

BabelNet, the LLOD cloud and the Industry

Roberto Navigli

DIPARTIMENTO
DI INFORMATICA



SAPIENZA
UNIVERSITÀ DI ROMA

Linguistic Computing Laboratory
<http://lcl.uniroma1.it>

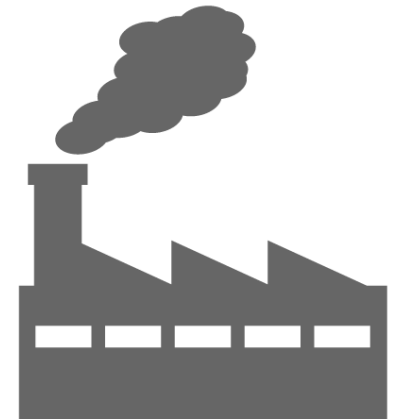
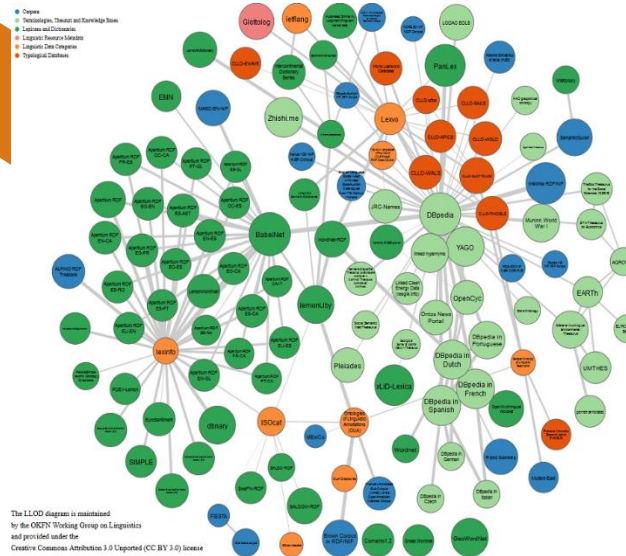
The Luxembourg BabelNet Workshop – Session 4

Session 4 – The Luxembourg BabelNet Workshop

[15.45-17.00, 2 March, 2016]

- BabelNet in the Linguistic Linked Open Data (LLOD) cloud
- Industrial applications
- Lightning talks II

BabelNet, the LLOD cloud and the Industry

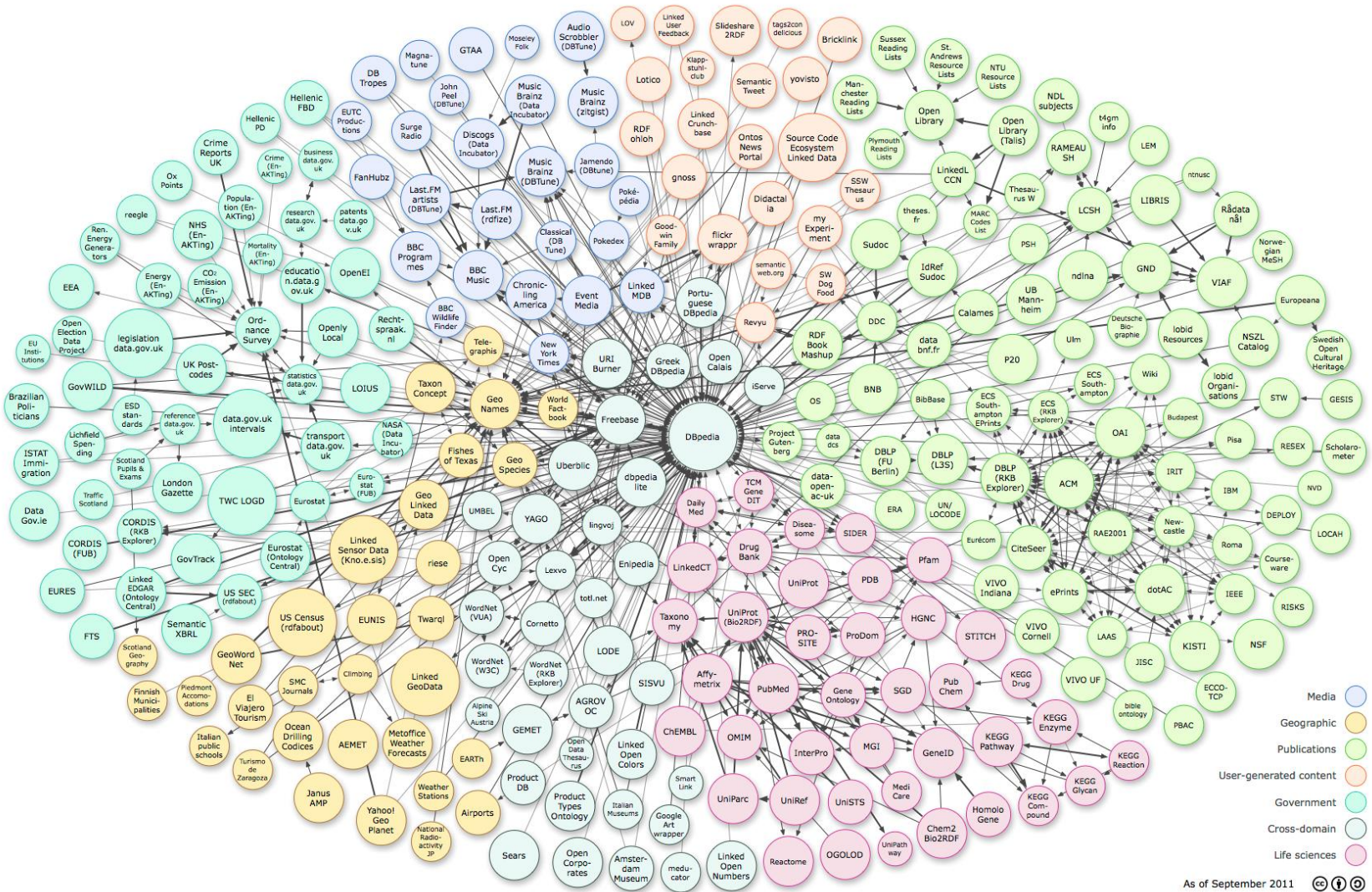


MAKE FRIENDS

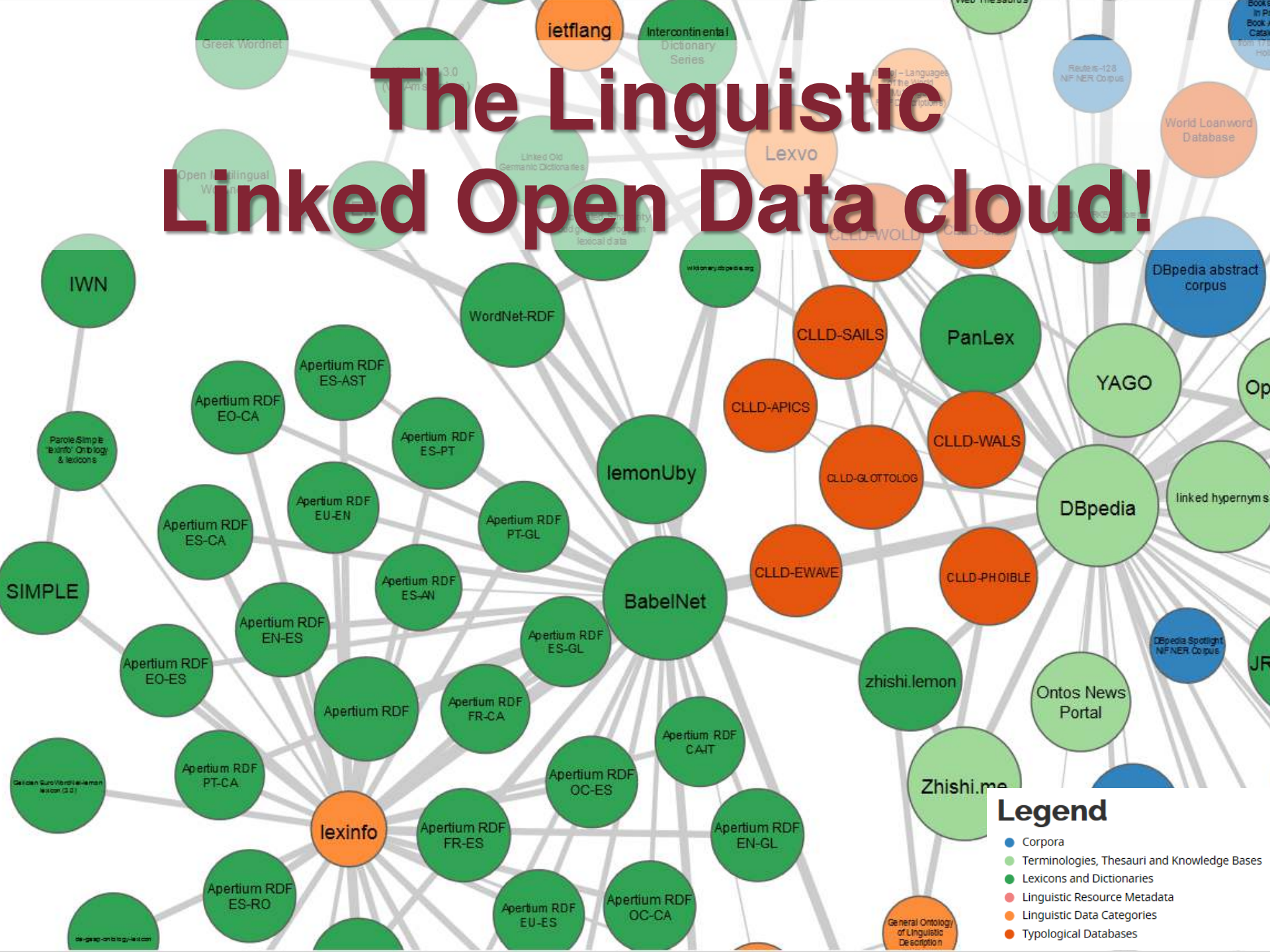
A meme featuring Woody and Buzz Lightyear from the movie Toy Story. Woody is on the left, looking slightly concerned. Buzz is on the right, wearing his iconic green and purple space suit and waving with a purple ring on his finger. The background is a simple, slightly blurred indoor setting.

