





Terminology and Dbpedia: Back to basics?

GUADALUPE AGUADO DE CEA

lupe@fi.upm.es

Ontology Engineering Group

http://www.oeg-upm.net/

Universidad Politécnica de Madrid

Outline

Introduction

 Wikipedia as a resource for translators and terminologists

 Dbpedia: exploring a new field for (computerbased) terminology

Open challenges and concluding remarks

Introduction

First stage: Terminology driven by pragmatic aims



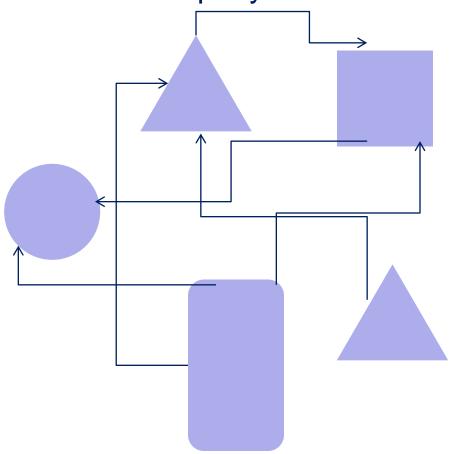
Wüster's ideas

Univocity



Unambiguous terms

Conceptual relations and construction of concept systems



Important issues in the first stage

Economic motivations



- International communication
- ISO TC 37 Committee, 1952







Is standardization important today?

- Other International workgroups
 - W3C
 - CEN, CENELEC, IEEE,

- Language: OWL, RDFS
- Metadata and specifications
- Terminologies: SKOS, FOAF
- Ontologies: CYC, SIMPLE,
- •



Sociocommunicative approaches

Second stage

- •L'aménagement linguistique, 1976
 - Sager, 1990, Assal/Gaudin 1991,
 Cabré 1993, Gambier 1994, Gaudin 2007
- Dictionary Building (Delisle 2008)
- Language planning in Spain:
 - Systematic coining terms
 - TermCat '84, Cabré 1993, Colomer, 2011
 - •UZEI, '86: Euskalterm
 - Termigal: '96





Shift in Translation and Terminology

- Sociopolitical reasons:
 - Increase in the number of countries in EU
 - Increase in the number of Translation Schools
 - Terminology was included in syllabuses
- Linguistic reasons
 - Dominant theories:
 - Cognitivism, Lakoff ('80), Langacker ('87) ...
 - Systemic-functional linguistics (Halliday, '85)
 - In translation: Hatim and Mason, ('90) Nord ('97) ...
 - In terminology: Temmermann (2000), Montero & Faber. (2008) ...

Computer-based Terminology

- Third stage
 - Corpus-based studies
 - Great development of technological tools applied to different tasks:
 - Terminology extraction
 - Ontology building
 - Definition enrichment
 - Pattern-based relations
 -
 - Collaboration between terminologists and technicians:
 - Bourigault/Jacquemin/L'Homme, 2001
 - Condamines et al, 2005
 - Temmerman & Kerremans 2003

Wikipedia

WIKIPEDIA

English

The Free Encyclopedia

Français

L'encyclopédie libre 303 000+ articles

日本語

フリー百科事典 221 000+ 記事

Svenska

Den fria encyklopedin 166 000+ artiklar

Português

A enciclopédia livre 146 000+ artigos

Deutsch

Die freie Enzyklopädie 413 000+ Artikel

Polski

Wolna Encyklopedia 241 000+ haseł

Nederlands

De vrije encyclopedie 205 000+ artikelen

Italiano

L'enciclopedia libera 165 000+ voci

Español

La enciclopedia libre 125 000+ artículos

Multilingual, web-based, freecontent encyclopedia projectOpen editable model

祖

Etymology

-pedia -> child, paideia

Pediatría (es)

Pediatrics (en)

