



Sistemas de Representación
y Procesamiento Automático
del Conocimiento

<http://swoogle.umbc.edu/>

Swoogle
semantic web search 2006

Consuelo Barberá

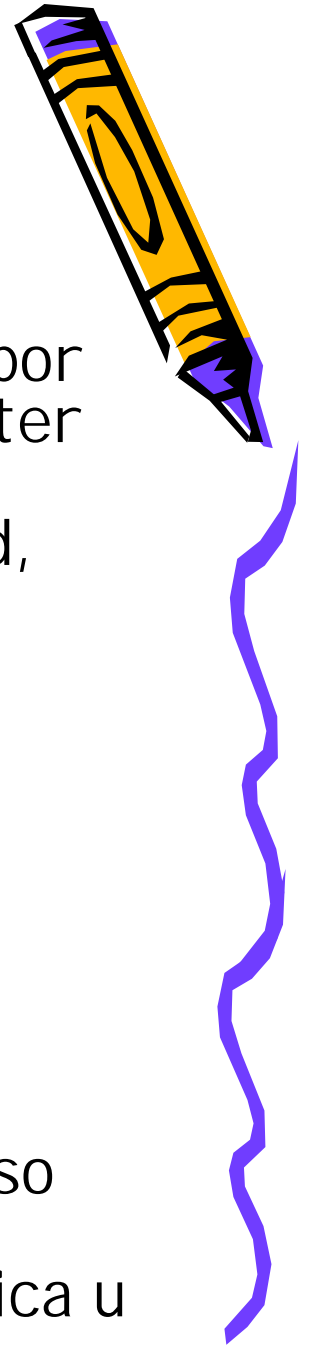
Mercé Millet

Emiliano Torres

Valencia, 22 de mayo de 2006

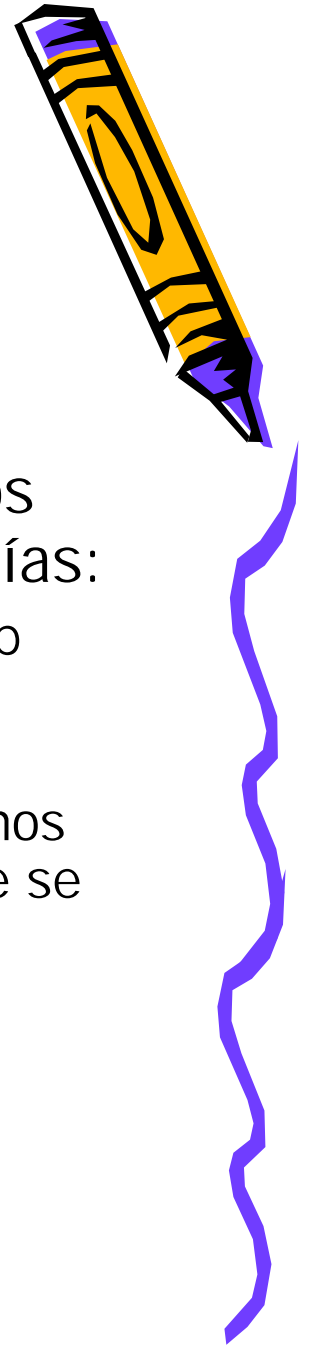
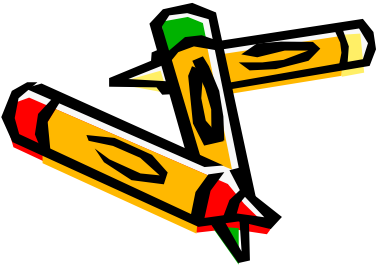
¿Qué es Swoogle?

- Un proyecto de investigación desarrollado por el grupo de trabajo "ebiquity", en el Computer Science and Electrical Engineering Department, de la Universidad de Maryland, Baltimore County, USA.
- Sistema de indexación y recuperación para Documentos de la Web Semántica
- Buscador de documentos para la Web Semántica
- Un Google para la Web Semántica aunque todavía no está dirigido al usuario final
- Un conjunto de servicios
- Herramienta para buscar, clasificar e incluso validar documentos y medir, controlar y analizar los vocabularios de la Web semántica u Ontologías

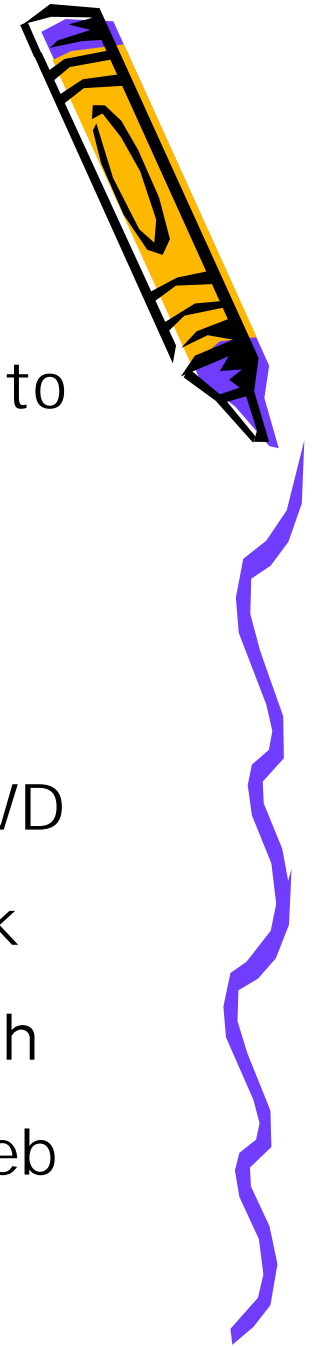


¿Para qué Swoogle?

