



Building the Multilingual Web of Data: A Hands-on tutorial

(ISWC 2014, Riva del Garda - Italy)

Multilingual Word Sense Disambiguation and Entity Linking on the Web based on BabelNet

Roberto Navigli, Tiziano Flati – Sapienza University of Rome

- **Roberto Navigli**, associate professor, Department of Computer Science, Sapienza University of Rome



- **Tiziano Flati**, PhD student, Department of Computer Science, Sapienza University of Rome



you will...

...receive a prize!!!

A BabelNet t-shirt!!!



[model is not included]

Part 1:

Identifying multilingual concepts and entities in text

- Web content is available in **many languages**
- Information should be **extracted and processed independently of the source/target language**
- This *could* be done automatically by means of high-performance **multilingual text understanding**



«**Thomas** and Mario are **strikers** playing in Munich»



Entity Linking: The task of **discovering mentions** of entities within a text and **linking them** in a knowledge base.

WSD: The task aimed at **assigning meanings to word occurrences** within text.

POLYSEMY

- Natural language is ambiguous
- The most frequent words have several meanings!
- **Our job:** model meaning from a computational perspective

- Monosemous words: **only one meaning**
 - Examples:
 - *plant life*
 - *internet*
- Polysemous words: **more than one meaning**
 - Example: *bar*
 - “a room or establishment where alcoholic drinks are served”
 - “a counter where you can obtain food or drink”
 - “a rigid piece of metal or wood”
 - “musical notation for a repeating pattern of musical beats”

- **Thesauri**
 - Groups words according to similar meaning
 - Relations between groups (e.g., narrower meanings)
 - Roget's Thesaurus (1911)
- **Machine Readable Dictionaries**
 - Enumerates all meanings of a word
 - Includes definitions, morphology, example usages, etc.
 - Oxford Dictionary of English, LDOCE, Collins, etc.
- **Computation Lexicons**
 - Repositories of structured knowledge about a word semantics and syntax
 - Include relations like hypernymy, meronymy, or entailment
 - WordNet

What if we choose BabelNet as our sense inventory?

BabelNet^{2.5}

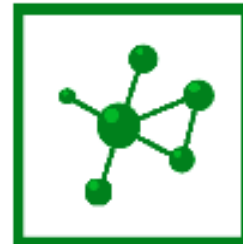
A very large multilingual encyclopedic dictionary and semantic network

BabelNet is both a **multilingual encyclopedic dictionary**, with lexicographic and encyclopedic coverage of terms in **50 languages**, and a **semantic network** which connects concepts and named entities in a **very large network of semantic relations**, made up of **more than 9 million entries**. [Read more...](#)

