjectProperty"
ns:foaf	Busca documentos que utilizan namespaces que contienen la subcadena "foaf"
ns:"http://xmlns.com/foaf/0.1/"	Busca documentos que utilizan el namespace "http://xmlns.com/foaf/0.1/"

- **Metadatos del documento propiamente dicho.**

hasEncoding: el charset o conjunto de caracteres fuente de codificación de D, por ejemplo, "utf-8", "iso-8859-1"

hasLength: tamaño del documento en bytes

hasMd5sum: la suma md5 del documento

hasFiletype: el tipo de archivo o formato analizado desde el propio URL del documento, por ejemplo, owl, rdf.

- **Metadatos RDF**

hasGrammar: la gramática RDF de D, su valor podría ser:

- (i) rx - RDF/XML;
- (ii) n3 – notación 3, y
- (iii) nt – Ntriples, ternas

hasCntTriple: el número de triples o ternas

hasOntoRatio: porcentaje de ternas que contribuyen a la definición

hasCntInstance: número de instancias de una clase (class-instances) excluyendo la definición de clase o de propiedad

Búsqueda Avanzada utilizando características Lucene

(Advanced Search using Lucene features)

Este servicio de búsqueda está impulsado por Lucene¹², y por tanto los agentes-usuario pueden también usar directamente sintaxis de consulta Lucene.

Ejemplos de algunas de las características más comunes de búsqueda:

food web	Busca documentos relevantes para "food" y "web". Por defecto, dos palabras clave de búsqueda están conectadas por la semántica AND.
food OR web	Busca documentos relevantes para "food" or "web". OR debe escribirse en mayúscula.
food NOT web	Busca documentos relevantes para food pero no para web. NOT debe escribirse en mayúscula.
modify~	fuzzy match recupera documentos que contiene exactamente esta cadena o variantes similares utilizando lógica borrosa

¹² **Lucene** es un producto desarrollado por Apache. Se trata de un IRS desarrollado íntegramente en JAVA especializado en la gestión de grandes colecciones documentales. Sintaxis de consulta de Lucene: <http://lucene.apache.org/java/docs/queryparsersyntax.html> [Consulta: 14-05-06]

Ejemplo de búsqueda y visualización de resultados.

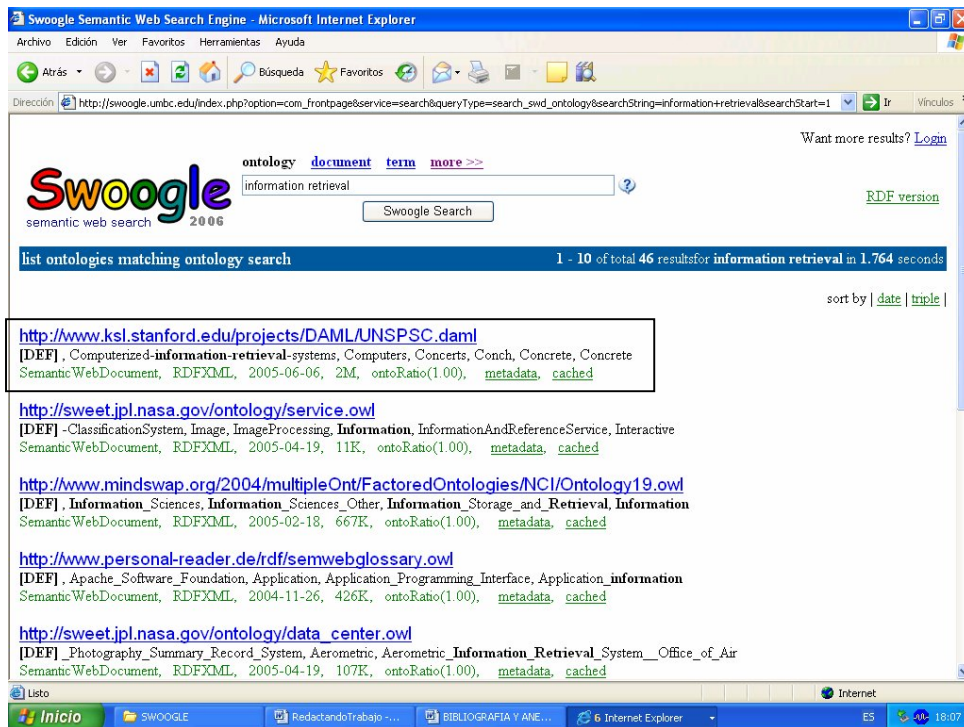


Figura 3. Ejemplo de la vista de la lista de ontologías recuperadas

3.2.3. Cómo buscar términos (referencias URI).

Swoogle permite buscar Términos, i.e. URIs de clases o propiedades definidas en SWDs. En este trabajo y en el contexto de Swoogle la forma abreviada de denotar un SWT será por convención, T.