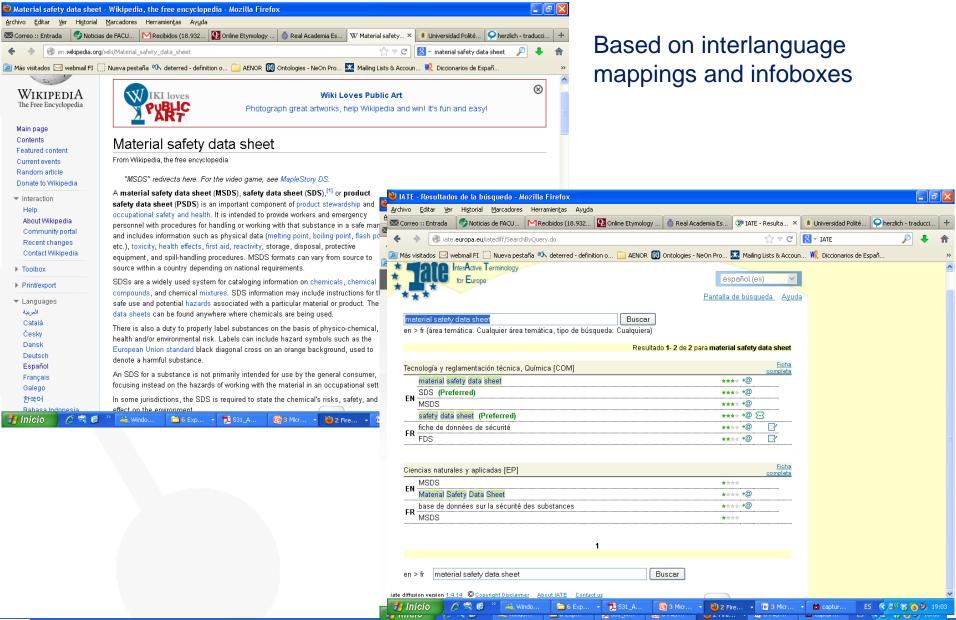
Pediatrique (fr)

Pädiatrie (de)

Evolution

- •Online ED: reference work arranged alphabetically 1640
- Webster's: course of general education

Interlanguage links



Spliceosome in Wikipedia

Iviodiller | Allicher Hilstorique





Crear una cuenta



The Free Encyclopedia

Main page

Featured content
Current events
Random article
Donate to Wikipedia

- Interaction
 Help
 About Wikipedia
 Community portal
 Recent changes
 Contact Wikipedia
- ▶ Toolbox
- Print/export
- ▼ Languages
 Català
 Česky
 Deutsch
 Eλληνικά
 Español



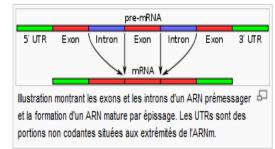






Le **splicéosome**, appelé particule d'épissage (en anglais, *splicing*), est un complexe dynamique de particules ribonucléoprotéiques (composées d'ARN et de protéines) et localisé dans le noyau des cellules. Son rôle est d'assurer l'excision des introns, des régions non-codantes de l'ARN pré-messagers et la suture des exons, qui correspondent aux parties codantes. C'est une étape essentielle du processus de maturation des ARN messagers, un mécanisme conservé chez tous les organismes eucaryotes.

Le splicéosome est en général composé de cinq particules ribonucléoprotéiques, appelées snRNP. Pour la majorité des introns nucléaires, les snRNP composant le splicéosome sont appelées U1, U2, U4, U5 et U6. Elles s'assemblent sur l'intron suivant un mécanisme précis et catalysent les différentes étapes de la réaction d'épissage.



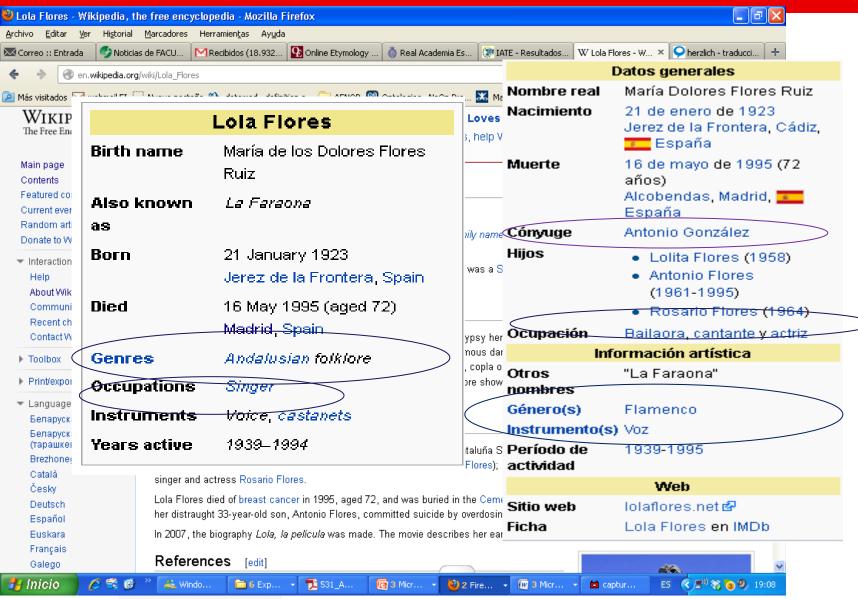
Sommaire [masquer]