FRIENDS EVERYWHERE

The Linked Data cloud



The Linguistic Linked Open Data cloud!



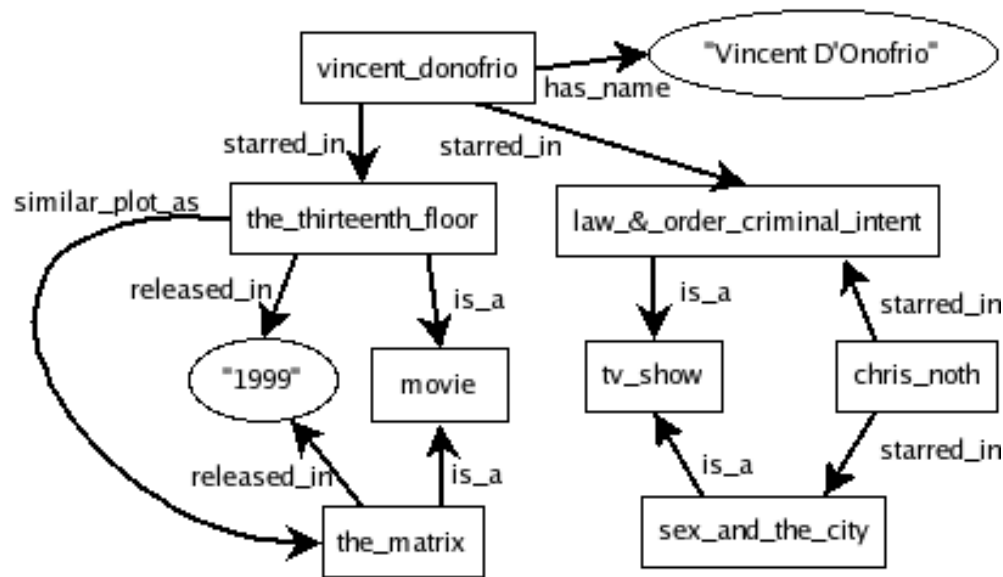
- Legend**
- Corpora
 - Terminologies, Thesauri and Knowledge Bases
 - Lexicons and Dictionaries
 - Linguistic Resource Metadata
 - Linguistic Data Categories
 - Typological Databases

RDF-Lemon encoding of BabelNet

- RDF representation based on:
 - Dublin Core:
 - <http://purl.org/dc/terms/> and <http://purl.org/dc/elements/1.1/>
 - lemon
 - <http://www.lemon-model.net/lemon#>
 - SKOS (Simple Knowledge Organization System)
 - <http://www.w3.org/2004/02/skos/core#>
 - LexInfo 2.0
 - <http://www.lexinfo.net/ontology/2.0/lexinfo#>
 - BabelNet model:
 - domain name: <http://babelnet.org/rdf/>
 - <http://babelnet.org/model/babelnet#>

What is RDF?

- The **Resource Description Framework** (RDF) is the W3C standard for **encoding knowledge**
- A general framework for describing any kind of resource ranging from concepts to real world objects such as Web sites, physical devices, etc.

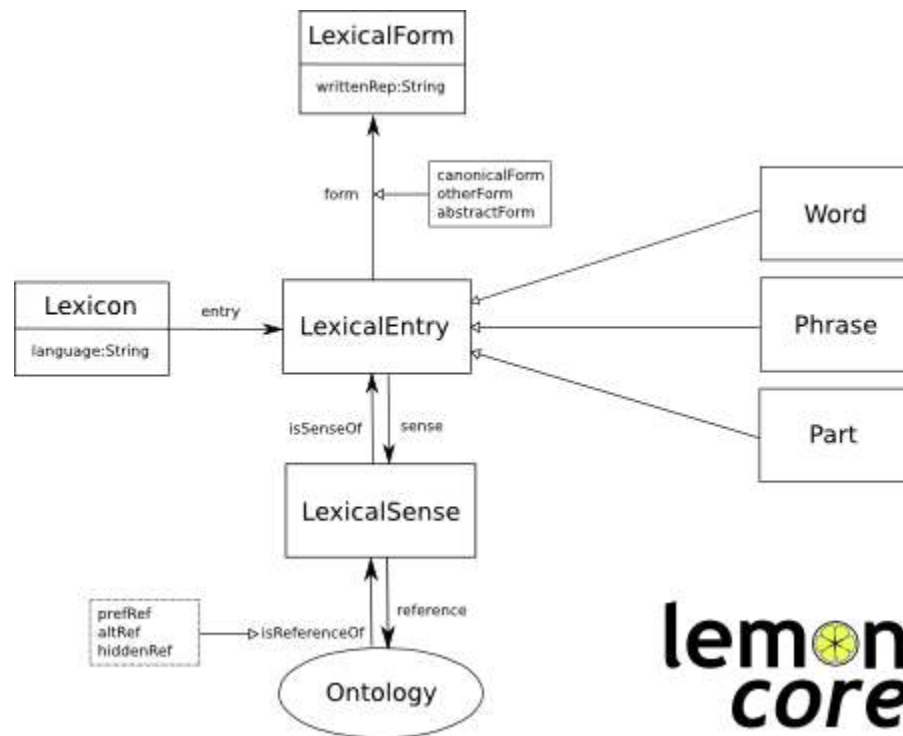


What is Dublin Core?

- A standard aimed at creating a digital "library card catalog" for the Web
- Dublin Core provides 15 metadata elements (data that describes data), including:
 - title (the name given to the resource)
 - creator (the person or organization who created it)
 - subject (the topic covered)
 - description (a textual outline of the content)
 - publisher (who made the resource available)
 - contributor (people who contributed content)
 - date (when the resource was made available)
 - type (a category for the content)
 - language (in what language the content is written)

What is Lemon?

- A model for encoding and structuring lexicons and machine-readable dictionaries and link them to the Semantic Web and the (Linguistic) Linked Open Data cloud



What is LexInfo

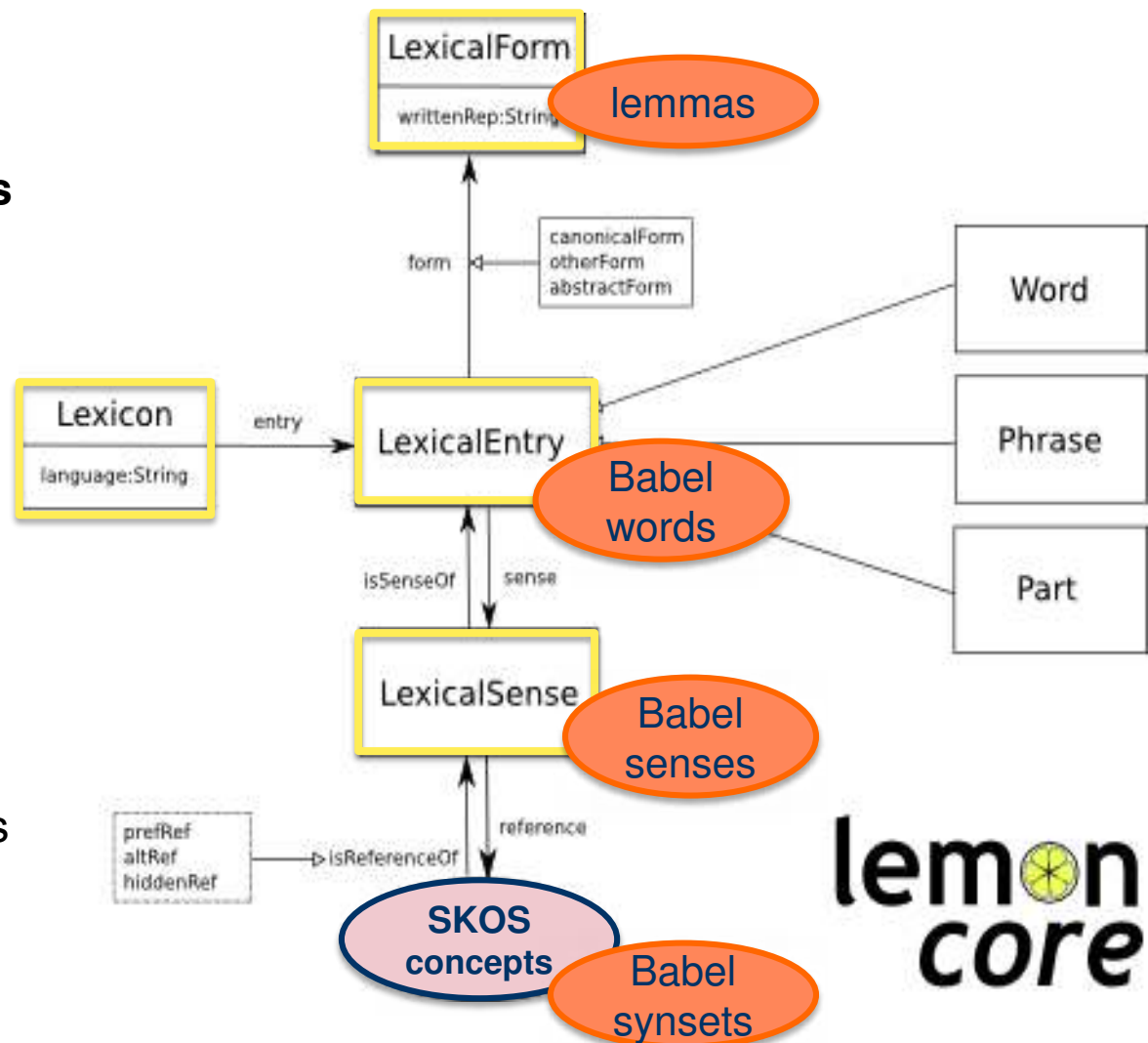
- LexInfo is a model for providing linguistic categories for the Lemon model
- Including:
 - Morphological information (tense, person, gender, etc.)
 - Part-of-speech tag information
 - Syntactic information (subject, directObject, etc.)
 - Lexical-semantic relations (antonymy, synonymy, collocations)

What is SKOS?