Componer una cadena de búsqueda (Composing a search string)

Por defecto, Swoogle posibilita a los usuarios especificar una cadena arbitraria de búsqueda para encontrar términos relevantes de la web semántica.

En vez de indexar solo el URI de un término, Swoogle busca información adicional: el URI del término, el namespace y el local-name. Y la descripción recogida desde sus definiciones distribuidas en las muchas ontologías disponibles.

Ejemplo:

foodweb	Busca terminos que son relevantes para "foodweb"
---------	--

Búsqueda avanzada usando restricciones de campo (Advanced Search using field constraints)

Swoogle también permite especificar restricciones sobre los siguientes campos de un término T de la web semántica:

1. **uri**: el URI de T, por ejemplo, "http://xmlns.com/foaf/0.1/Person"
2. **desc**: la descripción de T, obtenida de las definiciones de T en todas las ontologías que lo definen, por ejemplo, "A person"
3. **ns**: el namespace usado por T, por ejemplo, http://xmlns.com/foaf/0.1/
4. **localname**: el local-name utilizado por T, por ejemplo, "Person"

5. **label:** el lexema del local-name de T, por ejemplo, 'SeaFood' tiene dos etiquetas, 'sea' y 'food'
6. **type:** el tipo (o clase o propiedad) de que se trate, por ejemplo, owl.class"

Ejemplo:

uri:food	Busca términos que tienen food como parte de su URL
ns:food	Busca términos que tienen "food" como parte de su namespace
ns:"http://xmlns.com/foaf/0.1/"	Busca términos que tienen como namespace exactamente "http://xmlns.com/foaf/0.1/"
localname:food	Busca términos que tienen "food" como local-name
label:food	Busca términos que tienen "food" como lexema de su local-name
desc:food	Busca términos que tienen "food" como parte de su descripción definicional
type:owl.class	Busca términos que están definidos como una instancia de una clase en OWL, owl:Class

Búsqueda avanzada utilizando características Lucene

(Advanced Search using Lucene features)

Este servicio de búsqueda está impulsado por Lucene, por tanto, los agentes-usuario pueden también utilizar sintaxis de consulta Lucene. Algunas características comúnmente utilizadas son:

food web	Busca términos relevantes para "food" y "web". Por defecto, dos palabras clave de búsqueda están conectados con semántica AND
food OR web	Busca términos relevantes o para "food" o para "web" o para los dos.

3.2.4. Características de navegación. Ejemplos.

El igual que Google, Swoogle presenta una lista de entradas de 10 en 10. Por defecto, aparece ordenada por Ontoratio (normalizado entre 0 y 1) aunque puede ser ordenado por fecha o por triple.

En cada entrada se visualiza la URL del documento con su hiperenlace asociado, una línea donde aparece el texto buscado en contexto, el tipo de documento, por ejemplo, *SemanticWebDocument*, lenguaje del documento (i.e. RDFXML), fecha, tamaño, *ontoRatio*, un enlace a Swoogle *metadata* y otro a cache metadata.

El enlace de la referencia a los metadata ofrece la siguiente información: básica (por defecto), inlinks, outlinks, enlaces de los terminos y namespaces relacionados. Copia la idea de los Science citation index del ISI, Institut for Scientific Information.

Además te ofrece la posibilidad de validar la estructura con varios enlaces a validadores que ayudan a comprobar aspectos como la consistencia de la ontología.

En este caso se trata de una ontología. No tiene almacenado ningún caché metadata, pero muestra la consulta a la base de datos en SPARQL¹³. Cuando proporciona *cached metadata*, Swoogle muestra una tabla con el número, frecuencia, predicado y objeto correspondiente o valor.

