



Universidad
Carlos III de Madrid

The Semantic Web Overview

Luis Sánchez Fernández

Departamento de Ingeniería Telemática
Universidad Carlos III de Madrid

Short history of the Web

- 1990: Creation of World Wide Web infrastructure at CERN by Tim Berners-Lee
 - HTTP, HTML, first Web client, first Web server
- 1993: Mosaic, first graphic Web client
- 1994: Netscape Navigator
- 1996: Commercial use of WWW is generalized
- 1999: Tim Berners-Lee proposes the Semantic Web
- 2002: Weblogs and RSS → Web 2.0
- 6th October 2009: at least 8 billion indexable Web pages
- 23rd September 2010: at least 15 billion indexable Web pages
 - according to <http://www.worldwidewebsize.com/>

The problem of information overload

- The great success of the web has lead to one of its current problems: information overload
 - Difficult and time costly to find and update relevant information for people and companies
 - Ex.: keep an updated state of the art
 - Company employees can use up to 20% of their working time searching in the Web (Outsell Inc, 2002)

Web problems and pitfalls example: search engines

- We make queries and get Web pages that are not related to what we wanted
- We make queries and do not get Web pages that are related to what we wanted

Another example

- Search for images in flickr.com
 - Java (island/bird/coffee/programming language)
 - vela (Spanish) (in English: candle/sail)

Web problems and pitfalls example: search engines

- Some reasons of search engines problems:
 - polisemy/homonym
 - synonymy
 - multilinguism
- In summary: search engines are not based on meanings but in terms (syntactic search)

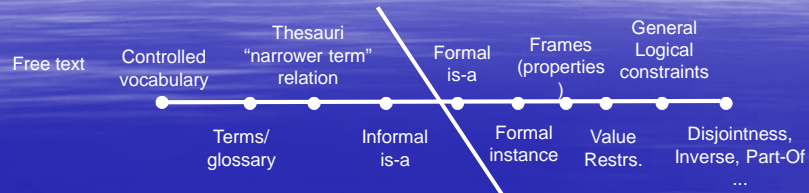
Semantic Search Engine v1.0

- We need a notation to identify meanings
 - example: (term, number) tuple
 - (“George Bush” , 1) → Bush senior
 - (“George Bush” , 2) → Bush junior
 - To attach to (parts of) a document the meaning of terms that are mentioned on it: **semantic annotation**

Semantic Annotation

- ¿How somebody knows which is the code of some concept?
 - By means of vocabularies shared by a community of users
 - example: [Wikipedia](#)
- ¿How somebody knows that the concept represented by some code is referenced in a given Web page?
 - semantic annotation
 - the Web is an open world
 - need for trustness mechanisms

Types of vocabularies



Lassila O, McGuinness D. The Role of Frame-Based Representation on the Semantic Web. Technical Report. Knowledge Systems Laboratory, Stanford University. KSL-01-02. 2001.

Controlled Vocabulary

- Example: a catalog

– <http://www.todocoleccion.net/catalogo.cfm>

Catálogo coleccionismo y antigüedades

1437 secciones | 5.458.358 lotes

A-B C-D E-F G-H I-J K-L M-N O-P Q-R S-T U-V W-X Y-Z

A-B subir

- Abrebotellas y Sacacorchos - Otros coleccionismos 2.280
- Accesorios - Instrumentos Musicales - Música, discos, instrumentos... 273
- Acciones Españolas - Papel 4.810
- Acciones Extranjeras - Papel 1.554
- Acrobática Infantil - Marco - Tebeos y Comics 28
- Action Man - Figuras de Acción - Juguetes, juegos y hobbies 560
- Aeroplanos, Zeppelines y Globos - Temáticas - Postales 2.335
- Agente Secreto - Ferma - Tebeos y Comics 27
- Aguamaniles, Zafas, Palanganas... - Porcelanas y Cerámicas - Antigüedades 334
- Airgam Boys - Figuras de Acción - Juguetes, juegos y hobbies 709
- Albúmina - Fotografía antigua 7.406
- Alfombras y Tapices - Antigüedades 488
- Alicia - Toray - Tebeos y Comics 412
- Alpha Flight - Forum - Tebeos y Comics 631
- Ambróticos, Daguerrotipos y Ferritipos - Fotografía antigua 1.121

Glossary

- <http://www.essentialsofmusic.com/glossary/glossary.html>

ONLINE GLOSSARY

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Click on a letter above to look up terms in the online glossary

absolute music Music that has no literary, dramatic, or pictorial program. Also [pure music](#).

a cappella Choral music performed without instrumental accompaniment.

accelerando Getting faster.

▶▶ **Example: Delibes, *Sylvia Suite*, "Pizzicato"**
Real Audio: [28k](#) | [56k](#) | [About this album](#)
This example remains steady in pulse, then slowly speeds up toward the end.

accent The emphasis on a beat resulting in that beat being louder or longer than another in a measure.