- **Simple Knowledge Organization System (SKOS)** is a specification for thesauri, taxonomies, subject heading lists, etc.
- Part of the Semantic Web family of standards built upon RDF
- **Objective:** easy publication and use of linked data vocabularies
- Provides **concepts** (skos:Concept) and **basic relations** (skos:related, skos:broader, skos:narrower)

BabelNet in Lemon-RDF

- the RDF resource consists of a set of **Lexicons**, one per language
- Lexicons gather **Lexical Entries** which comprise the forms of an entry; in our case: *words* of the Babel lexicon.
- **Lexical Forms** encode the surface realisation(s) of Lexical Entries; in our case: *lemmas* of Babel words.
- **Lexical Senses** represent the usage of a word as *reference* to a specific concept; in our case: Babel senses.
- **Skos Concepts** represent ‘units of thought’; in our case: Babel synsets.



Example

@prefix bn: <http://babelnet.org/rdf/>
@prefix lemon: <http://www.lemon-model.net/lemon#>
...

bn:babelNet_lexicon_EN

a **lemon:Lexicon** ;
dc:source <http://babelnet.org/>;
lemon:entry bn:Citrus_limón_EN , bn:lemon_EN , bn:Lemon_EN , bn:Lemons_EN , bn:lemon_tree_EN ...
lemon:language "EN".

bn:lemon_tree_EN

a **lemon:LexicalEntry** ;
rdfs:label "lemon_tree"@EN ;
dc:source <http://wordnet.princeton.edu/> ;
lemon:canonicalForm <http://babelnet.org/rdf/lemon_tree_n_EN/canonicalForm> ;
lemon:language "EN" ;
lemon:sense <http://babelnet.org/rdf/lemon_tree_EN/s00019309n> ;
lexinfo:partOfSpeech
lexinfo:noun .

<http://babelnet.org/rdf/lemon_tree_n_EN/canonicalForm>

a **lemon:Form** ;
lemon:writtenRep "lemon_tree"@EN .

<http://babelnet.org/rdf/lemon_tree_EN/s00019309n>

a **lemon:sense** ;
lemon:reference bn:s00019309n ;
lexinfo:translation <http://babelnet.org/rdf/柠檬树_ZH#00019309n> , <http://babelnet.org/rdf/limoeiro_PT/s00019309n> ,
<http://babelnet.org/rdf/citronnier_FR/s00019309n> , <http://babelnet.org/rdf/lemon_puno_TL/s00019309n> ...

bn:s00019309n

a **skos:Concept** ;
bn-lemon:synsetType "concept" ;
bn:wikipediaCategory
wikipedia-fr:Catégorie:Fruit_alimentaire , wikipedia-nl:Categorie:Wijnruitfamilie , wikipedia-en: Category:Tropical_agriculture , ... ;
lemon:isReferenceOf <http://babelnet.org/rdf/lemon_EN/s00019309n> , <http://babelnet.org/rdf/citron_FR/s00019309n> ,
<http://babelnet.org/rdf/Citronensaft_DE/s00019309n> , <http://babelnet.org/rdf/lemon_tree_EN/s00019309n> ...
lexinfo:membreHolonym
bn:s00037916n ;
skos:exactMatch dbpedia:Lemon , lemon-Omega:OW_eng_Synset_33386 , freebase:m.09k_b , lemon-WordNet31: 112732356-n ;
bn-lemon:definition "Die Zitrone oder Limone ist die etwa faustgroße Frucht des Zitronenbaums aus der Gattung der Zitruspflanzen."@DE , "A small evergreen tree that originated in Asia
but is widely cultivated for its fruit"@EN ,
"Le citron est un agrume, fruit du citronnier qui a un PH acide de 2,5."@FR ;
skos:narrower bn: s00019308n ;
skos:related bn:s00188641n , bn:s00052952n , bn:s00047566n ...;

BabelNet in RDF-Lemon

Statistics between the BabelNet versions

Version ↕	Release date ▲	Languages ↕	Sources ↕	Babel synsets ↕	Babel senses ↕	Concepts ↕	Named Entities ↕	Lexico-semantic relations ↕	Definitions ↕	Images ↕	RDF triples ↕
3.6	2016/01	271	13	13,801,844	745,856,326	6,066,396	7,735,448	380,239,084	40,705,588	10,767,833	1,958,820,772
3.5	2015/09	272	13	13,801,844	119,036,997	6,066,396	7,735,448	380,239,084	40,634,604	10,767,833	1,951,194,299
3.0	2014/12	271	7	13,789,332	117,204,438	6,418,418	7,370,914	354,538,633	40,328,194	10,960,634	1,932,845,662
2.5.1	2014/11	50	7	9,347,143	67,605,841	3,683,600	5,663,543	262,687,848	21,751,423	7,764,270	-
2.0.1	2014/03	50	5	9,348,287	50,282,923	3,684,512	5,663,775	262,687,848	17,961,157	7,764,270	1,138,337,378
1.1	2013/01	6	4	5,581,954	21,947,381	1,566,566	4,015,388	141,697,438	8,439,497	6,494,690	-

BabelNet and its links

- 2 billion triples

Links to:

- WordNet-RDF
- Wiktionary
- Apertium
- lemonUby
- Lexinfo
- DBpedia
- Zhishi.lemon
- **YAGO**
- **Freebase**



Online availability of BabelNet RDF

- SPARQL endpoint
 - virtuoso universal server
 - <http://babelnet.org/sparql/>
- Dereferencing
 - <http://babelnet.org/rdf/>
 - Pubby Linked Data Frontend
(<http://wifo5-03.informatik.uni-mannheim.de/pubby/>)

Let's see some examples of SPARQL queries

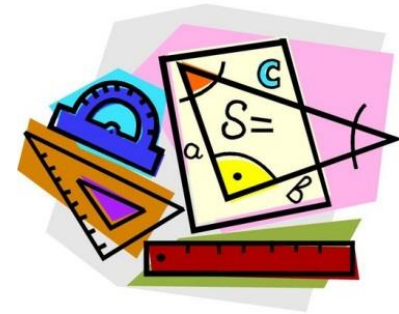
- Open a browser
- Enter the following URI: <http://babelnet.org/sparql/>



Example 1: Retrieve the senses of a given lemma

- Given a word, e.g. home, retrieve all its senses and corresponding synsets in all supported languages:

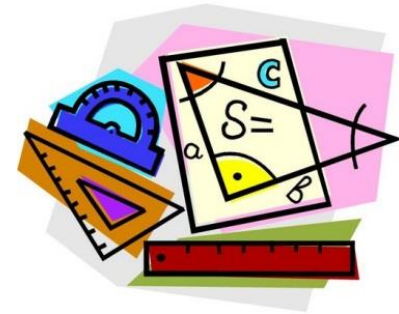
```
SELECT DISTINCT ?sense, ?synset WHERE {  
  ?entries a lemon:LexicalEntry .  
  ?entries lemon:sense ?sense .  
  ?sense lemon:reference ?synset .  
  ?entries rdfs:label "home"@en .  
}
```



Example 1: Retrieve the senses of a given lemma

- Given a word, e.g. home, retrieve all its senses and corresponding synsets in all supported languages:

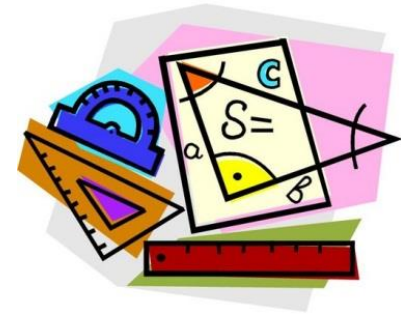
```
SELECT DISTINCT ?sense, ?synset WHERE {  
  ?entries a lemon:LexicalEntry .  
  ?entries lemon:sense ?sense .  
  ?sense lemon:reference ?synset .  
  ?entries rdfs:label "casa"@it .  
}
```



Example 2: Retrieve the translation of a given sense

- Given the URI of a given word sense:

```
SELECT ?translation WHERE {  
  ?entry a lemon:LexicalSense .  
  ?entry lexinfo:translation ?translation .  
  FILTER  
  (?entry=<http://babelnet.org/rdf/house\_EN/s00044994n>)  
}
```