¹³ SPARQL es un acrónimo recursivo del inglés *SPARQL Protocol and RDF Query Language*. Se trata de una recomendación para crear un lenguaje de consulta dentro de la **Web semántica** que, aunque en **2005** está en proceso de estandarización por el **DAWG** del **W3C**, está ya implementada en muchos lenguajes y bases de datos. Información disponible en Wikipedia. [Consultado 14-05-06]

Swoogle Semantic Web Search Engine - Microsoft Internet Explorer

Archivo Edición Ver Favoritos Herramientas Ayuda

← Atrás → Búsqueda Favoritos

Dirección http://swoogle.umbc.edu/index.php?option=com_frontpage&service=digest&queryType=digest_swd&searchString=http%3A%2F%2Fderpi.tuwien.ac.at%2F%7Eanc Ir Vinculos >>

Want more results? [Login](#)

Swoogle 2006 [basics](#) [out-links](#) [in-links](#) [related terms](#) [related namespaces](#) [RDF version](#)

semantic web search

show document's metadata

<http://derpi.tuwien.ac.at/~andrei/cerif/DC.MathNetRDFS.rdfs> document

View it using ... [W3C RDF validator](#) [Dumpont](#) [Pellet](#) [Hyperdam!](#)

Swoogle's Metadata

- [hasDateDiscovered](#): 2006-03-25
- [hasDatePing](#): 2006-03-25
- [hasPingState](#): PingModified
- [type](#): SemanticWebDocument
- [isEmbedded](#): false
- [hasGrammar](#): RDFXML
- [hasParseState](#): ParseSuccess
- [hasDateLastmodified](#): 2002-01-21
- [hasDateCache](#): 2006-03-25
- [hasEncoding](#): ISO-8859-1
- [hasLength](#): 7K
- [hasCntTriple](#): 128.00
- [hasOntoRatio](#): 0.84
- [hasCntSwr](#): 82.00
- [hasCntSwrDef](#): 61.00
- [hasCntInstance](#): 3.00

Inicio SWOOGLE SWOOGLEPREPRINT ... Swoogle Semantic W... ES 15:49

Swoogle Semantic Web Search Engine - Microsoft Internet Explorer

Archivo Edición Ver Favoritos Herramientas Ayuda

← Atrás → Búsqueda Favoritos

Dirección http://swoogle.umbc.edu/index.php?option=com_frontpage&service=digest&queryType=digest_swd&searchString=http%3A%2F%2Fderpi.tuwien.ac.at%2F%7Eanc Ir Vinculos >>

- [hasCntSwr](#): 82.00
- [hasCntSwrDef](#): 61.00
- [hasCntInstance](#): 3.00

Cached Metadata

[sparql query] SELECT ?predicate, ?object WHERE (< http://derpi.tuwien.ac.at/~andrei/cerif/DC.MathNetRDFS.rdfs > ?predicate ?object)

NO	frequency	?predicate	?object
1	1	dc:type	ontology
2	1	dc:title	"An Ontology for DC and Math-Net metadata set"
3	1	dc:subject	Conceptualization of DC and Math-Net
4	1	dc:rights	
5	1	dc:language	"OIL"
6	1	dc:identifier	"id"
7	1	dc:format	OIL
8	1	dc:description.release	"1.0"
9	1	dc:description	"The intention of this ontology is a development of Information Retrieval system which can retrieve provide data about research information resources There are several metadata standard to describe such resources: CERIF-RDFS, DC, Math-Net and others But information should be found without dependence on standard which it described in One of the way to do it is to provide ontology of each standard, describe terms of each standard in the terms of other standards"
10	1	dc:date	19 July 2001
11	1	dc:creator	"Andrei S. Lopatenko"

[news](#) o [faq](#) o [feedback](#) o [submit-url](#) o [swoogle2005](#)

Swoogle © 2004-2006, [sbiautv group](#) at UMBC
This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 2.0 License](#).

Inicio SWOOGLE SWOOGLEPREPRINT ... Swoogle Semantic W... ES 15:51

Figura 4. Ejemplo de metadatos para una entrada de los resultados de una consulta.

3.3. Otros servicios de Swoogle

3.3.1. Semantic Web Trackback Service

El servicio *Trackback*¹⁴ de Swoogle comparte una idea similar a la del **Internet Archive** (<http://www.archive.org/>). Swoogle guarda instantáneas de cada uno de los SWDs, como si fuese un archivo histórico. Desde el punto de vista de la documentación, es interesante para estudiar la evolución en el tiempo de un determinado recurso, su trazabilidad, y por otra parte viene a sumarse a las diversas iniciativas nacionales e internacionales que están surgiendo sobre el archivado de documentos web, en las varias versiones de su ciclo de vida, con finalidad de preservación digital.

Este servicio se puede utilizar mediante el siguiente formulario.

URL: <input type="text" value="http://protege.stanford.edu/plugins/owl/protege"/>
<input type="button" value="Take me back"/>

Por ejemplo para esa consulta la información recuperada es la siguiente:

About this URL

- The latest ping on **[2006-04-25]** shows its status is **[Succeed, unchanged]**.
- Its latest cached original snapshot is **[2006-03-21 (3373 bytes)]**
- Its latest cached NTriples snapshot is **[2006-03-21 (41 triples)]**.
- **more document metadata**

We have found **8** cached versions.