search



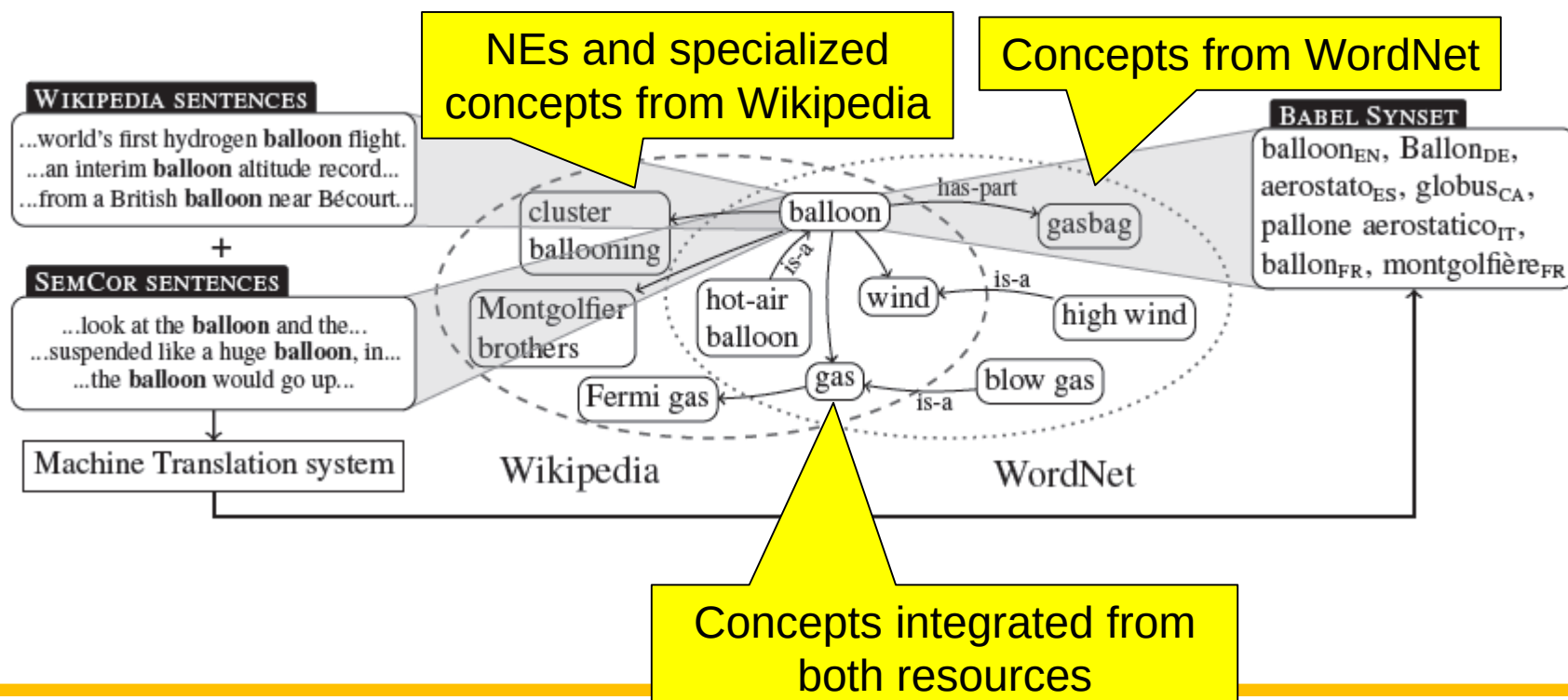
explore



publications

downloads

A wide-coverage **multilingual semantic network** including both encyclopedic (from Wikipedia) and lexicographic (from WordNet) entries



50 languages covered (including Latin!)

List of languages at <http://babelnet.org/stats>



50 languages covered (including Latin!)

9.3M Babel synsets (concepts and named entities)

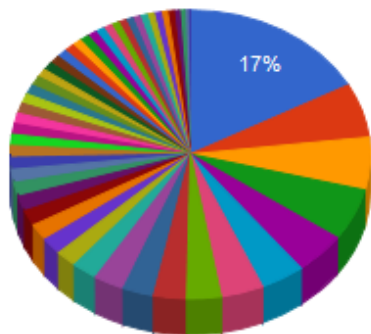
67M word senses

262M semantic relations (28 edges per synset on avg.)

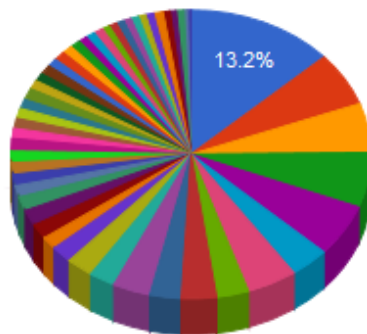
7.7M synset-associated images

21M textual definitions

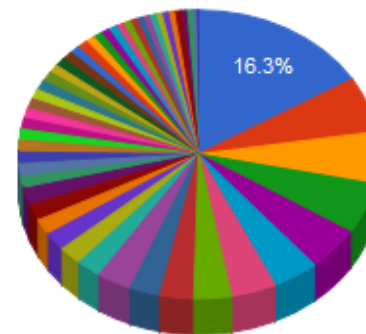
Lemmas by Language



Synsets by Language



Senses by Language



▲ 1/6 ▼

- Seamless integration of:
 - **WordNet 3.0**
 - **Wikipedia**
 - **Wikidata**
 - **Wiktionary**
 - **OmegaWiki**
 - **Open Multilingual WordNet** [Bond and Foster, 2013]
- Translations for all open-class parts of speech
- **1.1B** RDF triples available via SPARQL endpoint