Example 3: retrieve the license information about a given sense

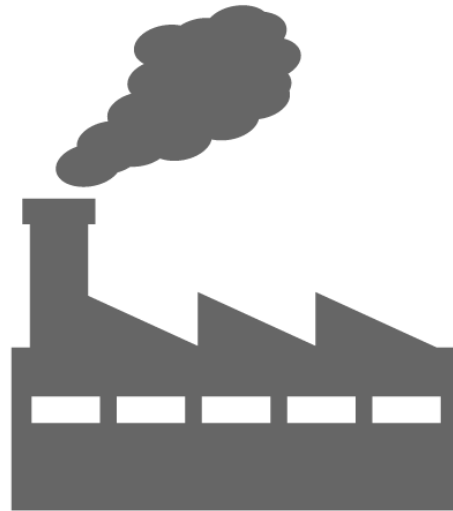
```
SELECT ?license WHERE {  
  ?entry a lemon:LexicalSense .  
  ?entry dcterms:license ?license .  
  FILTER  
    (?entry=<http://babelnet.org/rdf/Bill\_Gates\_%28Microsoft%29\_EN/s00010401n>)  
}
```

Example 4: retrieve textual definitions for a synset

```
SELECT DISTINCT ?language ?gloss ?license ?sourceurl
WHERE {
  ?url a skos:Concept .
  ?url bn-lemon:synsetID ?synsetID .
  OPTIONAL {
    ?url bn-lemon:definition ?definition .
    ?definition lemon:language ?language .
    ?definition bn-lemon:gloss ?gloss .
    ?definition dcterms:license ?license .
    ?definition dc:source ?sourceurl .
  }
  FILTER (?url=<http://babelnet.org/rdf/s03083790n>)
}
```

Industrial applications

- Concept and Named Entity Extraction
- Dictionary of the future
- New publishing initiatives
- Computer-Assisted Translation
- Multilingual Text Analytics



Industrial application: Concept and Named Entity Extraction

Who is interested: any company interested in extracting knowledge from their document base

Concept and Named Entity Extraction



- Compared to terminology extraction, which extracts **only terms**, here we extract **concepts** and **named entities**
- With BabelNet:
 - in **any language**
 - making the **outputs comparable across languages**

Concept and Named Entity Extraction: an example

Titolo

Natale 2015 a Milano: eventi, mostre e mercatini, ecco cosa fare in città

Testo:

Milano si veste a festa per il Natale 2015. Tante le occasioni per divertirsi, fare acquisti e concedersi qualche golosità da dicembre fino all'Epifania. Mercatini, bancarelle, piste di pattinaggio, alberi di Natale e mostre da non perdere. Ecco cosa fare in città. **MERCATINI** - A Milano non mancano le occasioni per fare acquisti in vista delle feste. Oltre ai tradizionali Oh bej oh bej (Tutte le informazioni sugli Oh Bej Oh Bej), in programma dal 5 all'8 dicembre in piazza Castello all'ombra della Castello Sforzesco, torna anche l'altrettanto amato appuntamento con L'Artigiano in Fiera, dal 5 al 13 dicembre a Rho Fiera Milano (Tutte le informazioni su L'Artigiano in Fiera). **ARTE**- L'Adorazione dei pastori di Rubens è la magnifica opera d'arte che sarà visitabile a Palazzo Marino nel periodo di Natale 2015. L'appuntamento con i capolavori dell'arte nel Municipio del Comune di Milano si rinnova dunque dal 3 dicembre al 10 gennaio 2016 in Sala Alessi. L'ingresso sarà libero e possibile tutti i giorni: i visitatori saranno accolti da storici dell'arte, coordinati da Civita, che faranno da guida nel percorso espositivo (Tutte le informazioni sulla mostra). **MUSICA E CONCERTI** - Una finestra aperta su piazza Duomo e un complesso jazz pronto a allietare il pomeriggio di milanesi e turisti con mezz'ora di canzoni natalizie. Una trasposizione musicale del calendario dell'avvento. È quanto ha organizzato palazzo Marino in collaborazione con Fondazione Milano (Tutte le informazioni sui concerti di piazza Duomo) **PISTE DI PATTINAGGIO** - Siete sportivi? Non riuscite a resistere al fascino del ghiaccio? Proprio all'ombra della Madonna durante il periodo natalizio verranno allestite quattro piste temporanee. Di seguito l'elenco con tutti gli impianti — sia «permanenti» che «temporanei» presenti in città (Leggi l'elenco delle piste di pattinaggio aperte per le feste). **ALBERI E LUCI DI NATALE** - Oltre al tradizionale albero in piazza Duomo, quest'anno largo anche a un Natale tutto di luci, nella piazza più all'avanguardia di Milano, piazza Gae Aulenti (Tutte le informazioni sull'Albero della Luce). Per una passeggiata magica non c'è che l'imbarazzo della scelta: sono 150 le vie illuminate per le feste. **ALLA DARSENA** - Una vera full immersion nell'atmosfera delle feste alla Darsena, con il Christmas Village. A disposizione di tutti Il Bosco delle Fiabe, la casa di Babbo Natale, ma anche pattinaggio sul ghiaccio e un albero di Natale sull'acqua. Tra le proposte anche un Mercatino di Natale che offrirà prodotti food e artigianali; allestimenti e luminarie natalizie con il CuboLed trasformato, per l'occasione, in Calendario dell'Avvento. (Tutte le informazioni sul Christmas Village) **GIARDINI MONTANELLI** - Al via la 9° edizione del fantastico Villaggio delle Meraviglie, un magico angolo nel cuore di Milano ricco di giochi e attrazioni per grandi e piccini. L'appuntamento d'eccezione che darà il via ufficiale a questa nuova stagione di divertimenti sarà però quello con l'amatissimo Geronimo Stilton “in pelliccia e baffi” che lunedì 7 dicembre dalle 15 intratterrà bambini e adulti con il suo “stratopico” spettacolo “Grande Ritorno nel Regno della Fantasia 2”. Il Villaggio - ideato e organizzato da Ambraorfei.it e Orizzonte Eventi, con il sostegno dell'Assessorato Benessere, Qualità della vita, Sport e Tempo Libero – animerà dal 5 dicembre al 10 gennaio 2016 la suggestiva area di Porta Venezia, all'interno dei magnifici giardini Indro Montanelli, diffondendo un clima di festa con spettacoli e speciali novità tutte da scoprire. (Tutte le informazioni)

Concept and Named Entity Extraction: an example

- Key terms extracted with statistical techniques:






Concept and Named Entity Extraction: an example



- Key conceptual categories:






Concept and Named Entity Extraction: an example

▲	Terminologia ◆	Immagine ◆	Definizione ◆
1	mercatino		Un mercato delle pulci è un mercato in cui si commerciano oggetti di poco valore e usati.
2	Natale		Il Natale è una festività cristiana che celebra la nascita di Gesù.
3	a milano		a milano
4	Milano		città capoluogo della Lombardia

Concept and Named Entity Extraction: an example

5	Calendario dell' Avvento		Calendario dell' Avvento
6	Calendario		<p>Il calendario è un sistema adottato dall'uomo per suddividere, calcolare e dare un nome ai vari periodi di tempo.</p>
7	Avvento		<p>L'avvento, in molti riti cristiani, è il tempo liturgico che precede il Natale ed è preparatorio allo stesso: nei riti cristiani occidentali segna l'inizio del nuovo anno liturgico.</p>

Concept and Named Entity Extraction: an example

7	Avvento		L'avvento, in molti riti cristiani, è il tempo liturgico che precede il Natale ed è preparatorio allo stesso: nei riti cristiani occidentali segna l'inizio del nuovo anno liturgico.
8	pattinaggio	NO IMAGE	The sport of gliding on skates
9	mostra		raccolta specialmente di opere d'arte o di prodotti agricoli, industriali o dell'artigianato, che vengono sistemati in appositi luoghi visitabili dal pubblico
10	bej		bej
11	palazzo marino		Palazzo Marino, opera dell'architetto perugino Galeazzo Alessi, è un palazzo nobiliare di Milano, sede dell'amministrazione comunale dal 19 settembre 1861.
12	oh		oh
13	Christmas Village		Christmas Village

Industrial application: The Dictionary of the Future

Who is interested: all producers of language
resources

Today's (Paper) Dictionaries

- They are **not hypertexts**
- Organized **alphabetically** (except for thesauri and analogical dictionaries)
- Not "**browsable**" by semantic correlation, synonymy or other criteria



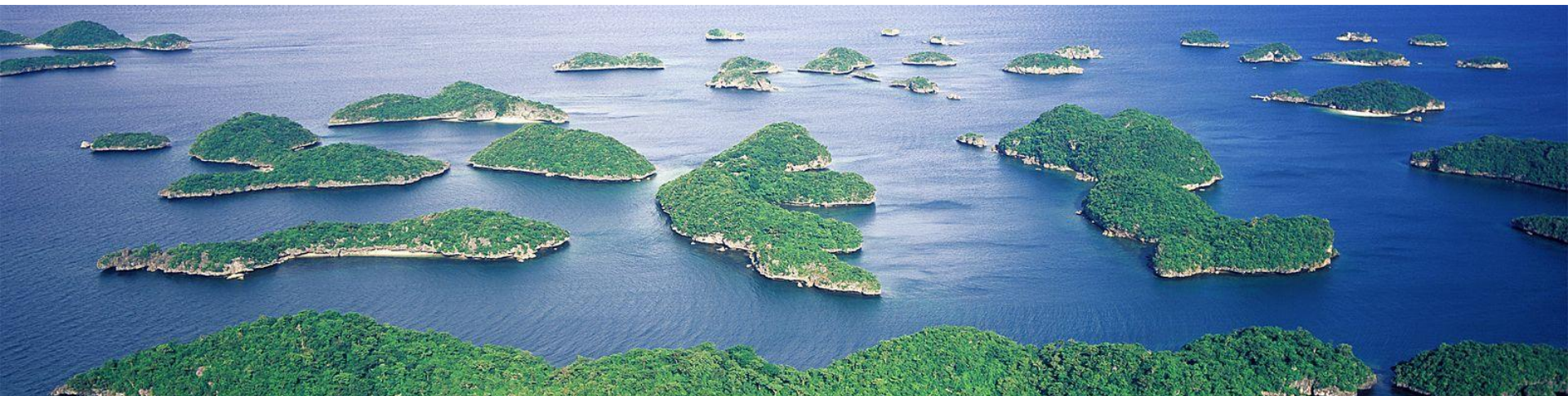
Machine-readable dictionaries

- Machine-readable dictionaries as conversions of paper dictionaries that connect **words** via hyperlinks
- **Problem 1: these links are not "disambiguated"**, as they do not take you to the appropriate meaning of the word
 - Es. If I click on *sveglio* (meaning both awaken and apt, smart), I will be taken to the entry of that word, but not on **its most suitable meaning**.