▶▶ **Example: Haydn, *Symphony No.94 "Surprise"*, third movement**
Real Audio: [28k](#) | [56k](#) | [About this album](#)
In this example, the accented beat occurs at regular intervals, on every third pulse: 1 - 2 - 3 - 1 - 2 - 3.

accompanato Accompanied; also a recitative that is accompanied by orchestra.

accordion A musical instrument with a small keyboard and free-vibrating metal reeds that sound when air is generated by pleated bellows. 🎵


Thesaurus

- Example: UNESCO Thesaurus

– <http://www2.ulcc.ac.uk/unesco/thesaurus.htm>

ULCC²

UNESCO Thesaurus: hierarchical list

 United Nations Educational, Scientific and Cultural Organization.

2.25 Chemical sciences

[Back to hierarchical index](#)

Chemical analysis

Narrower Term

- NT1 Biochemical analysis
- NT1 Chromatographic analysis
- NT1 Spectrochemical analysis
- NT1 Trace analysis

Chemical compounds

Narrower Term

- NT1 Acids
- NT1 Inorganic compounds
- NT2 Salt
- NT1 Organic compounds

Informal Is-a

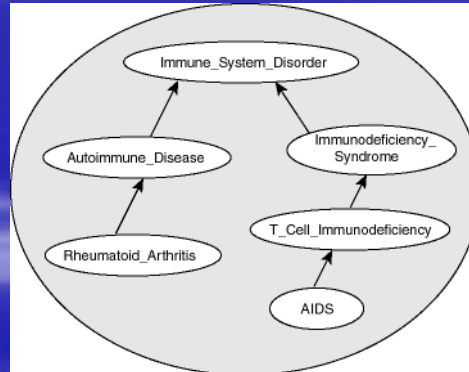
- Is-a: specifies that a concept is narrower than other
 - A Professors is-a Human
 - Luis is-a Professor
- Informal Is-a: although the relation is usually fulfilled there is not a 100% guarrantee
 - A Mammal is-a Not Egg-Lying Animal
 - Platypus

WordNet

- Developed at Princeton University
- Contains nouns, verbs, adjectives, adverbs
- Organized in synsets (synonyms lists + gloss)
- Meanings identified by tuples (term, number)
- Meanings identified by number
- Semantic relationships among synsets
- [WordNet](#)
- Multilingual version: EuroWordNet

Formal Is-a

- A hierarchy of concepts



Logical constraints

- Additional properties
- Axioms that model the relations between concepts and properties in the vocabulary

Ontology definition

- Gruber, Borst, 1993:
 - An ontology is a formal, explicit specification of a shared conceptualization

URIs

- Mechanism used to represent namespaces
- Used to identify **resources**
 - anything you can talk about in a Web document
- URI components:
 - URI scheme: http, ftp, urn, ...
 - name
- Types of URIs: URL, URN

URIref

- URI plus (optionally) *fragment identifier*
- Example
 - <http://www.example.org/index.html#section2>
 - URI: <http://www.example.org/index.html>
 - Fragment identifier: section2

Web problems and pitfalls example: search engines

- Questions and answers:
 - Which are the names of Spanish “Comunidades Autónomas”?
 - Who is the Spanish political party PSOE president?

Semantic Search Engine v2.0

- Semantic annotations extension
- Statements formal representation:
 - “Griñán is PSOE president”
 - isPresident(Griñan, PSOE)
 - “Andalucía is an Spanish Comunidad Autónoma”
 - isCCAA(Andalucía)

Semantic Search Engine v3.0

- Imagine the following query:
 - “We want to find PSOE members that have positions in some Comunidad Autónoma government”
- We search in Web (1.0) and find this information:
 - NOTE: real example by 2005 year

Semantic Search Engine v3.0

- Manuel Chaves is PSOE president
- PSOE is a political party
- Manuel Chaves is Andalucía government president
- Manuel Chaves is Juan Carlos Rodríguez Ibarra party comrade
- Andalucía is a Comunidad Autónoma
- Juan Carlos Rodríguez Ibarra is Extremadura government president
- Extremadura is a Comunidad Autónoma

Semantic Search Engine v3.0

- Domain knowledge:
 - $\forall X, Y, Z, \text{isPartyMember}(X,Z) \Leftarrow \text{isPartyMember}(Y,Z) \wedge \text{isPartyComrade}(X,Y)$
 - $\forall X, Y, \text{isPartyComrade}(Y,X) \Leftarrow \text{isPartyComrade}(X,Y)$
 - $\forall X, Z, \text{isPartyMember}(X,Z) \Leftarrow \text{isPartyPresident}(X,Z)$
 - $\forall X, Y, \text{isGovernmentMemberCCAA}(X,Y) \Leftarrow \text{isConsejeroCCAA}(X,Y)$
 - $\forall X, Y, \text{isGovernmentMemberCCAA}(X,Y) \Leftarrow \text{isPresidentCCAA}(X,Y)$
 - $\forall X, Y, \text{isPresidentCCAA}(X,Y) \Leftarrow \text{isPresident}(X,Y) \wedge \text{isCCAA}(Y)$
 - $\forall X, Y, \text{isPartyPresident}(X,Y) \Leftarrow \text{isPresident}(X,Y) \wedge \text{isParty}(Y)$