1. 2006-03-21: **Original Snapshot (3373 bytes), NTriples Snapshot (41 triples)**
2. 2006-01-29: **Original Snapshot (3373 bytes), NTriples Snapshot (41 triples)**
3. 2005-08-25: **Original Snapshot (3373 bytes), NTriples Snapshot (41 triples)**
4. 2005-07-16: **Original Snapshot (2439 bytes), NTriples Snapshot (35 triples)**
5. 2005-05-20: **Original Snapshot (2173 bytes), NTriples Snapshot (30 triples)**
6. 2005-04-10: **Original Snapshot (1909 bytes), NTriples Snapshot (28 triples)**
7. 2005-02-25: **Original Snapshot (1869 bytes), NTriples Snapshot (27 triples)**
8. 2005-01-24: **Original Snapshot, NTriples Snapshot (31 triples)**

3.3.2. Suscribir una URL

Este servicio¹⁵ permite suscribir una URL correspondiente a un documento de la Web Semántica o a un documento que contenga muchos enlaces a documentos de la Web Semántica. Para suscribir una URL basta rellenar una serie de datos en un formulario.

3.4. Estadísticas de Swoogle (Swoogle Today)

A diferencia la versión de Swoogle 2005, que ofrecía información mediante gráficos, actualmente sólo proporciona una tabla resumida de datos estadísticos *Swoogle hoy* (Swoogle Today)¹⁶.

admin_dt	2006-03-28 23:58:50	Datetime Watched
----------	---------------------	------------------

¹⁴ http://swoogle.umbc.edu/index.php?option=com_swoogle_service&service=trackback&Itemid=12

¹⁵ http://swoogle.umbc.edu/index.php?option=com_swoogle_service&service=submit&Itemid=13

¹⁶ http://swoogle.umbc.edu/index.php?option=com_swoogle_stats

url_total	2,927,960	Number of URLs being discovered
url_pinged	1,896,074	Number of URLs being pinged
total_sw_d	1,235,367	Number of Semantic Web Documents (regardless of embedded or containing some errors) be confirmed.
total_sw_d_strict	862,165	Number of error-free pure Semantic Web Documents
total_sw_d_embed	239,649	Number of documents (except SWDs, PDF, and JPEG) embedding Semantic Web Data
triple_total	246,531,769	Number of triples could be parsed from all Semantic Web Documents.

Las estadísticas ayudan a analizar el crecimiento de la Web Semántica. Como se puede observar en la siguiente tabla mes y medio después, por ejemplo, el número total de SWDs se ha incrementado en un 16,7%:

admin_dt	2006-05-14 23:58:40	Datetime Watched
url_total	3,659,478	Number of URLs being discovered
url_pinged	2,256,588	Number of URLs being pinged
total_sw_d	1,442,233	Number of Semantic Web Documents (regardless of embedded or containing some errors) be confirmed.
total_sw_d_strict	949,431	Number of error-free pure Semantic Web Documents
total_sw_d_embed	332,928	Number of documents (except SWDs, PDF, and JPEG) embedding Semantic Web Data
triple_total	282,451,406	Number of triples could be parsed from all Semantic Web Documents.

3.5. Gestión de usuario

3.5.1. Login/logout y gestión

Ofrece la posibilidad de registrarse y definir tu perfil de interés. Se puede acceder a través del menú de usuario a datos estadísticos de acceso en el apartado Site Traffic (Summary, Users, Resources, Browser, Referers, Domains, Languages, Keywords), sobre IP conectadas, qué consultas han realizado, frecuencias de uso y recursos, etc., a modo de un historial. Notesé la utilidad para el control de uso de la herramienta.

3.6. News

Es un enlace a la página de noticias del grupo de investigación:
<http://ebiquity.umbc.edu/blogger/index.php?cat=24&Itemid=5>

4. Resultados

Se ha diseñado una relación de cuatro grupos de búsquedas, utilizando consultas:

1. Para demostrar la idoneidad de los buscadores semánticos sobre los generales se ha buscado el término **person** en Google y Swoogle, ver tablas 1 y 2.
2. Para establecer qué información recoge cada búsqueda tipo (ontologías, document, term). Visto en epígrafes anteriores.
3. Relativas a los campos semánticos Information Science, Information Retrieval, Librarianship, Records Management y Semantic Web. Ver tablas 3 y 4. Como hecho destacable encontramos que no hay ninguna ontología, documento que utilice el término Librarianship.
4. Búsquedas relativas a fuentes en español. Tabla 5.
5. Para ver qué dominios están más presentes/representativos. Tabla 6.