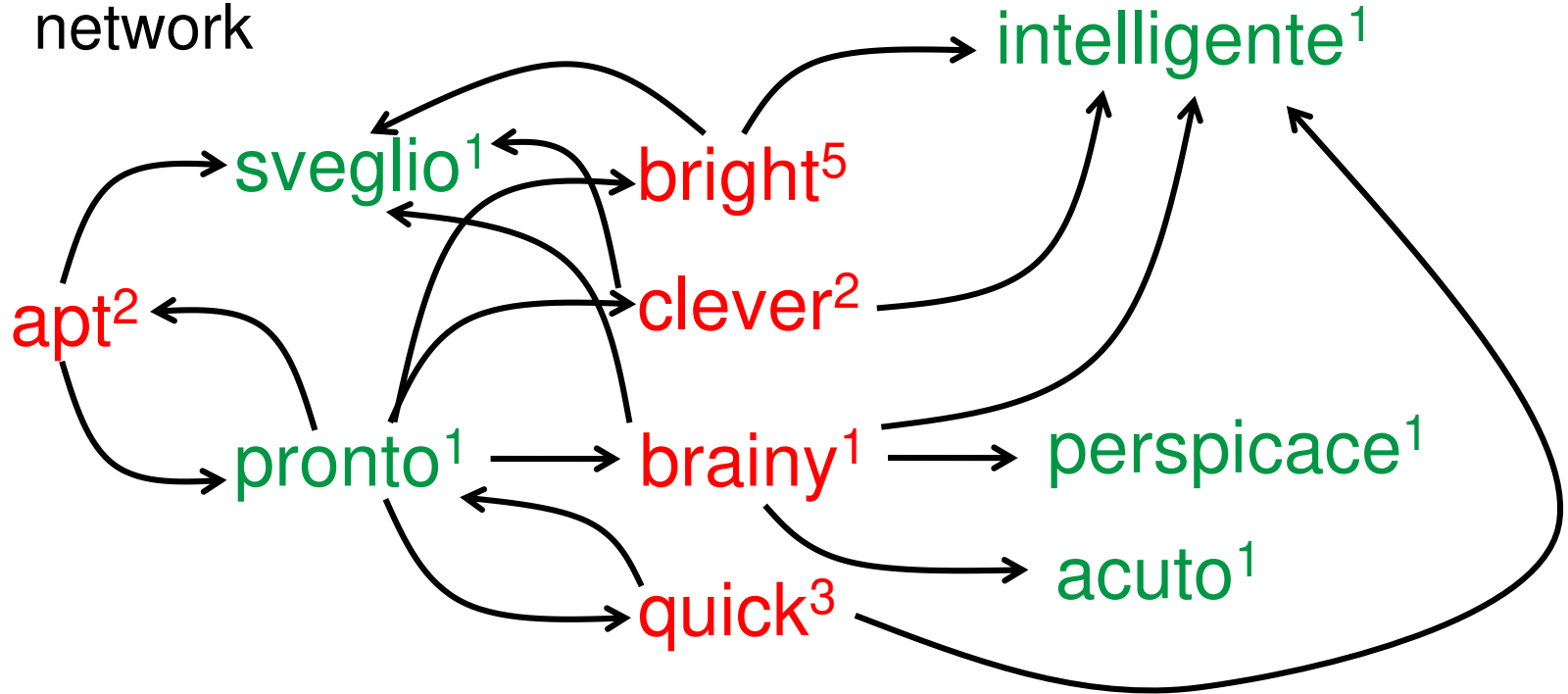
Dictionaries are like Islands without Bridges!

- **Problem 2:** monolingual and bilingual dictionaries are disconnected from each other
- But we could use the source language as a pivot to connect translations across languages (e.g. English-Italian, English-French, English-Chinese, etc.), so as to produce multilingual entries, like in BabelNet
- This requires **advanced disambiguation techniques**



What can Language Resource Producers do?

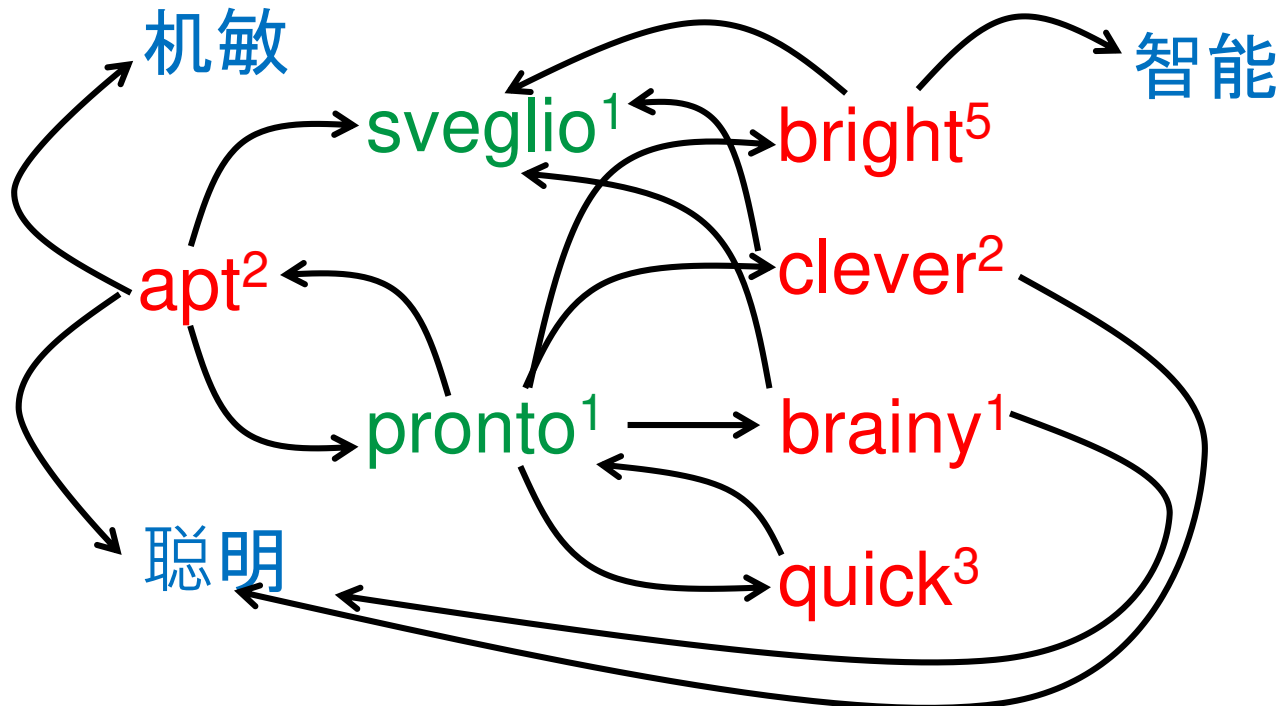
1. **Disambiguate translations of their bilingual dictionaries**, so as to create a browsable semantic network



2. Similarly with **phrases** and **example sentences**

What can Language Resource Producers do?

- Connect source language meanings across languages in the different bilingual dictionaries, obtaining a **multilingual dictionary and network**



- Beyond **innovative browsing experience** for the user, you can identify **missing meanings** or **potential mistakes**

What can Language Resource Producers do?

- Connect different monolingual dictionaries and thesauri, thus creating a **richer and unified experience** for the user
 - Think of **thousands of glossaries** on the Web and **EU resources**
- The user does not have to choose between a traditional dictionary, an analogical dictionary and a thesaurus, but will find **all the information aggregated in a single entry** (but keeping the links back to the respective resources)
- **Definition:** Il più comune degli aeromobili a sustentazione dinamica
- **Synonyms:** aeroplano, aereo, apparecchio
- **Actions:** pilotare, volare con, imbarcare, sbarcare, ecc.
- **Collocations:** a due piani, invisibile, canadair, ecc.

Industrial application: New Publishing Initiatives

Who is interested: all producers of educational and technical multimedia content

New Publishing Initiatives

- **Annotate text** with concepts, definitions and images
- **School texts**, also in multiple languages



Babelfy

LOG IN REGISTER

Era geologica. In geocronologia, un'era geologica è una delle suddivisioni della scala dei tempi geologici; nelle ere più recenti, è normalmente compresa tra due estinzioni di massa. Le ere geologiche sono (dalla più antica alla più recente): Eoarcheano, Paleoarcheano, Mesoarcheano, Neoarcheano, Paleoproterozoico, Mesoproterozoico, Neoproterozoico, Paleozoico, Mesozoico, Cenozoico. Solo l'èone Adeano non è suddiviso in ere.