- Swoogle es una herramienta utilísima para los desarrolladores de la Web semántica y para los agentes de software basado en estas tecnologías:
 - Para estudiar la magnitud y el crecimiento de la Web semántica
 - Para recopilar y buscar clases y propiedades (términos de la Web Semántica, SWTs) o las ontologías en que se conforman
 - Para apoyar herramientas de carácter semántico.

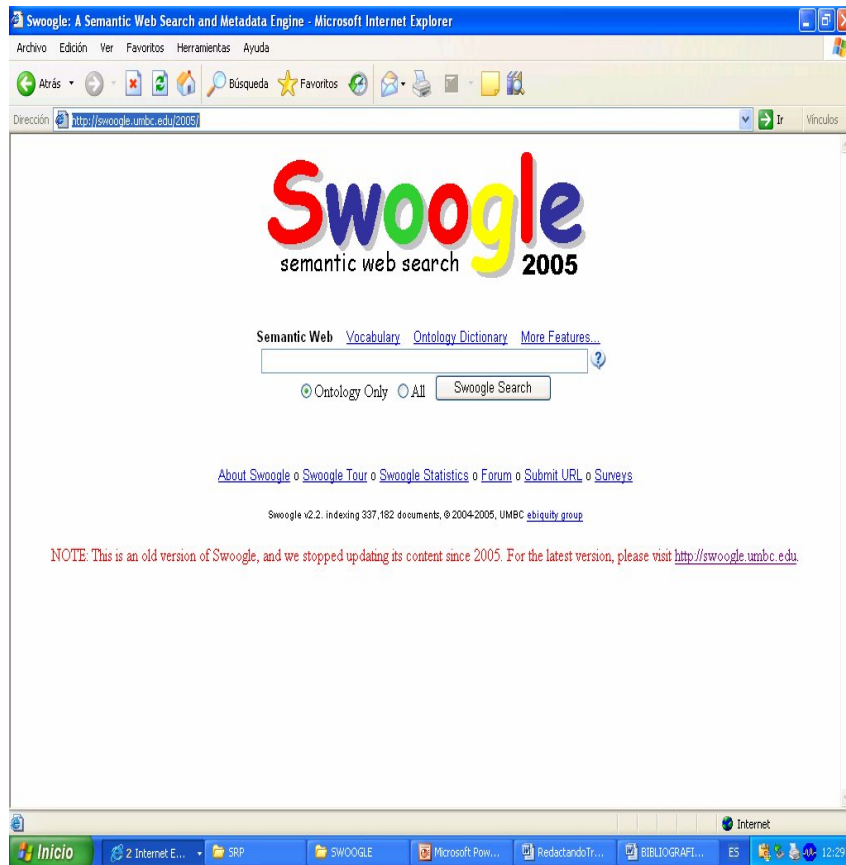
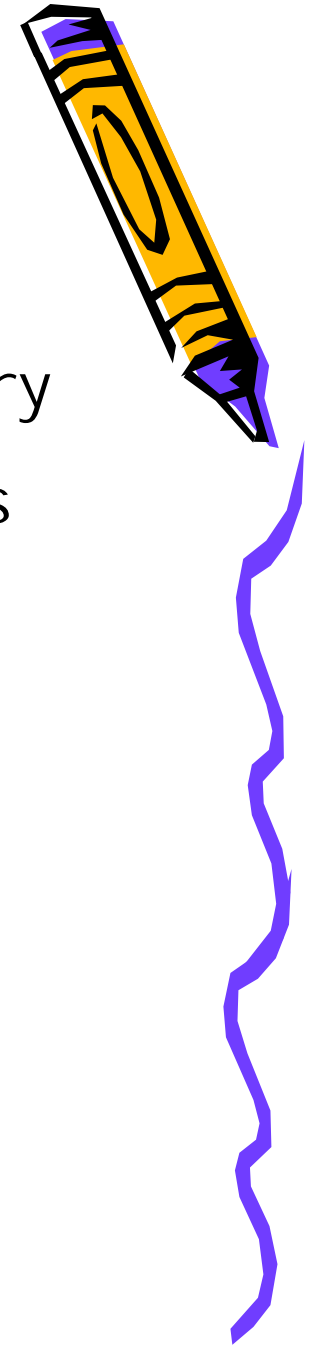


Swoogle 2004

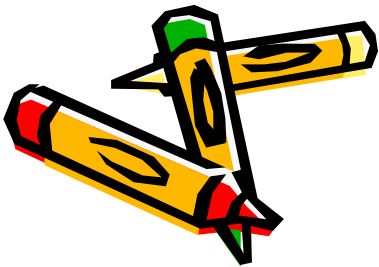


- Descubrimiento de SWD automatizado
- Creación y búsqueda de metadatos SWD
- Ontology rank
- Swoogle watch
- Interface Web