Algorithm

Note: this is just a sample algorithm

- Search in the (Semantic) Web for statements related to entities mentioned in the query
 - example: PSOE
- Apply logic reasoning to get new statements
- Repeat Web search over entities (subject, object) present in found statements
- Stop after a number of cycles

Step 1

- Search for statements where PSOE appears
- You get:
 - isPresident(Chaves,PSOE)
 - isParty(PSOE)
- Applying rules
 - $\forall X, Y, \text{isPartyPresident}(X,Y) \leftarrow \text{isPresident}(X,Y) \wedge \text{isParty}(Y)$
 - $\forall X, Z, \text{esPartyMember}(X,Z) \leftarrow \text{isPartyPresident}(X,Z)$
- You get:
 - $\text{isPartyPresident}(\text{Chaves}, \text{PSOE})$
 - $\text{isPartyMember}(\text{Chaves}, \text{PSOE})$

Step 2

- Search for statements where Chaves appears
- You get:
 - isPresident(Chaves,Andalucía)
 - isPartyComrade(Chaves,Ibarra)
- Applying rules
 - $\forall X, Y, Z, \text{isPartyMember}(X,Z) \leftarrow \text{isPartyMember}(Y,Z) \wedge \text{isPartyComrade}(X,Y)$
 - $\forall X, Y, \text{isPartyComrade}(Y,X) \leftarrow \text{isPartyComrade}(X,Y)$
- You get:
 - $\text{isPartyMember}(Ibarra,PSOE)$

Step 3

- Search for statements where Andalucía or Ibarra appear
- You get:
 - isCCAA(Andalucía)
 - isPresident(Ibarra, Extremadura)
- Applying rules
 - $\forall X, Y, \text{isPresidentCCAA}(X,Y) \leftarrow \text{isPresidente}(X,Y) \wedge \text{isCCAA}(Y)$
- You get:
 - $\text{isPresidentCCAA}(Chaves,Andalucía)$

Step 4

- Search for statements where Extremadura appears
- You get:
 - isCCAA(Extremadura)
- Applying rules
 - $\forall X, Y, \text{isPresidentCCAA}(X,Y) \Leftarrow \text{isPresident}(X,Y) \wedge \text{isCCAA}(Y)$
- You get:
 - $\text{isPresidentCCAA}(\text{Ibarra}, \text{Extremadura})$

Query result

- Chaves
- Ibarra

Semantic Search Engine v3.0: Components

- Knowledge base
- Reasoner
- These are the well known basic components of a knowledge-based system

Knowledge Base

- Domain knowledge model
- Information recovered from the Web
- The domain knowledge model can be built reusing available components
 - Ex.: a model for locations (concepts like city, country, etc.) can be used in different applications (book a travel, a semantic search application for news items, etc.)
 - Reusable knowledge components => ontologies

Ontology components

- Classes (ex. Party, Comunidad Autónoma)
- Instances (ex. PSOE, Andalucía)
- Properties (ex. isPartyMember)
- Rules
 - knowledge creation
 - restrictions
 - reactive rules
- Lightweight vs. heavyweight ontologies

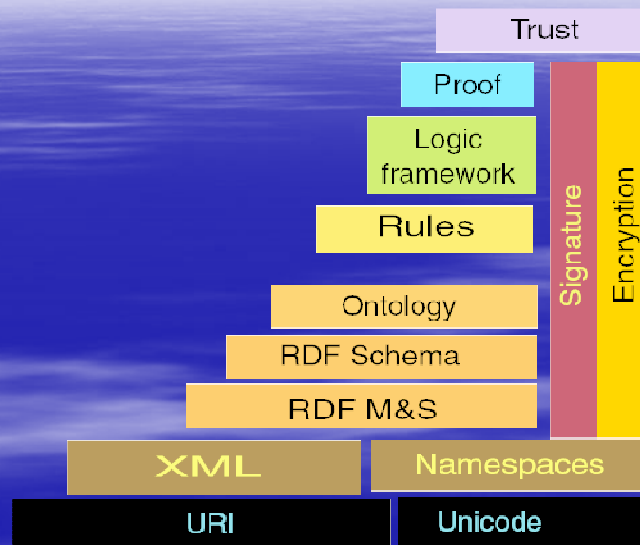
What is the Semantic Web

- Formal description (machine readable) of Web published contents

Technologies

- Semantic annotation
- Ontology engineering
- Reasoners

Proposal (T. Berners-Lee)



Applications

- Semantic search
 - Show user just relevant results
 - Question/answer functionality
 - Multimedia content search
- Information integration
 - Several, distributed sources (ex. databases) integrated by a common domain ontology + mappings
- Services semantic description will enable its discovery
- Web task automatization
 - Ex. comparing prices between several commercial sites
 - Filling of forms based on semantics instead of syntax

Material for next session

- Protégé Ontology editor
 - <http://protege.stanford.edu/>
- Pellet DL Reasoner
 - <http://clarkparsia.com/pellet>
- OWL tutorial
 - <http://www.co-ode.org/resources/tutorials/ProtegeOWLTutorial.pdf>