Google	Resultados
person filetype:rdf	81.400
person filetype:owl	537
person rdfs:Class filetype:rdf	100
person rdfs:Class filetype:owl	6

Tabla 1. Búsqueda de “person” en Google

Swoogle	person	person hasfileType: rdf	person hasfileType: owl	class person hasfileType: rdf	class person hasfileType: owl
Ontology	3.575	631	1.240	56	75
Document	519.521	73.371	3.849	645	1.733
Term	11.731	107*	96*	0*	0*

Tabla 2. Búsqueda de “person” en Swoogle (*no se puede aplicar el campo hasfileType)

Swoogle	information science	information retrieval	Librarianship	Records Management
Ontology	84	65	0	10
Document	127	70	0	17
Term	92	101	2	8

Tabla 3. Búsquedas en campos semánticos sobre documentación.

Swoogle	semantic web	swoogle semantic web	search engine semantic web
Ontology	463	1	3
Document	1.124	3	9
Term	1.488	0	0

Tabla 4. Búsquedas en campos semánticos sobre Web Semántica.

Swoogle	Cervantes	Fiesta	Valencia	Sida
Ontology	0	0	5	0
Document	1	3	13	37
Term	2	7	8	103

Tabla 5. Busquedas de documentos en español.

Se ha trabajado sobre la hipótesis de que las áreas más representadas son las de biomedicina, investigación y comunicación científica, government, commerce, learning, etc.

Swoogle	Medicine	Research	Government	Tourism	Environment	Commerce	Learning
Ontology	72	1.410	355	58	274	188	469
Document	144	5.369	775	140	816	244	2.401
Term	899	8.714	2450	2.294	1.603	211	993

Tabla 6. Busquedas de documentos de diversas áreas semánticas

5. Conclusiones

De los resultados y las impresiones obtenidas del estudio realizado sobre Swoogle, llegamos a las siguientes conclusiones:

- En primer lugar, el trabajo ha tenido principalmente un carácter introductorio, sin llegar a agotar todos los aspectos que requeriría un estudio evaluativo exhaustivo.
- En segundo lugar, se quiere recalcar la utilidad de herramientas de este tipo para la tarea común de desarrollo de la Web Semántica. Si bien es cierto que se ha observado que Swoogle, como proyecto que es, ha transcurrido con más o menos altibajos y la versión actual parece ya haber entrado en una fase de finalización del proyecto. Es comprensible si se tiene en cuenta que se trata de un proyecto financiado con fondos públicos. Hasta de ahora las herramientas habían sido de acceso libre como Protégé, pero recientemente se están apareciendo otras que no lo son¹⁷, como SemanticWorks de la compañía Altova.
- Se ha observado que hay un sesgo lingüístico respecto al uso de términos en otros idiomas diferentes del inglés, como cabría esperar, dado el origen del proyecto. Con toda probabilidad la situación cambiará en función de la propia difusión y del impacto de este tipo de herramientas en la comunidad científica.

Sin embargo, no es una situación paradójica, puesto que la propia naturaleza de las ontologías como estructuras, dado que su idea es servir de espacio común de entendimiento en dominios específicos, comporta que lo lógico sea que, a niveles más globales, se construyan en inglés como lengua franca, sin olvidar que la mayoría de lenguajes de marcado también se nutren de una morfología proveniente del inglés. Por lo tanto, deberemos considerar las peculiaridades de este idioma y la manera de enfocarlas por los distintos proyectos, herramientas y buscadores.

- Esta situación se podría mejorar sin demasiada dificultad, si se deseara incorporar ontologías para un área lingüística e idiomática determinada, ya que, actualmente los rastreadores incorporan de una forma automática los términos de interés a la estructura. No podemos olvidar tampoco que una de las áreas de la IA desde donde más se está trabajando en este sentido es en las técnicas de Procesamiento del Lenguaje Natural (PLN).
- Se ha constatado que existen una serie de dominios mejor representados en Swoogle como son ontologías de bioinformática y medicina, economía, medio ambiente, y dominio tecnológico e industrial, así como Web Semántica, puesto que es en estos campos donde más iniciativas podemos encontrar.
- Desde el punto de vista del profesional de la información actual, Swoogle supone una herramienta que podría considerarse de utilidad para la organización del conocimiento en un dominio determinado, ya que los tradicionales tesauros¹⁸, originariamente pensados para representar el contenido de los documentos, no cubren necesidades que las ontologías sí vienen a resolver en parte. El disponer de una herramienta que ayude a recuperar ontologías y estructuras semánticas facilitaría la gestión no solo de los contenidos sino de toda la organización del conocimiento, es decir, acciones, hechos, transacciones e interoperabilidad entre agentes, tanto para un mismo dominio en diferentes lugares como para sistemas multiagente.
- Aplicaciones más concretas de herramientas como Swoogle serían:

¹⁷ BENJAMINS, J. Richards; CONTRERAS, Jesús: "Semantic Web: out of the Labs, into the Market". Santiago de Compostela, XI Conferencia de la Asociación Española para la Inteligencia Artificial, 16-18 nov. 2005, CAEPIA 2005, Vol. II, pp. 386-390.

¹⁸ GARCÍA JIMÉNEZ, Antonio: "Instrumentos de representación del conocimiento: tesauros versus ontologías". *Anales de Documentación*, núm. 7, 2004, pp. 79-95.

- Contribuir a la mejora de la comunicación científica y tecnológica, al acceso abierto al conocimiento y la cultura, mejora de interoperabilidad, al trabajo cooperativo y colaborativo.
 - Contribuir también a la economía, al desarrollo del e-government, al e-learning, al entretenimiento y al estudio de las redes sociales y comunidades.
 - Evaluación de calidad, métrica (ontométrica) y comparación de las ontologías existentes para el dominio concreto al que pretenden servir¹⁹.
 - Cibermetría y webmetría²⁰.
- La presentación de resultados en Swoogle nos parece mejorable en cuanto a servicio de recuperación con posibilidad de visualización de ontologías. Podría introducirse, además de las ordenaciones actuales, una ordenación de resultados por categorías (clusters) o una clasificación similar a un directorio.
 - En las búsquedas, en concreto al concatenar cadenas de caracteres, la ayuda que da no coincide con los resultados de los experimentos, por ejemplo cuando se utiliza el operador AND (opción por defecto) realiza un OR. Además, en la búsqueda para términos no contempla la posibilidad del NOT booleano ni la búsqueda fuzzy. Esta técnica de búsqueda fuzzy si estaba disponible en la versión de 2005, incluido el diccionario.
 - El manual de Swoogle no está actualizado a la versión de 2006 y quedan puntos por desarrollar.

Sin embargo, y a pesar de todas las dificultades indicadas, la iniciativa y la valoración global del proyecto estudiado se ha estimado como muy positiva para el desarrollo de la Web Semántica y en especial la propuesta como trabajo para la adquisición de conocimiento en el marco de la asignatura, nos ha resultado pedagógicamente satisfactoria, si bien, un estudio más detallado sería adecuado para el contexto de un Proyecto Final de Carrera.

En particular se proponen las siguientes posibles líneas de continuidad en el estudio del proyecto:

- Estudio y seguimiento del crecimiento de la Web Semántica a través de Swoogle.
- Comparación de Swoogle con otros buscadores similares.
- Estudio de la evolución de la presencia y ausencia de los dominios específicos relativos a la Documentación, o de otras áreas de conocimiento y de la difusión del proyecto entre la profesión de la información en España y a nivel internacional. Recepción, aceptación e impacto del proyecto.
- Exportabilidad del modelo de Swoogle y estudio de iniciativas similares en entornos concretos.

¹⁹ LOZANO TELLO, Adolfo: Métrica de idoneidad de ontologías. Universidad de Extremadura. Dpto de Informática. Abril, 2002. Disponible en: <http://quercusseg.unex.es/adolfo/tesis.htm> [consulta: 18-05-2005].

²⁰FABA PÉREZ, Cristina; GUERRERO BOTE, Vicente P.; MOYA ANEGÓN, Félix de: Fundamentos y técnicas ciber métricas. Badajoz, junta de Extremadura, 2004. Disponible en: <http://www.juntaex.es/consejerias/idt/dgsi/Publicaciones/tecnicascibermetricas.pdf> [consulta:21-05-2006].

6. Bibliografía

BERNERS-LEE, Tim; HENDLER, James; LASSILA, Ora. (2001). "The semantic web". *Scientific American*, 284 (5), pp. 34-43.

BENJAMINS, J. Richards; CONTRERAS, Jesús: "Semantic Web: out of the Labs, into the Market". Santiago de Compostela, *XI Conferencia de la Asociación Española para la Inteligencia Artificial, 16-18 nov. 2005*, CAEPIA 2005, Vol. II, pp. 386-390.

BIDDULPH, Matt: "Crawling the Semantic Web". Disponible en: http://www.idealliance.org/papers/dx_xmle04/papers/03-06-03/03-06-03.pdf [consulta: 19-05-2006].