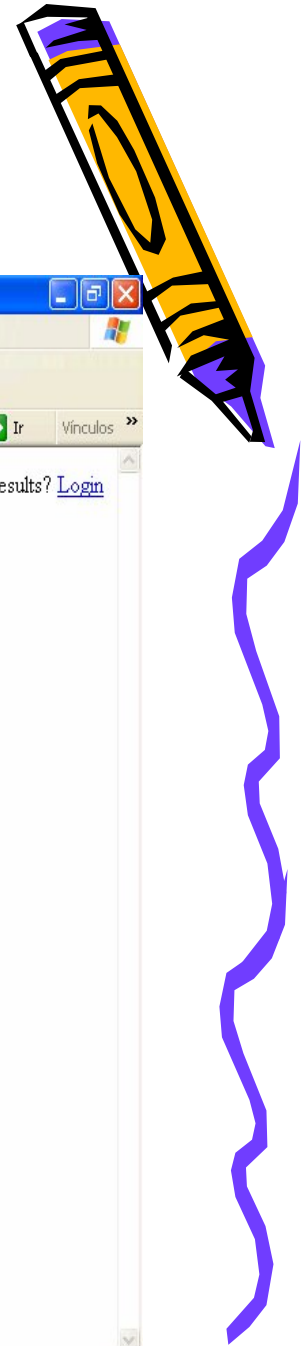
Swoogle 2005



- Ontology dictionary
- Swoogle statistics
- Web service interface (WSDL)
- Bag of URI ref I R search



Swoogle 2006



Swoogle Semantic Web Search Engine - Microsoft Internet Explorer

Archivo Edición Ver Favoritos Herramientas Ayuda

Atrás Búsqueda Favoritos Ir Vinculos

Dirección <http://swoogle.umbc.edu/>

Want more results? [Login](#)

Swoogle

semantic web search 2006

[ontology](#) [document](#) [term](#) [more >>](#)

Swoogle Search

Searching over 10,000 ontologies

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

Swoogle © 2004-2006, [eBiquity Group](#) at UMBC
This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 2.5 License](#).

Inicio 2 Internet E... SRP SWOOGLE Microsoft Pow... RedactandoTr... BIBLIOGRAFI... ES 12:30

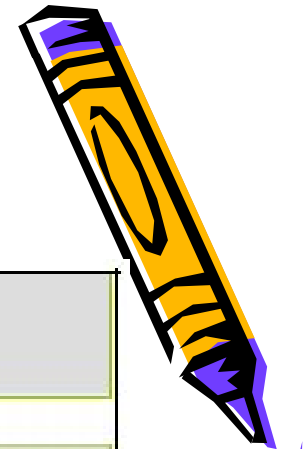
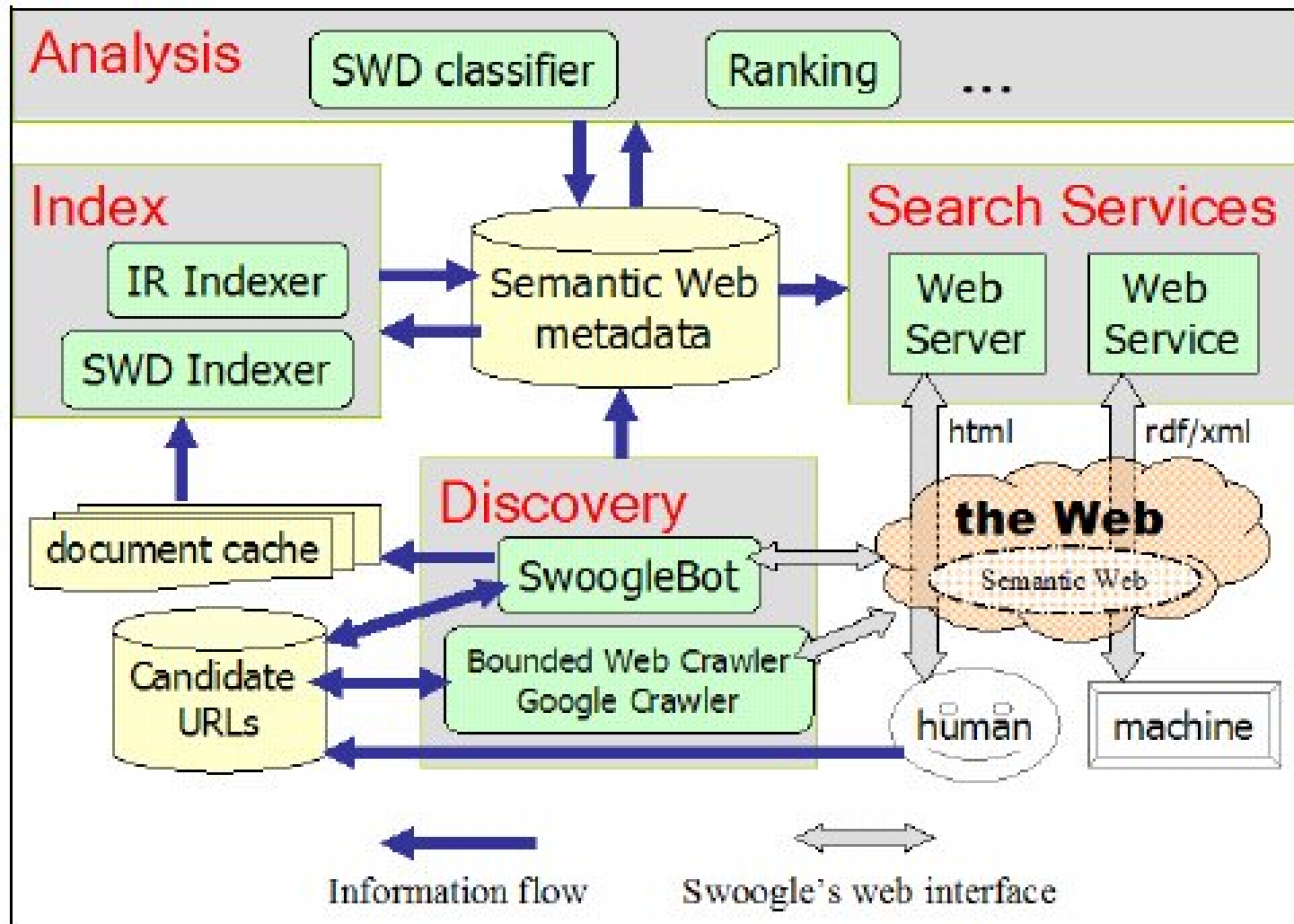