- 1 Rôle
- 2 Composition
- 3 Mécanisme
- 4 Voir aussi

Rôle [modifier]

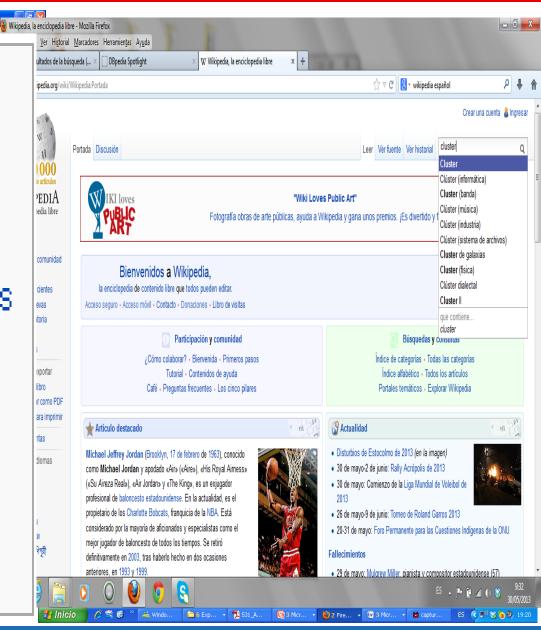
Très souvent chez les eucaryotes, on trouve au milieu des chaines codantes d'ARN (les exons), des portions non-codantes, les introns, qui ne contiennent aucune information permettant la traduction de la protéine. Ces introns doivent être enlevés de l'ARN pré-messager et les exons doivent ensuite être suturés, un processus qu'on appelle l'épissage. L'ARN pré-messager contient des signaux spécifiques dans sa séquence qui permettent de réaliser ce processus : à l'intérieur de l'intron doivent figurer un site d'épissage en 3', un site d'épissage en 5' et un point de branchement. Le site d'épissage 5' ou site donneur comprend très souvent une séquence GU à l'extrémité 5' de l'intron. Le site d'épissage en 3' ou site accepteur d'épissage se termine presque toujours avec une séquence AG. En amont de l'extémité AG se trouve une région riche en pyrimidines (C et U), le tractus polypyrimidine. En amont du tractus polypyrimidine est situé le point de branchement, qui comprend une adénine.