Meaning: newspaper¹ • ID: bn:00057563n • Type: Concept

[\[details\]](#) [\[explore\]](#)

Senses:

-   newspaper¹, paper³
-   صحيفة, جريدة,  報紙,  journal,  εφημερίδα,  יומון,  giornale,  ペイパー, ペーパー, ペーブル, 新紙, 新聞, 新聞紙,  diario, periódico
-   Newspaper,  صحيفة,  報紙,  Presse écrite,  Lokalzeitung, Zeitung,  Εφημερίδα,  עיתון,  समाचारपत्र,  Giornale,  新聞,  Газета,  Prensa escrita
-   Newspapers, Illustrated newspaper, News paper, Microdaily, News sheet, Print news media, National Newspaper,  Broadsheet newspapers, Newspaper Reading, Sunday newspaper supplement, Local newspaper, News-sheet, Newspaper articles, Shimbun,  الصحيفة, الصحف, صحيفة, جريدة,  報刊, 報章, 报纸,  Presse écrite, Financement de la presse,  Lokalblatt, Lokalpresse, Wochenschrift, Zeitungsartikel,  Εφημερίδες,  עיתון יומי, יומון, עיתונות כתובה, עתון,  अखबार, समाचार पत्र,  Giornale (editoria), Giornali,  日刊新聞, 日刊紙, 新聞界, 新聞社,  Газеты, Еженедельник,  Periódicos, Semanario, Publicaciones periodicas, Prensa periodica, Diario (publicación), Periódico (publicación), Periodico (publicacion), Periódico (Publicación), Prensa diaria, Periódico, Prensa periódica, Publicaciones periódicas, Periódico (medio de comunicación), Periodismo escrito, Periodico (publicación), Periodicos, Publicacion periodica, Periódico escrito, Diario (publicacion), Periodico, Periódico (prensa), Publicacion periódica, Diario escrito, Diario (medio de comunicación)
-   gazette, newspaper,  journal,  Zeitung, Tageszeitung,  עתון,  giornale,  新聞,  газета,  diario, periódico



- 🇬🇧 gazette, newspaper, 🇫🇷 journal, 🇩🇪 Zeitung, Tageszeitung, 🇮🇱 יתון, 🇮🇹 giornale, 🇯🇵 新聞, 🇷🇺 газета, 🇪🇸 diario, periódico
- 🇬🇧 newspaper, 🇩🇪 صحيفة, 🇨🇳 報紙, 🇫🇷 journal, 🇩🇪 Zeitung, 🇬🇷 Εφημερίδα, 🇮🇱 יתון, 🇮🇹 समाचारपत्र, 🇮🇹 giornale, 🇯🇵 新聞, 🇷🇺 газета, 🇪🇸 prensa escrita
- 🇬🇧 صحيفة محلية, 🇩🇪 صحيفة, 🇨🇳 报纸, 🇫🇷 journal, journaux, journal local, 🇩🇪 zeitung, zeitung, lokalzeitung, 🇬🇷 εφημερίδα, εφημερίδες, τοπική εφημερίδα, 🇮🇱 יתון, יתון, יתון, 🇮🇹 समाचार पत्र, स्थानीय समाचार पत्र, 🇮🇹 giornale, giornali, giornale locale, 🇯🇵 新聞, 地元の新聞, 🇷🇺 газеты, местной газете, 🇪🇸 periódico, periódicos, periódico local
- 🇬🇧 صحيفة, 🇩🇪 ورقة, 🇨🇳 报纸, 🇫🇷 journal, 🇩🇪 papier, zeitung, 🇬🇷 χαρτί, εφημερίδα, 🇮🇱 יתון, 🇮🇹 कागज, 🇮🇹 giornale, 🇪🇸 periódico