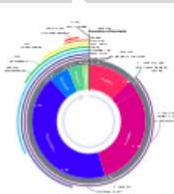
Enable partial matches:

ITALIAN

BABELFY!

expanded view | compact view

In **geocronologia**, un'era **geologica** è una delle **suddivisioni** della **scala dei tempi geologici**; nelle ere



geocronologia

The longest division of geological time

geologica

Of or relating to or based on geology



suddivisioni

A residential area of similar dwellings built




scala dei tempi geologici

The longest division of



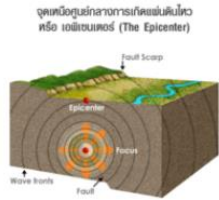
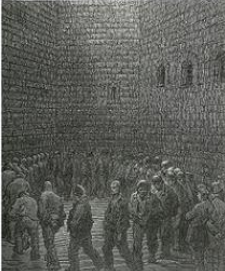
Consider the semantic annotation of free text

Absah, 41, as she searched for her 11 offspring in Banda Aceh, Indonesia, the city closest to the epicentre of Sunday's earthquake Like many, many prisoners before me, both well-known and unknown, I should say Thank you to prison.

<< 1 2 >>			
BabelNetID	Sense	BabelNet definition	Image
bn:00090460v	searched	Try to locate or discover, or try to establish the existence of	
bn:00525133n	Banda Aceh, Indonesia	Banda Aceh is the capital and largest city in the province of Aceh, Indonesia.	
bn:00946641n	Aceh, Indonesia	Aceh ; Atjeh ; or Acheh is a special region of Indonesia.	
bn:03335997n	city	A city is a relatively large and permanent human	



Consider the semantic annotation of free text

Absah, 41, as she searched for her 11 offspring in Banda Aceh, Indonesia, the city closest to the epicentre of Sunday's earthquake Like many, many prisoners before me, both well-known and unknown, I should say Thank you to prison.

bn:00031140n	epicentre	The point on the Earth's surface directly above the focus of an earthquake	 <p>A diagram showing a cross-section of the Earth's crust. A red starburst labeled 'Focus' is located underground. A vertical line extends from the focus to the surface, where it is labeled 'Epicenter'.</p>
bn:00028190n	Sunday	First day of the week; observed as a day of rest and worship by most Christians	 <p>An illustration of a man in a suit sitting in a chair, looking towards a woman in a long purple dress who is standing near a window. A small dog is on the floor in the foreground.</p>
bn:00029448n	earthquake	Shaking and vibration at the surface of the earth resulting from underground movement along a fault plane or from volcanic activity	 <p>A 3D diagram of an earthquake. It shows a fault line with a 'Fault Scarp' on the surface. Below the surface, the 'Epicenter' is marked on the fault plane, and the 'Focus' is shown deeper underground. Concentric circles represent 'Wave fronts' radiating from the focus. Thai text at the top reads: 'จุดศูนย์กลางการเกิดแผ่นดินไหว หรือ เอพิเซนเตอร์ (The Epicenter)'.</p>
bn:00015767n	prisoners	A person who is confined; especially a prisoner of war	 <p>A black and white photograph showing a group of men, likely prisoners of war, standing in a line in a camp. They are wearing simple, worn clothing and are surrounded by stone walls.</p>

Consider the semantic annotation of free text

Mikhail Khodorkovsky, the Russian tycoon and former chief executive of Yukos, who is charged with fraud and tax evasion I think there are one or two worse players, but they're having difficulties admitting it Alexander McCall Smith, creator of The No. 1 Ladies' Detective Agency, says that he finds therapy in playing bassoon for the Really Terrible Orchestra I was thinking of starting a PhD, but I've realised I can't use my degrees because no one will give me a job.

BabelNetID	Sense	BabelNet definition	Image
bn:02354051n	Mikhail Khodorkovsky	Mikhail Borisovich Khodorkovsky is a Russian businessman and former oligarch, as well as philanthropist, public figure and author.	
bn:00110004a	Russian	Of or pertaining to or characteristic of Russia or its people or culture or language	
bn:00008639n	tycoon	A very wealthy or powerful businessman	

Example:

You are running a booking or reviewing web site

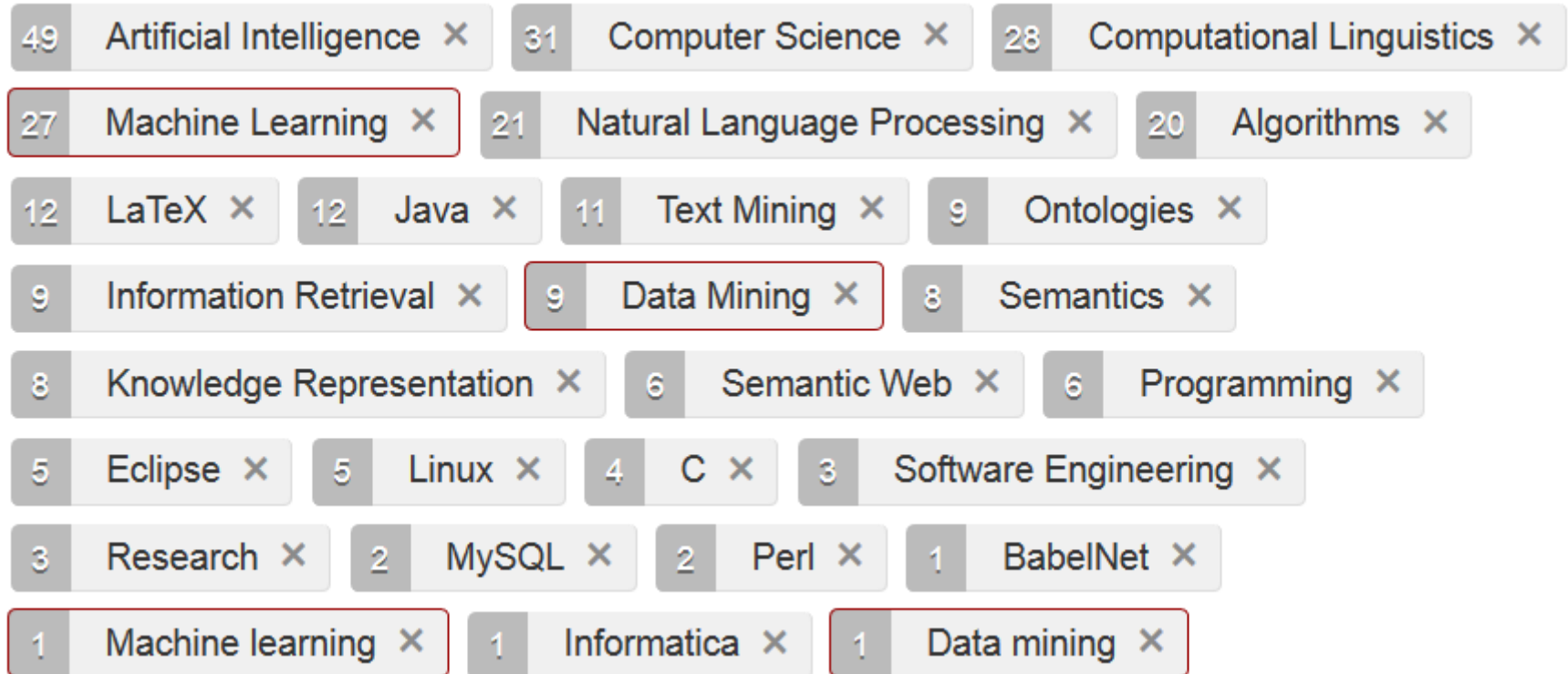
Assume you want to **semantically analyze your reviews**:

- "Una serata stupenda all'insegna della cucina romana" -Recensito 2 settimane fa
- E' impressionante come questo locale mantenga elevatissima la qualità e la freschezza degli ingredienti, grande la passione nella cucina e prezzi assolutamente competitivi! Dei tantissimi locali da me visitati fino a oggi, questo ha di gran lunga il miglior rapporto qualità prezzo a Roma (e, oserei dire, tra i migliori in Italia). E' vero, ogni tanto non hanno il menù, ma il conto è, ieri come sempre, impeccabile (tra i 25 e i 30 euro a testa per due primi di paccheri carciofi e guanciaie, un secondo di pesce - un'ombrina freschissima - e 4 contorni di verdure di stagione, tra cui carciofi, broccoletti, cicoria). Fantastici! Continuate così!

What can you do?

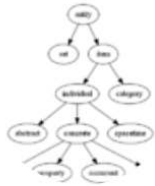
- People can search **by dish independently of their language**
 - What restaurants serve spaghetti with cheese and pepper?
 - Quels restaurants servent des artichauts?
- People can **search by similarity**
 - Give me all the restaurants that cook dishes containing ingredients similar to paccheri al pomodoro?
- People can have **explanations of the ingredients or dishes cooked**
- People can **browse and explore dishes**

Same can be done with unstructured tags



Same can be done with unstructured tags

artificial intelligence , computer science , computational linguistics , machine learning , natural language processing



artificial intelligence

The branch of computer science that



computer science

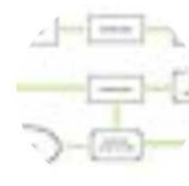
The branch of engineering science

computational linguistics

The use of computers for linguistic research and applications

computational

Of or involving computation or computers



machine learning

Machine learning is a scientific discipline that deals with the



natural language processing

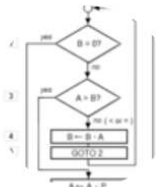
The branch of information science

natural language processing , algorithms , latex , java , text mining , ontologies



natural language processing

The branch of



algorithms

A precise rule (or set of rules) specifying



latex

LaTeX is a document preparation system

java

A platform-independent object-



text mining

Text mining, also referred to as text



ontologies

(computer science) a rigorous and

ontologies , information retrieval , data mining , semantics , knowledge representation , ser

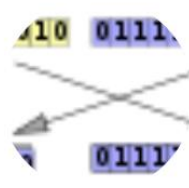


ontologies

(computer science) a



information retrieval



data mining

Data processing using



semantics

The study of language

knowledge representation

Knowledge representation and reasoning is the field of artificial intelligence

representation

A representation of the

ser

The

Same can be done with unstructured tags

- Because Babel synsets are multilingual, I can now search the database **in any language**
 - For instance, programmazione instead of programming
- I can exploit the semantic network structure, to find people who are **semantically closest** to my interests or to skills I am looking for

Industrial application: Computer-Assisted Translation

Who is interested: all producers of computer-assisted translation software

Computer-Assisted Translation (CAT)







Computer Assisted Translation (CAT)



Interactive translation, showing possible translations from Chinese to English. Definitions are grouped by pinyin pronunciation.