Swoogle 2006

- Nuevo hardware. Se ejecuta sobre tres máquinas.
- Más datos. Tiene unos 950K de documentos en sus índices, comparados con los 340K en 2005.
- Mejor ranking.
- Mejor rastreador.
- Servicio Web. Dispone de más servicios Web.
- 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.
- Caché y archivo.
- Un servicio de usuarios registrados.



Arquitectura de Swoogle



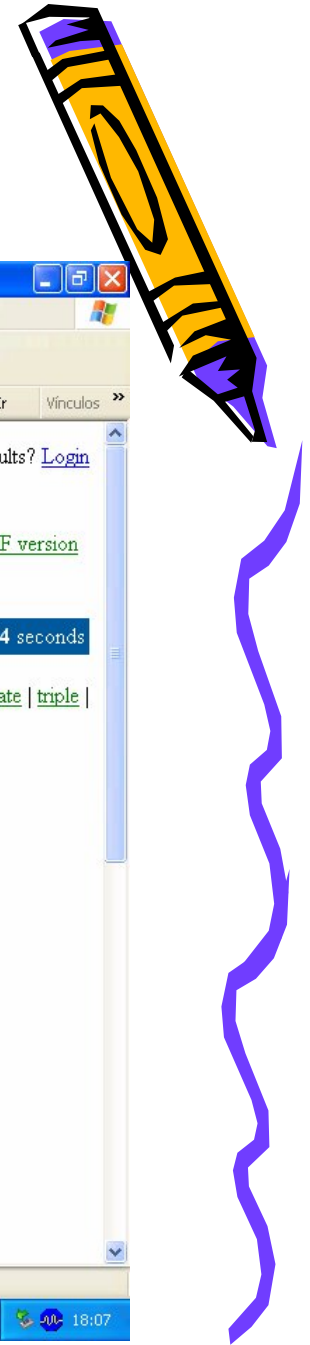
¿Cómo buscar en Swoogle?



- 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 tesoro, o un número de teléfono. Siempre serán instanciados en Swoogle con sus URI s.
- Anteriormente esta funcionalidad se ofrecía a través de un Diccionario, el Ontology Dictionary.



Ejemplo de búsqueda



Swoogle Semantic Web Search Engine - Microsoft Internet Explorer

Archivo Edición Ver Favoritos Herramientas Ayuda

Atrás Búsqueda Favoritos Ir

Dirección http://swoogle.umbc.edu/index.php?option=com_frontpage&service=search&queryType=search_swd_ontology&searchString=information+retrieval&searchStart=1 Ir Vinculos

Want more results? [Login](#)

Swoogle 2006
semantic web search

ontology **document** [term](#) [more >>](#)

information retrieval Swoogle Search [RDF version](#)

list ontologies matching ontology search 1 - 10 of total 46 results for information retrieval in 1.764 seconds

sort by [date](#) | [triple](#) |

<http://www.ksl.stanford.edu/projects/DAML/UNSPSC.daml>
[DEF] , Computerized-information-retrieval-systems, Computers, Concerts, Conch, Concrete, Concrete
SemanticWebDocument, RDFXML, 2005-06-06, 2M, ontoRatio(1.00), [metadata](#), [cached](#)

<http://sweet.jpl.nasa.gov/ontology/service.owl>
[DEF] -ClassificationSystem, Image, ImageProcessing, Information, InformationAndReferenceService, Interactive
SemanticWebDocument, RDFXML, 2005-04-19, 11K, ontoRatio(1.00), [metadata](#), [cached](#)

<http://www.mindswap.org/2004/multipleOnt/FactoredOntologies/NCI/Ontology19.owl>
[DEF] , Information_Sciences, Information_Sciences_Other, Information_Storage_and_Retrieval, Information
SemanticWebDocument, RDFXML, 2005-02-18, 667K, ontoRatio(1.00), [metadata](#), [cached](#)