Glosses:

-   A daily or weekly publication on folded sheets; contains news and articles and advertisements; "he read his newspaper at breakfast"
-   Scheduled publication containing news of events, articles, features, editorials, and advertising
-   A publication (usually published daily or weekly and printed on cheap, low-quality paper) that contains news and other articles.
-   A newspaper is a scheduled publication containing news of current events, informative articles, diverse features, editorials, and advertising.
-   الصحيفة أو الجريدة هي إصدار يحتوي على أخبار ومعلومات وإعلانات، وعادة ما تُطبع على ورق زهيد الثمن.
-   定期刊物
-   Périodique présentant des nouvelles et autres informations diverses
-   Publication (en général publiée chaque semaine ou chaque mois et imprimé sur du papier de basse qualité) qui contient des actualités et d'autres articles.
-   La presse écrite désigne, d'une manière générale, l'ensemble des moyens de diffusion de l'information écrite, ce qui englobe notamment les journaux quotidiens, les publications périodiques et les organismes professionnels liés à la diffusion de l'information.
-   Periodisch erscheinendes Druckerzeugnis mit aktuellem und universellem Inhalt
-   Eine Publikation (meist täglich oder wöchentlich erscheinend und auf kostengünstigem Papier von niedriger Qualität), die Nachrichten und andere Artikel enthält.
-   Als Lokalzeitung wird eine Tageszeitung oder Wochenzeitung bezeichnet, die sich im Wesentlichen auf das Geschäft der Nachrichten-, Sport- und Veranstaltungsberichterstattung sowie auf Kleinanzeigen einer Region beschränkt und in der Regel lokale Begebenheiten in satirisch-feuilletonistischer Form als Lokalglosse oder Spitze "aufspießt".
-   Das Wort Zeitung war ursprünglich der Begriff für eine beliebige Nachricht; die Bedeutung hat sich jedoch im Laufe des 18. Jahrhunderts geändert.

		Das Wort Zeitung war ursprünglich der Begriff für eine beliebige Nachricht; die Bedeutung hat sich jedoch im Laufe des 18 Jahrhunderts geändert.
		Ως εφημερίδα χαρακτηρίζεται οποιαδήποτε έντυπη περιοδική έκδοση της οποίας η περιεχόμενη ύλη αφορά κατά πλειονότητα ειδησεογραφία τρεχόντων γεγονότων της περιόδου στην οποία εκδίδεται.
		פרסום נייר יומי או שבועי, המכיל חדשות, מאמרים ופרסומות; "עיתון ערב"; "הוא נוהג לקרוא את העיתון בעת ארוחת הבוקר"
		עיתון הוא כתב-עת היוצא לאור, בדרך כלל, בתדירות קבועה ובדפוס, או באופן אלקטרוני.
		समाचार पत्र या अखबार, समाचारों पर आधारित एक प्रकाशन है।
		Pubblicazione periodica o anche aperiodica in forma cartacea
		Un giornale è una pubblicazione periodica o anche aperiodica in forma cartacea.
		Печатное периодическое издание
		Газёта — печатное периодическое издание, выходящее под постоянным названием и не реже одного раза в месяц.
		Publicaciones impresas que se diferencian en función de su periodicidad
		Publicación (realizada diaria o semanalmente e impresa sobre papel de baja calidad) que contiene noticias y otros artículos.
		El concepto prensa escrita se refiere a publicaciones impresas que se diferencian en función de su periodicidad, que puede ser diaria, semanal, mensual, o anual.
Categories:		
		Newspapers, Ephemera, Journalism, Media formats, Newspaper publishing, Paper products, Periodicals, Printing,
		صحف, طباعة, صحافة, كلمات من وضع مجمع اللغة العربية
		报纸, 新聞, 報紙發行,
		Presse écrite,
		Zeitung, Druckerzeugnis,
		עיתונים, עיתונות,
		समाचार पत्र, संचार के साधन,
		Giornali,
		Medios de comunicación, Medios escritos,
DBPedia:		
	Newspaper	



1. From **six** to **50** languages;
2. From **two** resources to **six**;
3. From **5** million to **9.3** million synsets;
4. From **50** million to **68** million word senses;
5. From **140** million semantic relations to **262** million semantic relations





BabelNet

Type a text or a term...