Infoboxes



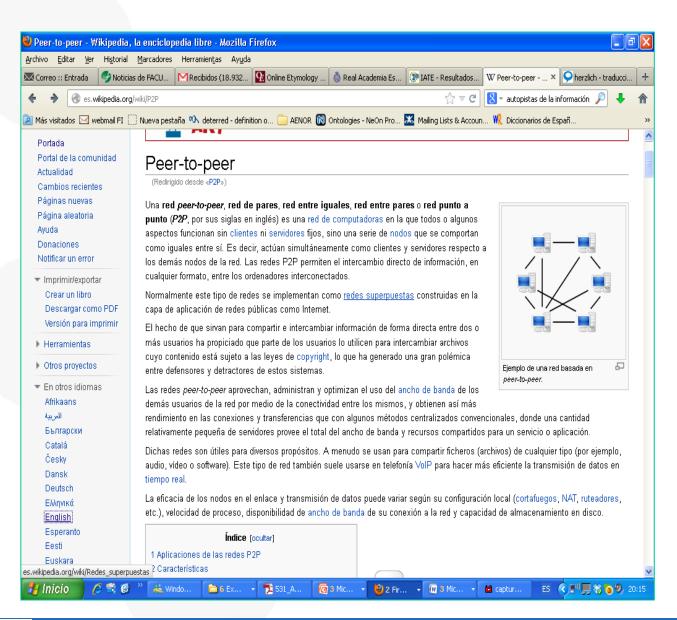
Cluster: Disambiguation





Engineer ing@roup

Redirect pages

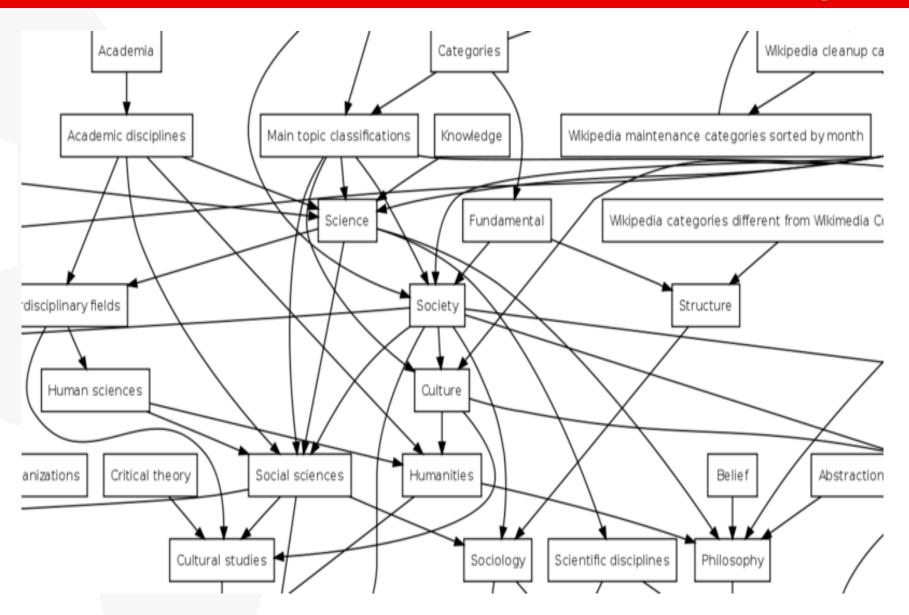


Red de pares: Y
Red entre iguales: Y
Red entre pares: Y
Red punto a punto: N
P2P: Y

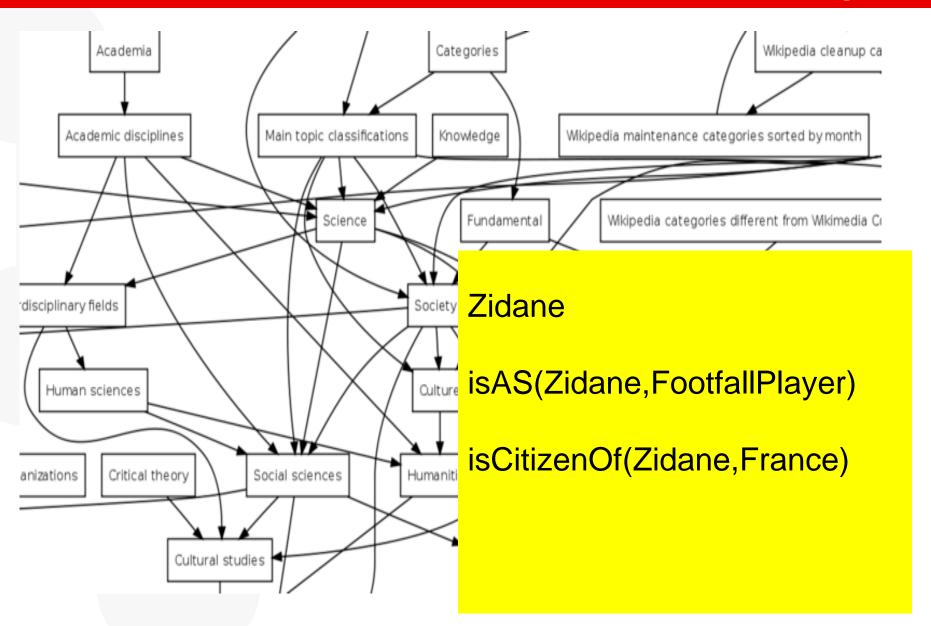
Pair á pair : Y

- General subject outlines (outlines of knowledge).
 - Academic disciplines
 - Figurative system of human knowledge
 - Propædia Outline of Knowledge
 - Subject classification systems
 - Academic classification systems
 - Fields of doctoral studies (United States).
 - Joint Academic Classification of Subjects
 - Library classification systems
 - Colon classification
 - Cutter Expansive Classification
 - Bliss bibliographic classification
 - Dewey Decimal classes
 - Library of Congress Classification
 - Universal Decimal Classification

Categories II



Categories

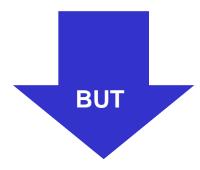




It is an **ontology** that contains almost 4 (3.77) million "things" including

- Persons, places
- creative works (music albums, films and video games)
- organizations such as companies and educational institutions)
- Species and more than 5,000 diseases,