Computer Assisted Translation (CAT)

- From XTM's system:

A	B	C	D	E
Src Seg.	en-GB	hu-HU	Tgt Seg.	Probability
	Unmatched  Not sure  Merged  Very good 			
0	Before Adam by Jack London	NNCL1041-452v1.0	0	0.0000
1	1906	JACK LONDON	1	0.6250
2	"These are our ancestors, and their history is our history."	Ádám előtt	2	0.2764
		Fordította: SZÁSZ IMRE A fedői és a belső illusztráció VISZT GYÖRGY munkája (c) Szász Imre, 1985 Hungarian translation	3	0.2802
		ISBN 963 11 3917 4	4	0.0000
		Móra Ferenc Ifjúsági Könyvkiadó, Budapest Felelős kiadó: Sziládi János igazgató Alföldi nyomda (5783.66-15-2), Debrecen, 1985 Felelős vezető: Benkő István igazgató Felelős szerkesztő: Szathmáry Éva Műszaki vezető: Szakálos Mihály Műszaki szerkesztő: Beszédés Natasa 67 000 példány, Terjedelem: 10,47 (A/5) ív.	5	0.0000
		IF 5331 "Ezek a mi őseink, az ő történetük a mi történetünk.	6	0.0000
3	Remember that as surely as we one day swung down out of the trees and walked upright, just as surely, on a far earlier day, did we crawl up out of the sea and achieve our first adventure on land."	Ne feledjük el: amilyen biztos az, hogy egy napon leereszkedtünk a fákról, és egyenesen kezdtünk járni, éppoly biztos, hogy egy sokkalta régebbi napon kimásztunk a tengerből, és átöltük első kalandunkat a szárazföldre."	7,8	1.0000
4	CHAPTER I	1.	9	1.0000
5	Pictures!	Képek!	10	1.0000
6	Pictures!	Képek!	11	1.0000
7	Pictures!	Képek!	12	1.0000
8	Often, before I learned, did I wonder whence came the multitudes of pictures that thronged my dreams; for they were pictures the like of which I had never seen in real wake-a-day life.	Míg meg nem tudtam, gyakran csodálkoztam, honnan származik az a rengeteg kép, ami előzőnlötte álmaimat; mert ilyen képeket sose láttam a nappal, az ébrenlét valóságában.	13	0.8925
9	They tormented my childhood, making of my dreams a procession of nightmares and a little later convincing me that I was different from my kind, a creature unnatural and accursed.	Már gyerekkoromban gyötörték, álmaimat lidércnyomások sorozatává tették, s kissé később azt a meggyőződést ültették el bennem, hogy más vagyok, mint a fajtám: természetellenes, elátkozott lény.	14,15	0.9269
10	In my days only did I attain any measure of happiness.	Csak nappal érezhettem némi kis boldogságot.	16	0.5510
11	My nights marked the reign of fear - and such fear!	Éjszakáim a félelem uralmát jelentették - és micsoda félelemét!	17	1.0000
12	I make bold to state that no man of all the men who walk the earth with me ever suffer fear of like kind and degree.	Merem állítani, hogy senki emberfia, aki velem együtt tapossa a földet, nem érzett még ilyenfajta és ilyen fokú félelmet.	18	0.7099

Industrial application: Multilingual Text Analytics

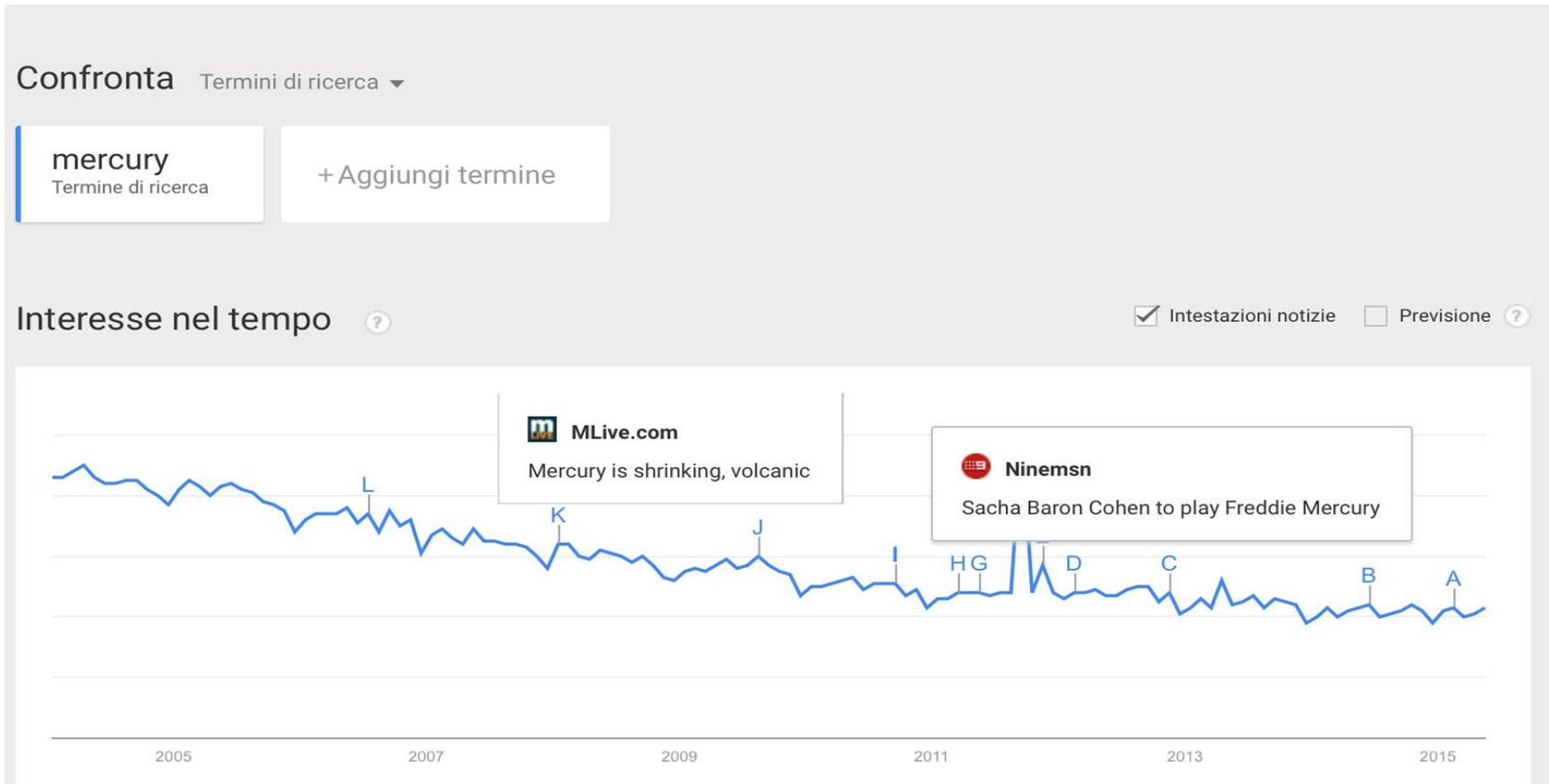
Who is interested: media content providers
and analysts, etc.

A close-up photograph of a computer keyboard. The central focus is a single white key with the word "Analytics" printed on it in a black, monospace-style font. The key is slightly raised and has rounded corners. Surrounding it are other keys, including one with a vertical line and another with a diagonal line, all in a similar white and grey color scheme. The lighting is soft, creating subtle shadows and highlights on the keys' surfaces.