ENGLISH

BABEL

A very large multilingual encyclopedic dictionary and semantic network

 [PREFERENCES](#)



BabelNet

piano|

ENGLISH

BABEL

A very large multilingual encyclopedic dictionary and semantic network

PREFERENCES

Noun

- Noun
- Adverb
- Adjective



piano¹

Fortepiano designates the early version of the piano, from its invention by the Italian instrument maker Bartolomeo Cristofori around 1700 up to the early 19th century....

ID: 00035986n | Concept



piano²

In music, dynamics normally refers to the volume of a sound or note, but can also refer to every aspect of the execution of a given piece, either stylistic or functional....

ID: 00062202n | Concept

Piano (play)

Piano is a 1990 play by Trevor Griffiths, adapted from the 1977 film *An Unfinished Piece for Mechanical Piano*, itself based on the incomplete





My daughter I scratched fence skewers processed oblivion sci-fi electric-wave fx

Enable partial matches: ☐

Word sense disambiguation and entity linking together!

ENGLISH

BABELFY!

[expanded view](#) | [compact view](#)

my daughter i scratched fence skewers processed oblivion



daughter

a female human offspring; "her daughter cared for her in her old age"

scratched

cause friction; "my sweater scratches"



fence

a barrier that serves to enclose an area



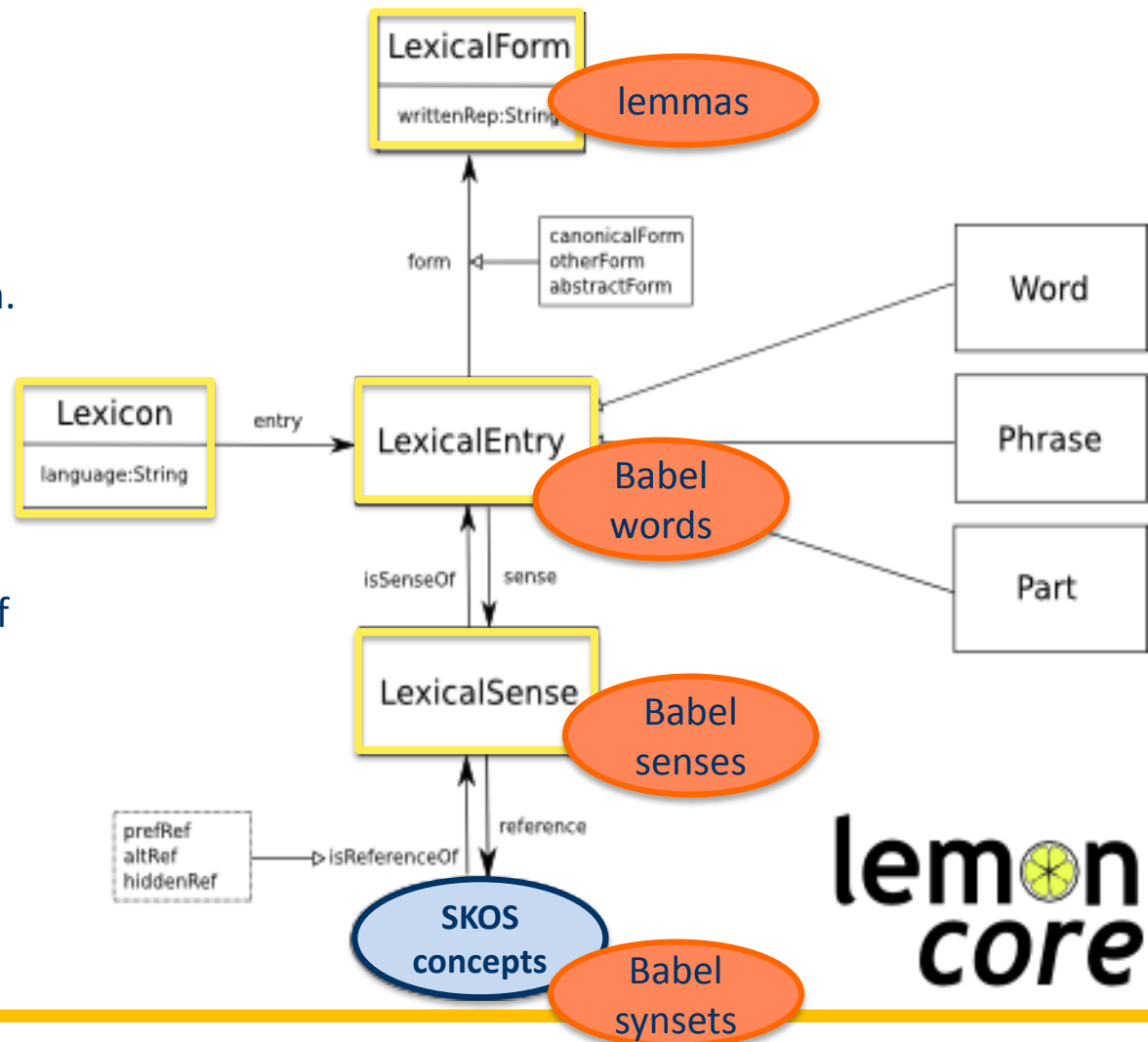
oblivion

Oblivion is a 2013 American science fiction comedy film

Part 2:

BabelNet-lemon: BabelNet as multilingual linked data!

- 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.



- Links towards **encyclopedic datasets**:

Wikipedia pages (36M):

`bn-lemon:wikipediaPage on lemon:LexicalSense`

Wikipedia categories (46M):

`bn-lemon:wikipediaCategory on skos:concept`

DBpedia pages (4M):

`skos:exactMatch on skos:concept`

DBpedia categories (15M):

`bn-lemon:dbpediaCategory on skos:concept`

- Links towards **lexical datasets**:

- lemon-Wordnet (117k): `skos:exactMatch on skos:concept`

- lemon-Uby, OmegaWiki (en) (15k): `skos:exactMatch on skos:concept`

- RDF dumps
 - format: n-triples
 - We produced dumps for both URIs and IRIs
- SPARQL endpoint
 - virtuoso universal server
 - <http://babelnet.org:8084/sparql/>
- Dereferencing
 - babelnet.org/2.0/
 - Pubby Linked Data Frontend
(<http://wifo5-03.informatik.uni-mannheim.de/pubby/>)

Hands-on Session: BabelNet





Go to:

<http://babelnet.org:8084/sparql/>

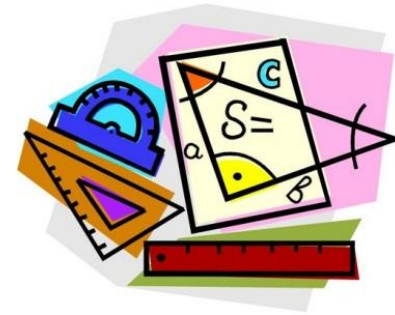
1. Open directory **babelnet-lemon**
2. Launch **launch_fuseki.sh** or **launch_fuseki.bat**
3. Open **<http://localhost:3030/>**
4. Click on **Control Panel**
5. Click on **Select**
6. Within the SPARQL query text area **put your SPARQL queries**
7. [List of possible words in file **freqs.top10K.words.txt**]



- PREFIX bn: `<http://babelnet.org/2.0/>`
- PREFIX bn-lemon: `<http://babelnet.org/model/babelnet#>`
- PREFIX lemon: `<http://www.lemon-model.net/lemon#>`
- PREFIX skos: `<http://www.w3.org/2004/02/skos/core#>`
- PREFIX lexinfo: `<http://www.lexinfo.net/ontology/2.0/lexinfo#>`
- PREFIX rdfs: `<http://www.w3.org/2000/01/rdf-schema#>`
- PREFIX dc: `<http://purl.org/dc/elements/1.1/>`
- PREFIX dcterms: `<http://purl.org/dc/terms/>`