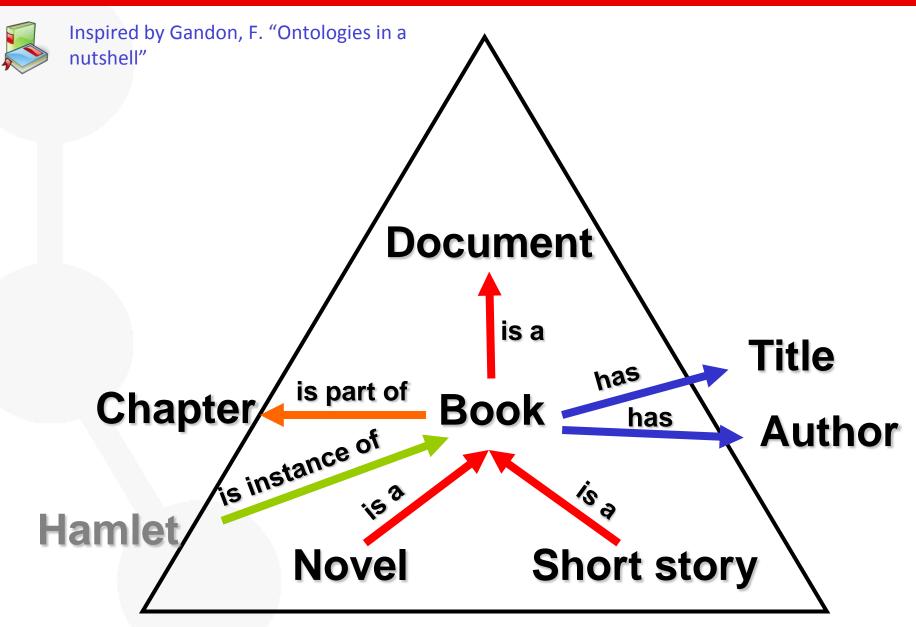
- Provide consensual knowledge
- Reused and shared across software applications and by groups of people



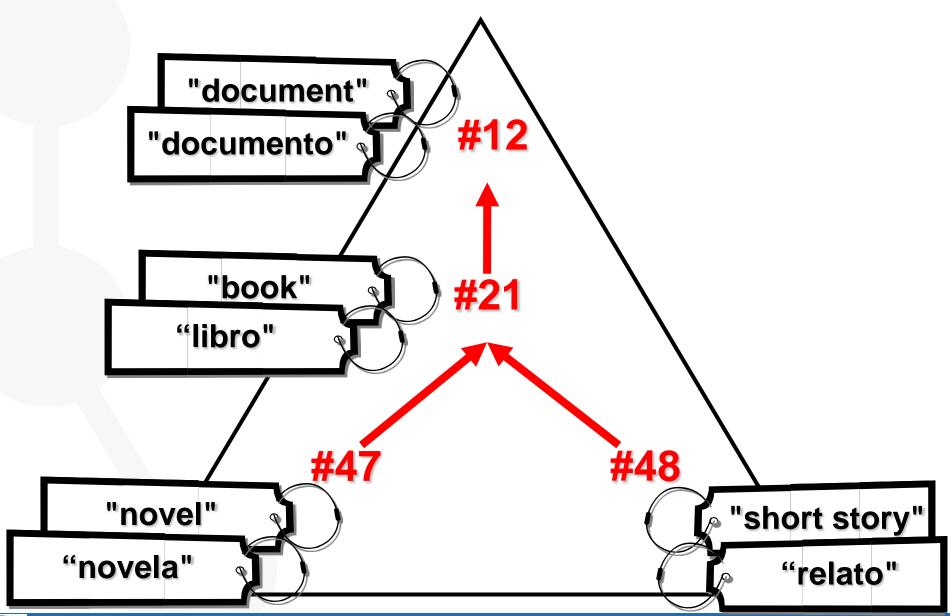
- Generated from the English Wikipedia
- Conceptualization initially in one language
- Certain problems with other languages
 - mismatches between concepts
 - Different granularity
 - Different conceptualizations

- A machine readable markup language such as XML
- Metadata, set of data that gives information about other data -> to exchange data among machines and understandable by humans.
- A common language, OWL, that expands
 RDFS, to express ontologies.