Analytics

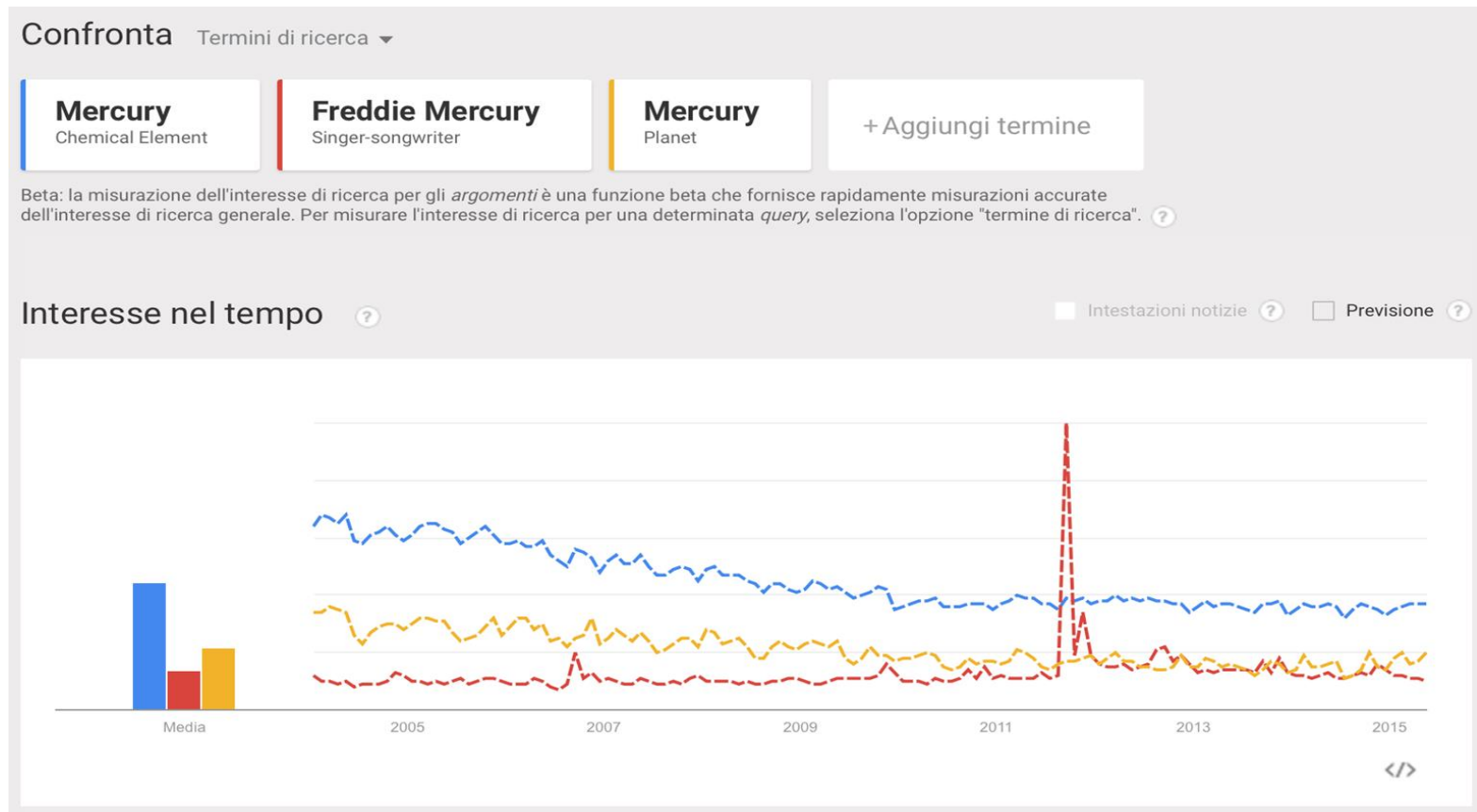
Trends are ambiguous

- The senses of the word “mercury” are **conflated** and plotted together.



Dream: multilingual semantic analytics

- **Goal:** semantic analytics of text **in any language** for both **named entities** and **concepts**



Multilingual News Analytics and Comparison



ANSA.it Tlc Fai la ricerca

Cronaca Politica Economia Regioni + Mondo Cultura Tecnologia

PRIMOPIANO • HI-TECH • INTERNET & SOCIAL • TELECOMUNICAZIONI • SOFTWARE & APP

ANSA.it • Tecnologia • Tlc • Usa: hacker viola computer aereo, arrestato

Usa: hacker viola computer aereo, arrestato

Cnn, esperto cybersicurezza prende controllo motore durante volo



International Edition London, United Kingdom 12° Sign in | MyCNN

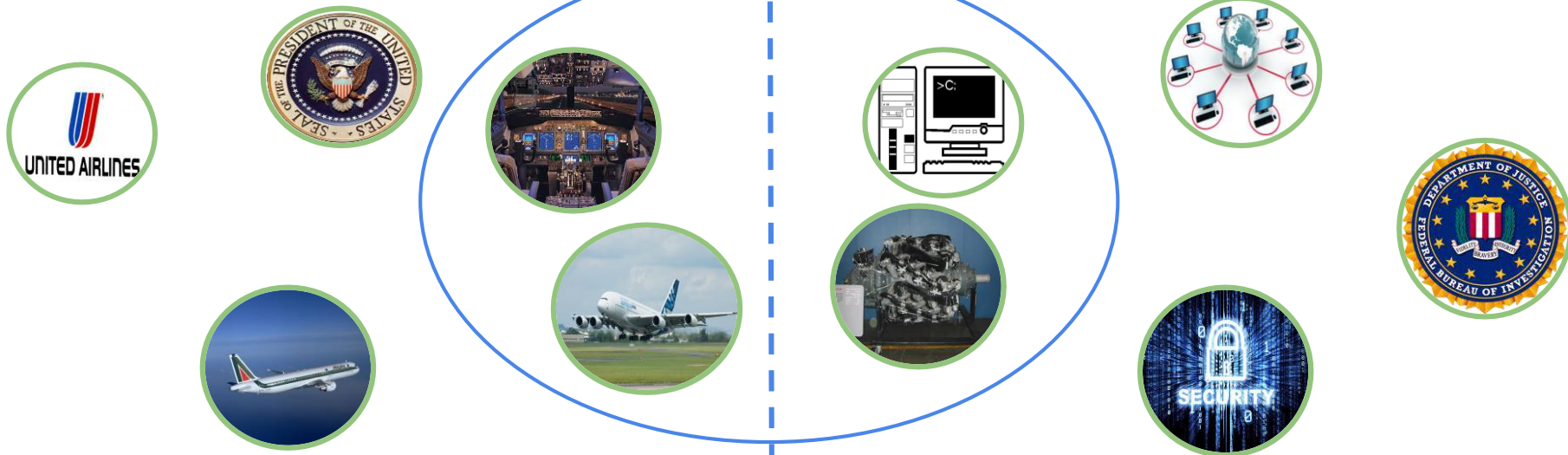
News Regions Video TV Features Opinions More... Search CNN

U.S. China Asia Middle East Africa Europe Americas

FBI: Hacker claimed to have taken over flight's engine controls

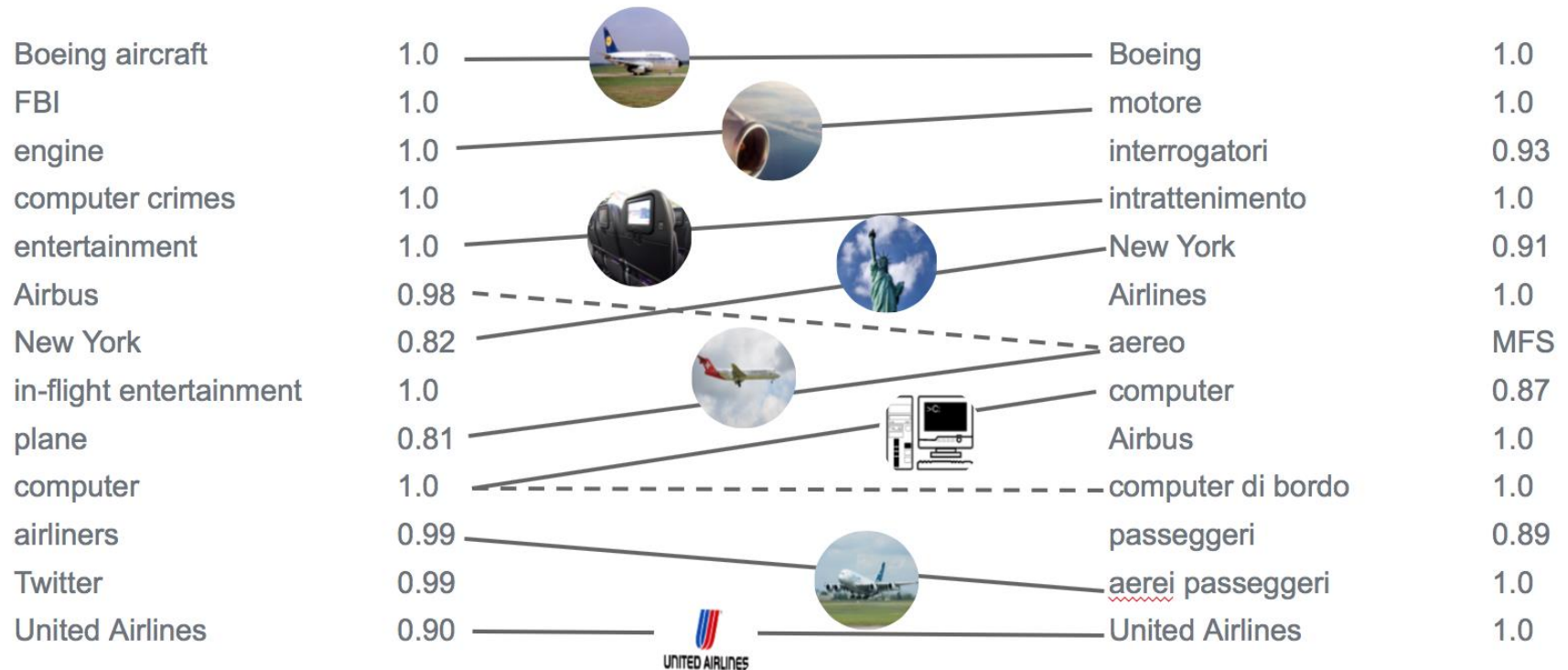
By Evan Perez, CNN

Updated 1305 GMT (2005 HKT) May 18, 2015



- Extract the most important **concepts** and **Named Entities** in virtually **any language** (271 covered currently)

Cont'd



- Important **semantic common ground**
- But also: **complementarity!**

NOW...

**WHAT WILL YOU DO WITH
BABELNET TONIGHT?**

Thanks or...





SAPIENZA
UNIVERSITÀ DI ROMA

Roberto Navigli

Linguistic Computing Laboratory

<http://lcl.uniroma1.it>

@RNavigli