<http://www.personal-reader.de/rdf/semwebglossary.owl>
[DEF] , Apache_Software_Foundation, Application, Application_Programming_Interface, Application_information
SemanticWebDocument, RDFXML, 2004-11-26, 426K, ontoRatio(1.00), [metadata](#), [cached](#)

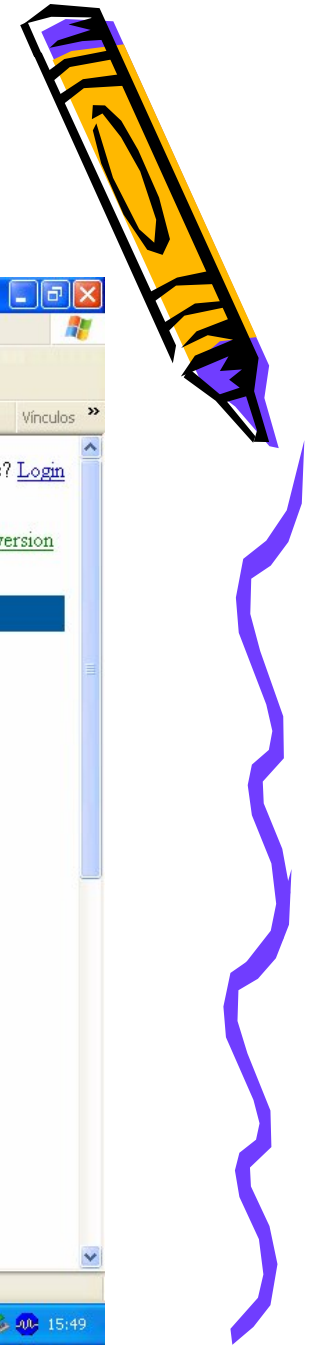
http://sweet.jpl.nasa.gov/ontology/data_center.owl
[DEF] _Photography_Summary_Record_System, Aerometric, Aerometric_Information_Retrieval_System_Office_of_Air
SemanticWebDocument, RDFXML, 2005-04-19, 107K, ontoRatio(1.00), [metadata](#), [cached](#)

Listo Internet

Inicio SWOOGLE RedactandoTrabajo -... BIBLIOGRAFIA Y ANE... 6 Internet Explorer ES 18:07



Ejemplo de búsqueda



Swoogle Semantic Web Search Engine - Microsoft Internet Explorer

Archivo Edición Ver Favoritos Herramientas Ayuda

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

semantic web search 2006

navigational paths for this document

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

show document's metadata

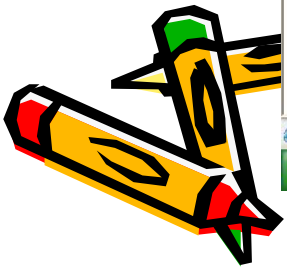
<http://derpi.tuwien.ac.at/~andrei/cerif/DC.MathNetRDFS.rdfs> document