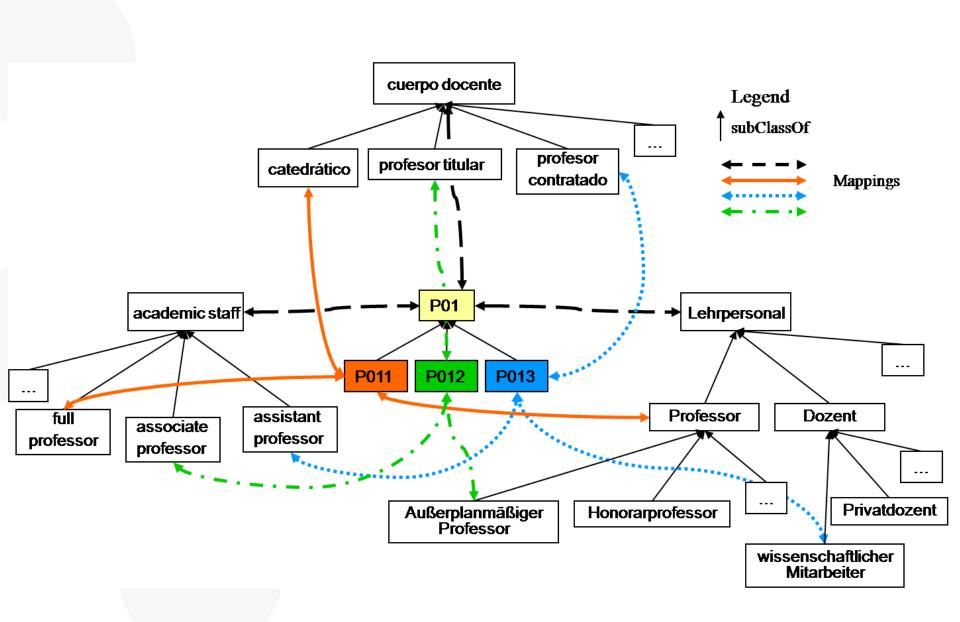
Example of a "toy" domain ontology



Multilingual ontologies



Example: culturally-influenced domain



Splicing in Dpbedia

ES _ [15:47]



TSIB 2013

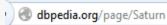
e

ĒC

How can Dbpedia be exploited by terminologists?

- Non- experts
 - A lot of work is required to complete other languages infoboxes
 - Mappings between languages are needed
 - Dbpedia provides tooos to edit these mappings
- Applying NLP techniques
 - Question answering
 - Discourse strategies to create new queries
 - More user-friendly interfaces in NLP

Saturno



About: http://es.dbpedia.org/resource/Saturno

About: http://es.dbpedia.org/resource/Sombrero Saturno

An Entity of Type: Thing, from Named Graph: http://es.dbpedia.org, within Data Space: es.dbpedia.org