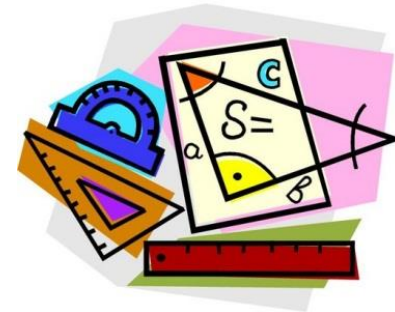
- Given a word, e.g. home, retrieve all its senses and corresponding synsets in all supported languages:

```
SELECT DISTINCT ?sense ?synset WHERE {
  ?entries a lemon:LexicalEntry .
  ?entries lemon:sense ?sense .
  ?sense lemon:reference ?synset .
  ?entries rdfs:label ?term .
  FILTER (str(?term)="home")
  } LIMIT 10
```



- We can restrict to a given language, e.g. English:

```
SELECT DISTINCT ?sense ?synset WHERE {
  ?entries a lemon:LexicalEntry .
  ?entries lemon:language "EN" .
  ?entries lemon:sense ?sense .
  ?sense lemon:reference ?synset .
  ?entries rdfs:label ?term .
  FILTER (str(?term)="home")
} LIMIT 10
```



- For instance, given the sense
http://babelnet.org/2.0/home_EN/s00044488n

```

SELECT ?translation WHERE {
    ?entry a lemon:LexicalSense .
    ?entry lexinfo:translation ?translation .
FILTER
    (str(?entry)="http://babelnet.org/2.0/home_EN/s00044488n")
}
    
```



- For instance, given the sense:

- http://babelnet.org/2.0/home_EN/s00044488n

```
SELECT ?license WHERE {
```

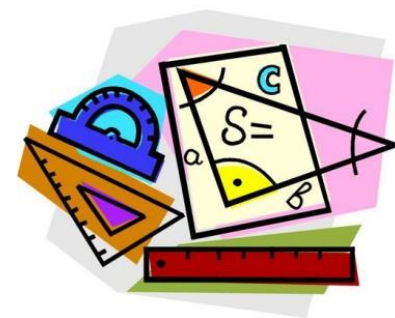
```
  ?entry a lemon:LexicalSense .
```

```
  ?entry dcterms:license ?license .
```

```
  FILTER
```

```
    (str(?entry)="http://babelnet.org/2.0/home_EN/s00044488n")
```

```
}
```



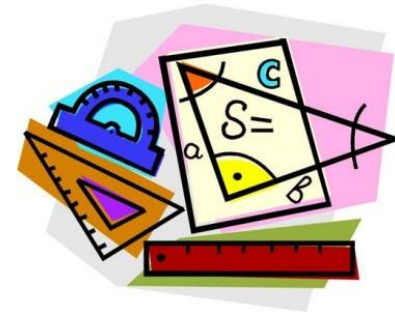
- For instance, given the synset <http://babelnet.org/2.0/s000000356n>

```
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 (str(?url)="http://babelnet.org/2.0/s000000356n")
}
```



- For instance, given the synset:
 - <http://babelnet.org/2.0/s00000356n>

```
SELECT ?broader WHERE {
    ?entry a skos:Concept .
    OPTIONAL { ?entry skos:broader ?broader }
    FILTER (str(?entry)="http://babelnet.org/2.0/s00000356n")
}
```



Part 2:

WSD and Entity Linking

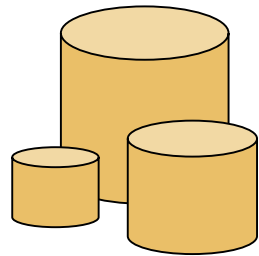
Together!