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

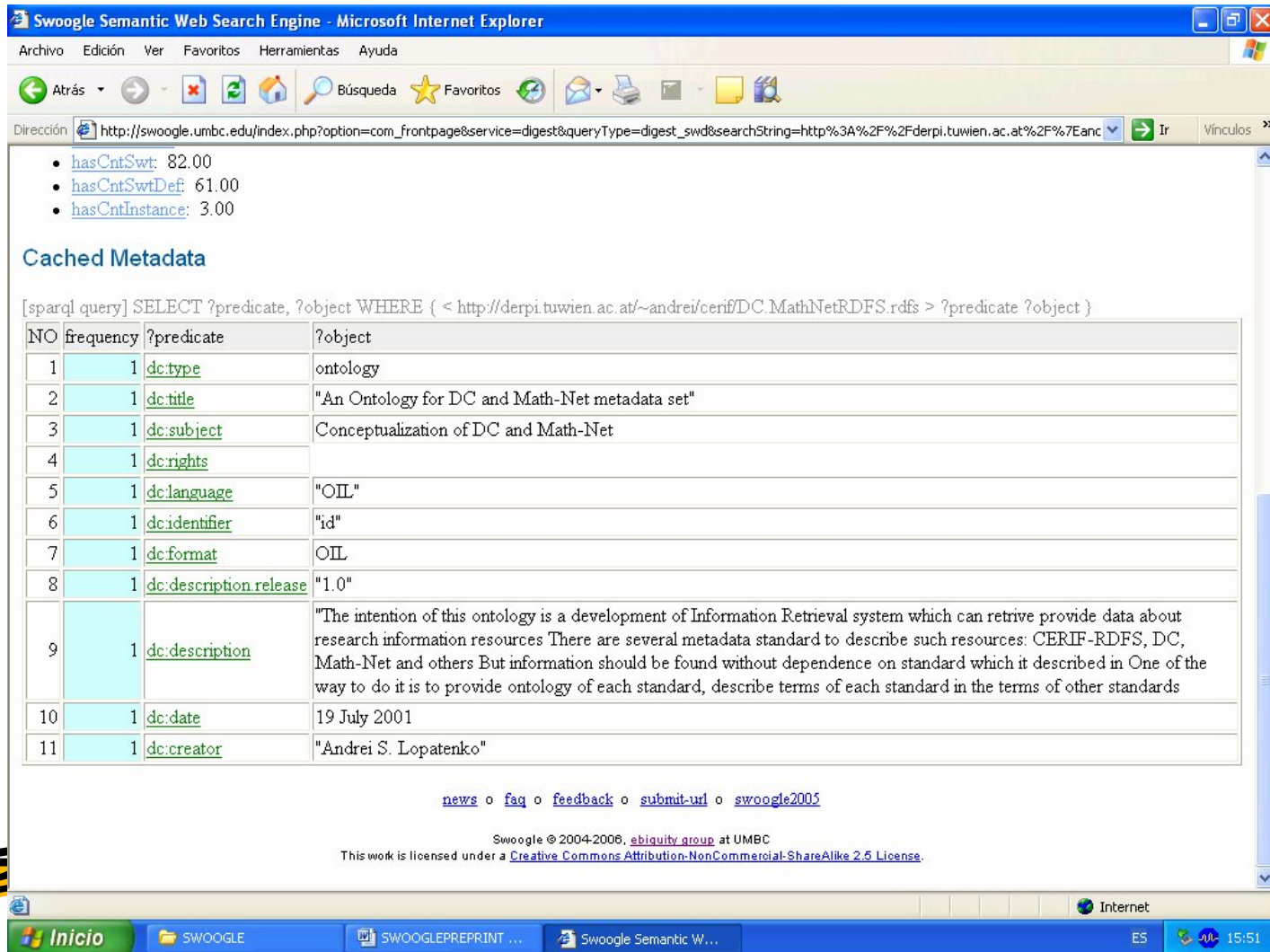
Swoogle's Metadata

- [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
- [hasCntSwf](#): 82.00
- [hasCntSwfDef](#): 61.00
- [hasCntInstance](#): 3.00

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



Ejemplo de búsqueda



• [hasCntSwi](#): 82.00
• [hasCntSwiDef](#): 61.00
• [hasCntInstance](#): 3.00

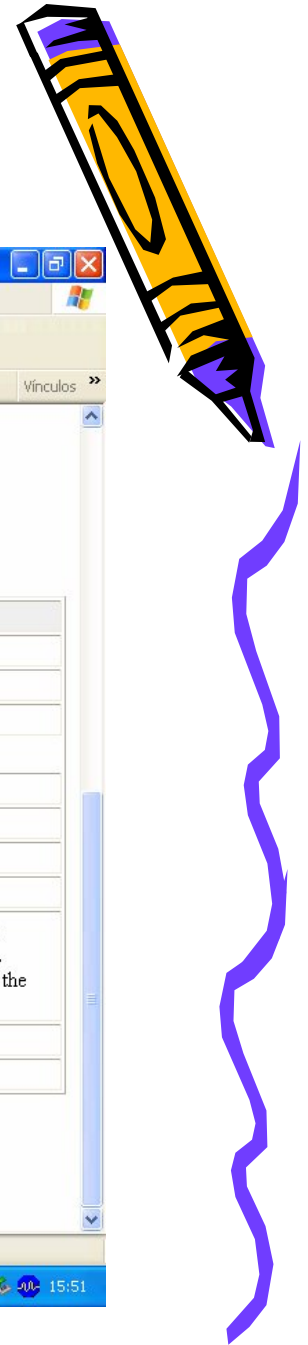
Cached Metadata

[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"

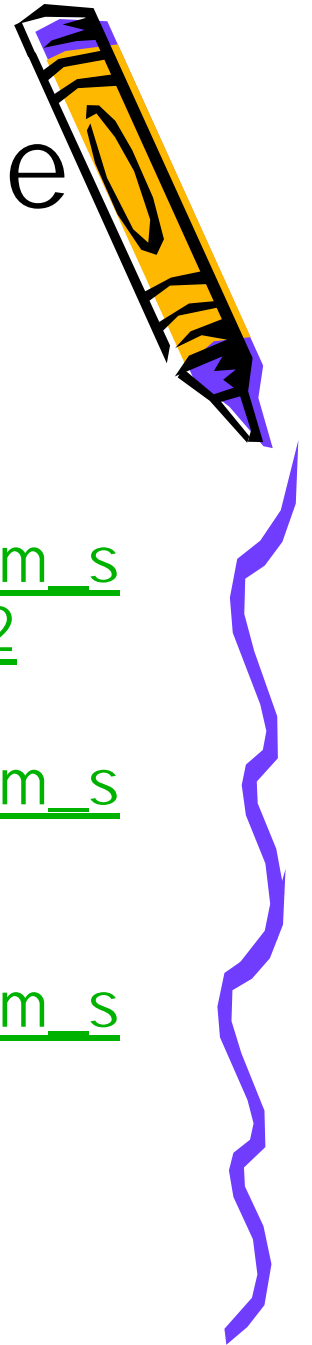
[news](#) o [faq](#) o [feedback](#) o [submit-url](#) o [swoogle2005](#)

Swoogle © 2004-2006, [sbluigitr group](#) at UMBC
This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 2.5 License](#).