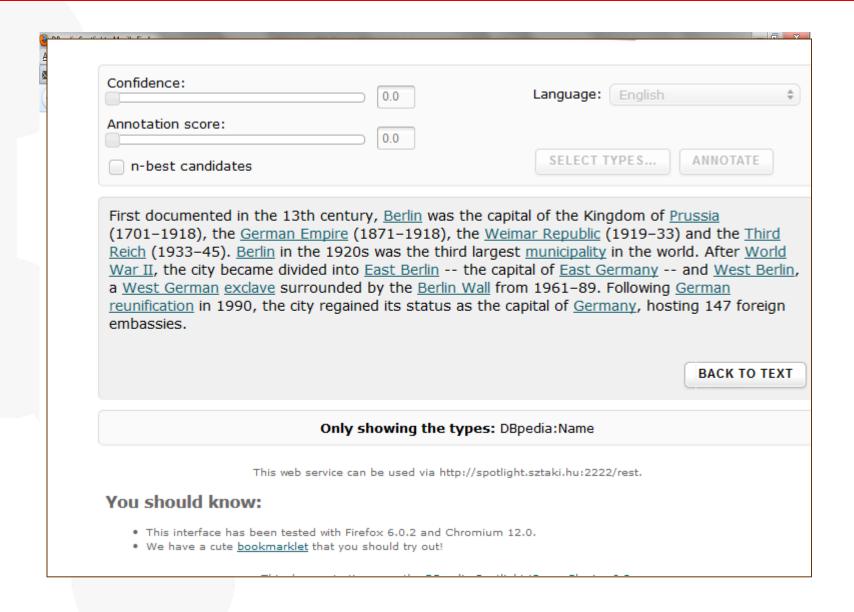
Property			Value
dbpedia-owl:wikiPageWikiLink			dbpedia:Sombrero_de_teja
is dbpedia-owl:wikiPageDisambiguates of			 dbpedia:Saturno
is dbpedia-owl:wikiPageWikiLink of			dbpedia:Capelodbpedia:Saturno
dbpprop:adjectives	 http://pianetarynames.wr.usgs.gov/ http://www.solarviews.com/eng/sati http://www.flickr.com/photos/vontor Saturnian, Cronian 		 dbpedia:Zaturno dbpedia:Saturno_(Buenos_Aires) dbpedia:Sombrero_Saturno dbpedia:Saturno
dbpprop:align	■ right	is dbpedia-owl:wikiPageDisambiguates of	■ dbpedia:Saturn
dbpprop:angularSize dbpprop:argPeri dbpprop:ascNode dbpprop:atmosphere dbpprop:atmosphereComposition dbpprop:avgSpeed dbpprop:axialTilt dbpprop:b	 15 (xsd:integer) 336 (xsd:integer) 114 (xsd:integer) yes ~96% hydrogen ~3% helium ~0.4% 9.69 27 (xsd:integer) Solar System/Saturn 	is dbpedia-owl:wikiPageWikiLink of	 dbpedia:Viento dbpedia:Febe (satélite) dbpedia:Ojo (ciclón) dbpedia:Archipiélago de las Berlengas dbpedia:Universo matemático dbpedia:Reyes Demonio (personajes de Digimon) dbpedia:Dead Space 2 dbpedia:29P/Schwassmann-Wachmann dbpedia:El mañana es ayer (Star Trek: La serie original) dbpedia:Yo, Mudd (Star Trek: La serie original) dbpedia:Omega Supreme
dbpprop:caption	 Saturn in natural color, photographe The rings of Saturn are the most m False-color UV image of Saturn's o 	http://es.dbpedia.o	rg/Wiki.jsp?page=Ackno

Extracting Name Entities:

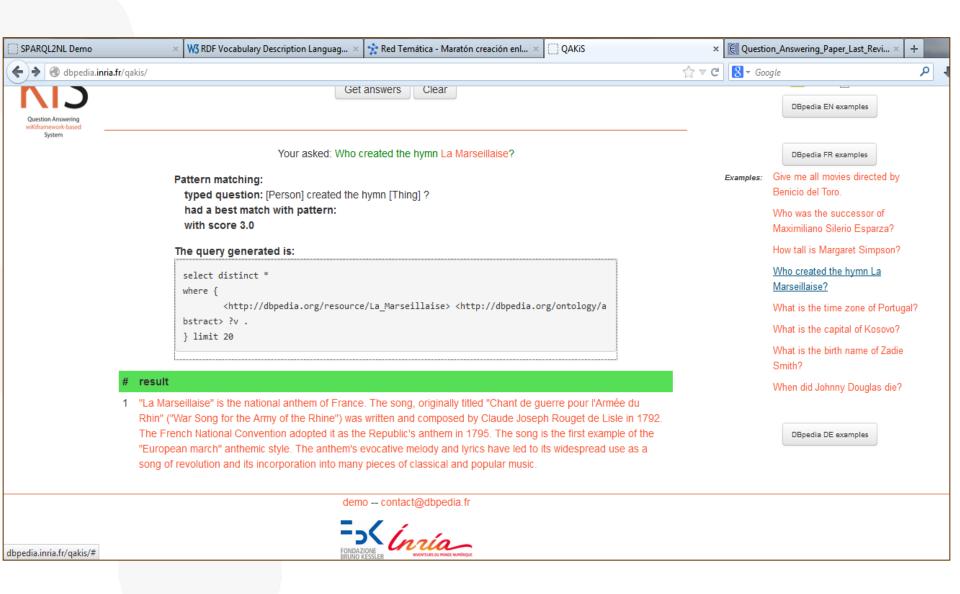
http://dbpedia-spotlight.github.io/demo

- Question- answering systems
- http://dbpedia.inria.fr/qakis
- NL queries in Spanish