DING, Li et al., "Swoogle: A Search and Metadata Engine for the Semantic Web", *Proceedings of the Thirteenth ACM Conference on Information and Knowledge Management*, November 2004. Disponible en: <http://ebiquity.umbc.edu/paper/html/id/183/Swoogle-A-Search-and-Metadata-Engine-for-the-Semantic-Web> [consulta: 18-05-2005].

DING, Li, et al.: "Finding and Ranking Knowledge on the Semantic Web". En: *Proceedings of the 4th International Semantic Web Conference, Galway IE, November 2005*, Springer-Verlag. [en prensa]. Disponible en: http://ebiquity.umbc.edu/_file_directory_/papers/197.pdf [consulta: 20-05-2006].

FABA PÉREZ, Cristina; GUERRERO BOTE, Vicente P.; MOYA ANEGÓN, Félix de: *Fundamentos y técnicas cibernéticas*. Badajoz, junta de Extremadura, 2004.

FININ, Tim; DING, Li: *Swoogle Manual*. Version 3.1 (last modified: March 3, 2006) [Consulta: abril-mayo 2006].

FININ, Tim; DING, Li: *Search Engines for Semantic Web Knowledge*. *Proceedings of XTech 2006: Building Web 2.0*, Amsterdam, 16-19 May 2006. [en prensa]. Disponible en: http://ebiquity.umbc.edu/_file_directory_/papers/268.pdf [consulta:19-05-2006].

GARCÍA JIMÉNEZ, Antonio: "Instrumentos de representación del conocimiento: tesauros versus ontologías". *Anales de Documentación*, núm. 7, 2004, pp. 79-95.

GARCÍA PEÑALVO, Francisco José: "Web Semántica y Ontologías". En: GARCÍA PEÑALVO, Francisco José; MORENO GARCÍA, M.N. *Tendencias actuales en desarrollo de aplicaciones Web*. Salamanca, Universidad de Salamanca, 2004.

GÓMEZ-PÉREZ, Asunción et al.: "Legal ontologies for the spanish e-government". Santiago de Compostela, *XI Conferencia de la Asociación Española para la Inteligencia Artificial, 16-18 nov. 2005*, CAEPIA 2005, Vol. I, pp. 263-272.

LAMARCA LAPUENTE, María Jesús: *Hipertexto: el nuevo concepto de documento en la cultura de la imagen*. Tesis doctoral. Director: Felix del Valle Gastaminza. Madrid, Universidad Complutense de Madrid, 2006. Disponible en: <http://www.hipertexto.info/index.htm> [consulta: 14-05-2006].

LOZANO TELLO, Adolfo: *Métrica de idoneidad de ontologías*. Tesis doctoral. Directora: Asunción Gómez-Pérez. Universidad de Extremadura. Departamento de Informática. Abril, 2002. Disponible en: <http://quercusseg.unex.es/adolfo/tesis.htm> [consulta: 18-05-2005].

MÉNDEZ RODRÍGUEZ, Eva: *Metadatos y recuperación de información: Estándares, problemas y aplicabilidad en bibliotecas digitales*. Gijón, TREA, 2002.

MÉNDEZ RODRÍGUEZ, Eva: "Cuando la búsqueda se vuelve semántica". *Thinkepi*, 6 de febrero de 2006. En línea: <http://www.thinkepi.net/repositorio/cuando-la-busqueda-se-vuelve-semantica-swoogle/> [consulta: 29-4-2006].

RUSSELL, Stuart; NORVIG, Peter: *Inteligencia Artificial. Un enfoque moderno*. 2ª edición. Madrid, etc., Pearson Educación, 2004.

Anexo: Glosario

A continuación se describen algunos términos con la definición apropiada para Swoogle.

Class

Una clase hace referencia a un **rdfs:Resource** el cual es una instancia de **rdfs:Class**. En Swoogle, dada una terna (X,Y,Z), X es una clase accesible si y solo si se cumplen las siguientes restricciones:

- X no es anónimo
 - Y es **rdf:type**
 - Z es una de las siguientes clases -- **rdfs:Class**, **owl:Class**, **owl:Restriction**, **owl:DataRange**, **daml:Class**, **daml:Datatype**, y **daml:Restriction**.
-

DAML

DARPA's Agent Markup Language

FOAF

FOAF (Friend of A Friend) es un vocabulario RDF para describir personas y relaciones entre personas que fundamenta algunas de las Redes sociales de hoy en día como LiveJournal o MyOpera,

Individual

Un "individual" se refiere a un **rdfs:Resource** el cual es una instancia de **owl:Thing** o **daml:Thing**. En Swoogle, dada una terna (X,Y,Z), X es un individuo accesible si y solo si se cumplen las siguientes restricciones:

- X no es anónimo
 - Y es **rdf:type**
 - Z es cualquiera de: **owl:Thing**, **daml:Thing** or una clase cuyo namespace no es ninguno de estos { owl, rdf, rdfs, daml}.
-

Ontology Ratio

La ontology ratio muestra la proporción de las clases definidas en todos los términos definidos en un SWD.