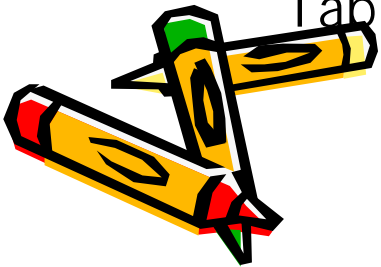
Otros servicios de Swoogle

- Semantic Web Trackback Service (Internet Archive)
http://swoogle.umbc.edu/index.php?option=com_swoogle_service&service=trackback&Itemid=12
- Suscribir una URL
http://swoogle.umbc.edu/index.php?option=com_swoogle_service&service=submit&Itemid=13
- Estadísticas de Swoogle
http://swoogle.umbc.edu/index.php?option=com_swoogle_stats&Itemid=8



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.



Resultados



- Se ha diseñado una relación de cuatro grupos de búsquedas, utilizando consultas:
- Para demostrar la idoneidad de los buscadores semánticos sobre los generales se ha buscado el término *person* en Google y Swoogle.

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

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*



- Para establecer qué información recoge cada búsqueda tipo (ontologías, document, term). Visto en epígrafes anteriores.

Resultados



3. Relativas a los campos semánticos
Information Science, Information Retrieval,
Librarianship, Records Management y

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

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



Resultados

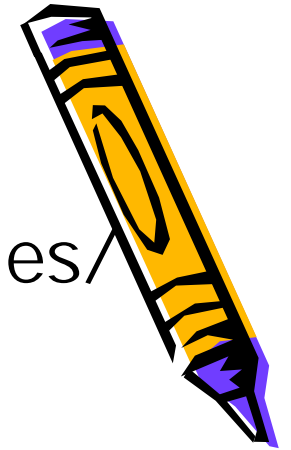
4. Búsquedas relativas a fuentes en español..



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

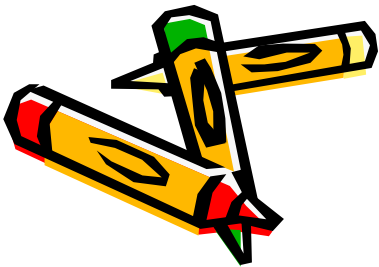


Resultados



5. Para ver qué dominios están más presentes/
representativos

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



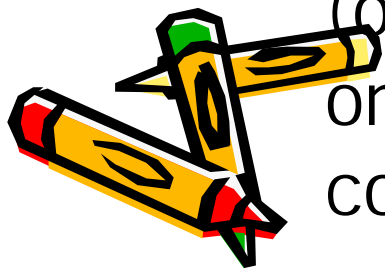
Conclusiones generales

- Destacar la utilidad de herramientas de este tipo para el desarrollo de la Web Semántica.
- Swoogle ha entrado en una fase de finalización del proyecto.
- Se ha observado un sesgo lingüístico.
- Dominios mejor representados.
- Herramienta de utilidad para el profesional de la información en la organización del conocimiento.



Conclusiones: Documentación

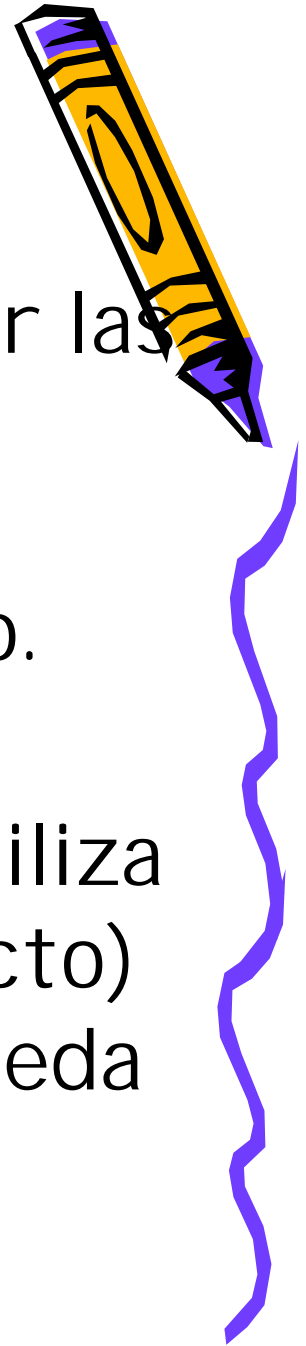
- Aplicaciones más concretas:
 - Contribuir a la mejora de la comunicación científica y tecnológica, y el trabajo cooperativo y colaborativo.
 - Contribuir a la economía, al desarrollo del e-government, al e-learning, al entretenimiento y al estudio de las redes sociales y comunidades.
 - Evaluación de calidad, métrica (ontométrica) y comparación de las ontologías existentes para el dominio concreto al que pretenden servir.



Conclusiones:

Aspectos mejorables

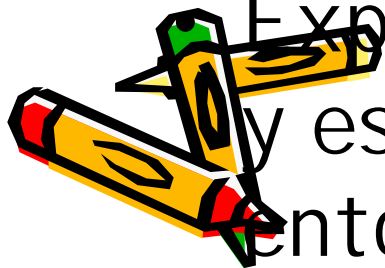
- Presentación de resultados: ampliar las ordenaciones actuales por categorías (clusters) o una clasificación similar a un directorio.
- Algunos operadores no hacen lo esperado por ejemplo cuando se utiliza el operador AND (opción por defecto) realiza un OR. Además, en la búsqueda para términos no contempla la posibilidad del NOT booleano ni la búsqueda fuzzy.