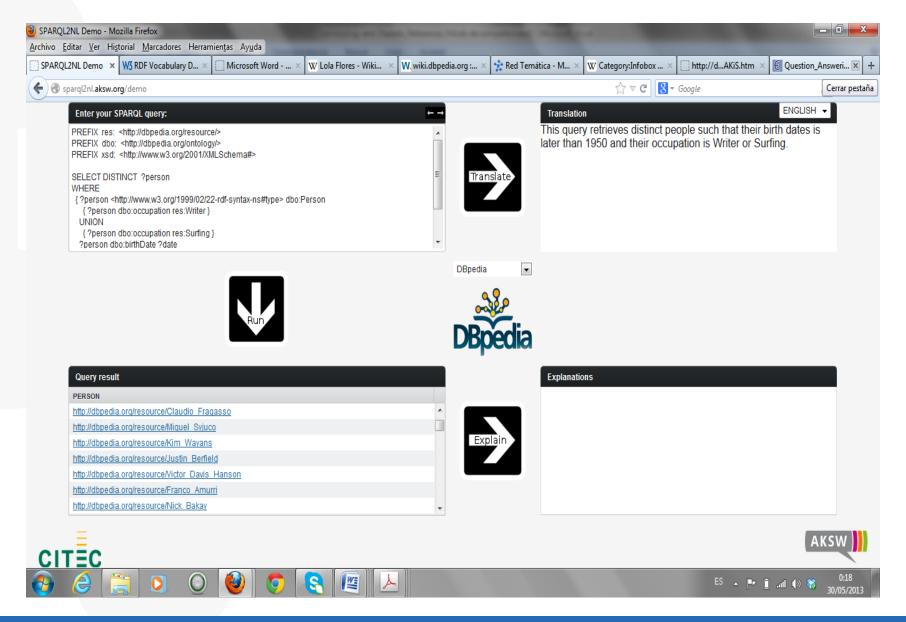
Name Entities extraction



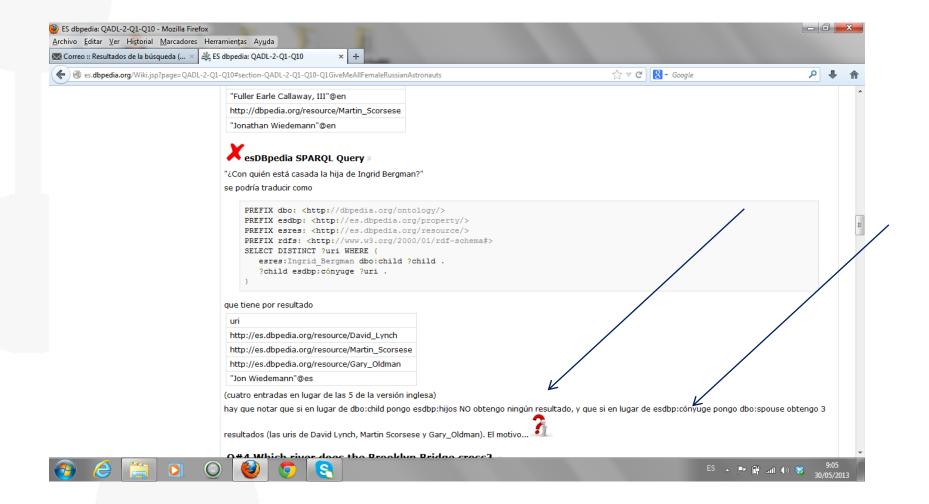
Question-answering application



SPARQL 2NL



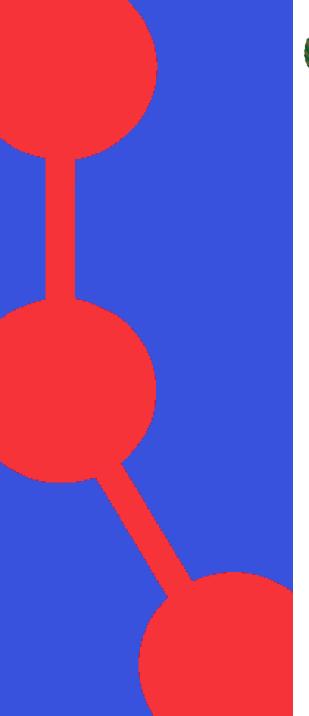
Some queries in Spanish



32

Concluding remarks and open challenges

- Evolution of Terminology according to needs, theories and users
- Nowadays, need of developing new tools to exploit available information
- Collaboration between technicians and linguists is needed
- Integratation of linguistic, communicative, cognitive and knowldge representation with the economic, industrial technological and professional needs in communication.
- New representation models have developed that try to integrate more terminologial information: W3C-, Ontolex







Terminology and Dbpedia: Back to basics?

GUADALUPE AGUADO DE CEA

lupe@fi.upm.es

Ontology Engineering Group

http://www.oeg-upm.net/

Universidad Politécnica de Madrid