Dado un SWD foo, su ontology ratio R(foo) se calcula mediante la siguiente ecuación:

$$R(\text{foo}) = (|C(\text{foo})| + |P(\text{foo})|) / (|C(\text{foo})| + |P(\text{foo})| + |I(\text{foo})|)$$

donde C(foo), P(foo), I(foo) se refiere a los conjuntos de clases definidas, propiedades definidas y individuos definidos en foo respectivamente.

Property

Una propiedad referencia a un **rdfs:Resource** el cual es una instancia de **rdf:Property**. En Swoogle, dado un triple (X,Y,Z), X es una propiedad accesible si y solo si se encuentran las siguientes restricciones:

- X no es anonimo
- Y es **rdf:type**
- Z es cualquiera de Meta-properties -- **rdf:Property**, owl:ObjectProperty, owl:DatatypeProperty, owl:AnnotationProperty, owl:FunctionalProperty, owl:InverseFunctionalProperty, owl:SymmetricProperty, owl:OntologyProperty, owl:TransitiveProperty, daml:Property, daml:ObjectProperty, daml:DatatypeProperty, daml:TransitiveProperty, y daml:UniqueProperty.

Semantic Web Document (SWD)

Swoogle considera que un Semantic Web document es un documento representado como un grafo RDF. Actualmente hay tres codificaciones sintácticas aceptadas para grafos RDF, a saber '**RDF/XML**', '**N-Triples**' y '**N3**'.

Aunque no está requerido, los SWDs que emplean estas gramáticas son 'rdf', el 'nt' y el 'n3'. Observe que no consideramos un documento XHTML con RDF enbebido como un SWD. Esta definición se puede relajar en el futuro cuando aparezcan estándares para encajar RDF en XHTML y cuando nuevas codificaciones para RDF lleguen a ser populares.

Semantic Web Ontology (SWO)

Una Semantic Web Ontology es un tipo especial de SWD el cual define muchas **Classes** y **Properties**. Este consiste en información que está almacenada en el T-box de un Description Logic system. En la práctica, un SWD es un SWO con un **Ontology Ratio** no menos de 0.8.

Semantic Web Database (SWDB)

Una Base de Datos Semántica es un tipo especial de SWD el cual no define una **Class** o **Property**. Esto corresponde a formularios que son almacenados en el A-box en un Description Logic system. En la practica, un SWD es un SWDB si su **Ontology Ratio** es 0.

Term

Un termino se refiere a un **rdfs:Resource** nodo en un SWD.

En RDF se definen las sentencias en el formato de una 3-upla o triple (sujeto, predicado, objeto), donde:

Sujeto

Es el Recurso sobre el cual nos estamos refiriendo.

Predicado

Es el Recurso que indica que es lo que estamos definiendo.

Objeto

Puede ser un Recurso o un Literal que podria considerarse el valor de lo que acabamos de definir.

A simple vista pareciera que no es mucho lo que decimos con esto, pero cuando tenemos un conjunto de triples es mayor la expresividad que tenemos. Por ejemplo,

para señalar que cuál es el título de este documento podríamos decir que: el **título** del recurso <http://f14web.com.ar/inkel/que-es-la-web-semantica> es **¿Qué es la Web Semántica?**, que expresado en RDF sería:

```
(<http://f14web.com.ar/inkel/que-es-la-web-semantica>,  
<http://purl.org/dc/elements/1.1/title>,  
"¿Qué es la Web Semántica?")
```

Ahora que hemos visto como se compone un triple de RDF, veamos como sería su sintaxis RDF/XML:

```
<?xml version="1.0" encoding="iso-8859-1"?>  
<rdf:RDF  
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"  
  xmlns:dc="http://purl.org/dc/elements/1.1/">  
  
  <rdf:Description  
    rdf:about="http://f14web.com.ar/inkel/que-es-la-web-semantica">  
    <dc:title>¿Qué es la Web Semántica?</dc:title>  
  </rdf:Description>  
  
</rdf:RDF>
```