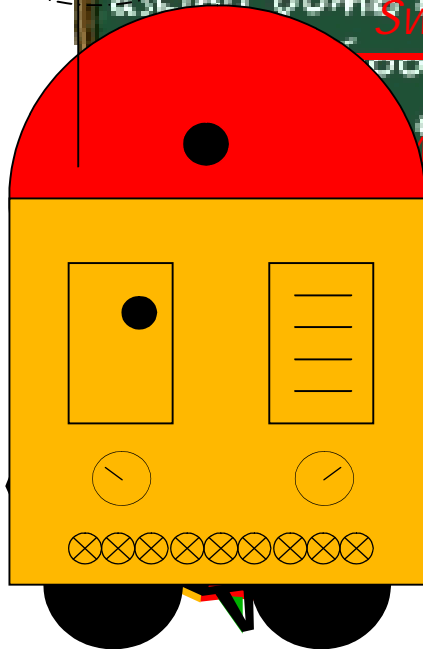
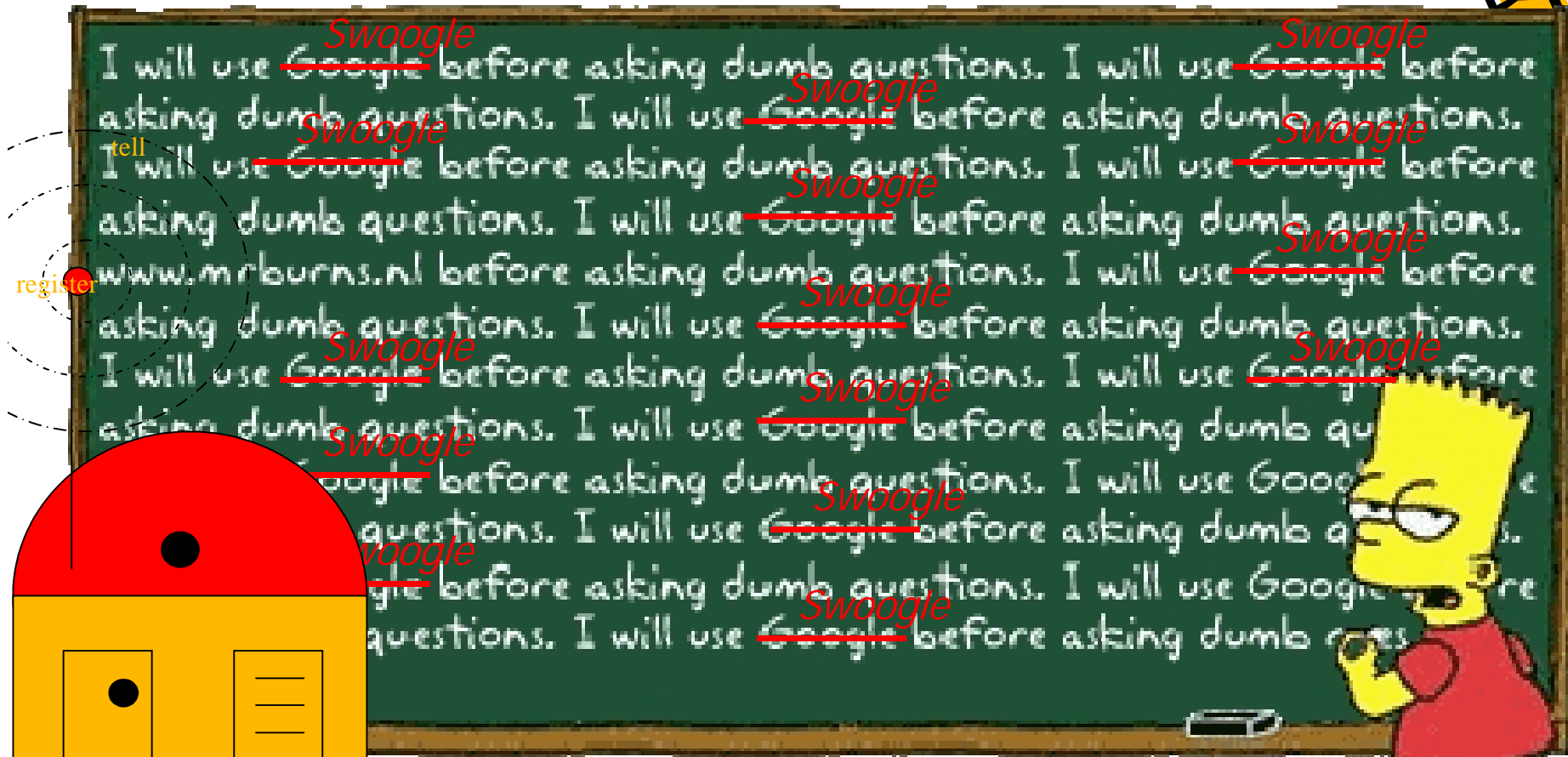
Posibles líneas de continuidad

- 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.

Exportabilidad del modelo de Swoogle y estudio de iniciativas similares en entornos concretos.



¿Basta con Google?



Se necesita un Google para el conocimiento de la Web Semántica que sirva a usuarios y agentes

Gracias por vuestra atención

- Comentarios, dudas...

Consuelo mabarmon@fiv.upv.es

Mercé mermilfu@fiv.upv.es

Emiliano emtormar@fiv.upv.es