Back to the original problem

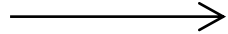


strikers
(target word)

“Thomas and Mario are **strikers** playing in Munich”
(context)



knowledge



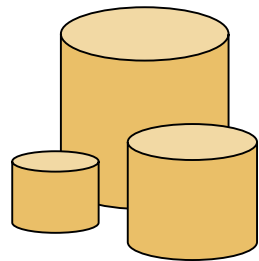
WSD
system



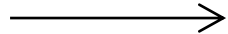
sense of target word

Thomas
(target mention)

“Thomas and Mario are strikers playing in Munich”
(context)



knowledge



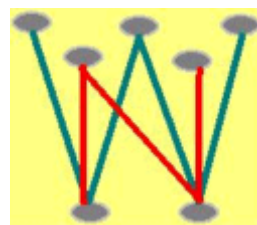
EL
system



Named Entity

- EL encompasses a set of **similar tasks**:
- **Named Entity Disambiguation**, that is the task of linking entity mentions in a text to a knowledge base
- **Wikification**, that is the automatic annotation of text by linking its relevant fragments of text to the appropriate Wikipedia articles.

- In both tasks, WSD and EL, **knowledge-based** approaches have been shown to perform well.
- What about **multilinguality**?
- Which kind of resources are available out there?



Open
Multilingual
WordNet



a multilingual tree
encyclopedia

Wiktionary
['wɪkʃənɹɪ] *n.*,
a wiki-based Open
Content dictionary
Wileo ['wɪl kærɪ]



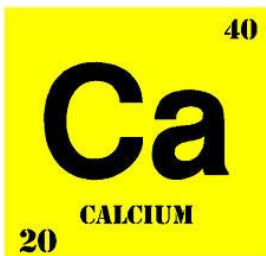
WordNet
A lexical database for English

- The main **difference** between WSD and EL is
the kind of inventory used



1. Concepts

- *Calcio* in Italian can denote different concepts:



2. Named Entities

- The text *Mario* can be used to represent different things such as the video game character or a soccer player (Gomez) or even a music album



Meaning: kick¹ • ID: bn:00012171n • Type: Concept [\[details\]](#) [\[explore\]](#)

Senses:

- kick¹, boot⁷, kicking²
- رِغْل, coup de pied, κλοτσιά, λάκτισμα, calcio, pedata, scalcia, キック, 一蹴り, 足蹴, 蹴り
- Kick, 脚法, Coup de pied, Tritttechnik, בעיטה, Calcio (pedata), 蹴り技, Удар ногой, Patada
- Kicks, Back Kick, Kicking, Push Kick, Hook Kick, 腳法, 踢 (武術), 足技, Coups de pied, Coup De Pied, かかと落とし, キック, 踵落とし, 蹴り, 蹴る, ネリチャギ, Patadas, Patear
- kick, boot, בעיטה, пинок
- kick, Coup de pied, Tritttechnik, בעיטה, calcio, 蹴り技, Удар ногой, Patada
- 踢, coup de pied, coups de pied, kick, treten, κλωτσιά, κλωτσιές, בעיטה, लात, kicks, लात मार, calcio, calci, キック, удар, patadas

Glosses:

- A hit or strike with the leg or foot or knee.
- A blow with the foot; a kick.
- Physical strike in combat sports
- The act of delivering a blow with the foot; "he gave the ball a powerful kick"; "the team's kicking was excellent"
- In combat sports and hand-to-hand combat, a kick is a physical strike using the foot, leg, or knee.
- Un coup de pied en sport de combat est une frappe réalisée avec une action du membre inférieur dont la surface d'impact est le pied, Dans certaines pratiques de combat, on parle de techniques de jambe; elles sont la partie émergente de leur activité, ce qui est le cas pour le taekwondo moderne.



Meaning: calcium¹ • ID: bn:00006850n • Type: Concept [\[details\]](#) [\[explore\]](#)

Senses:

- calcium¹, Ca¹, atomic number 20¹
- الكالسيوم, كالسيوم, العنصر العشرون, calcium, ασβέστιο, תי"ס, calcio, Ca, カルシウム, Ca
- Calcium, Calcium in biology, كالسيوم, 钙, 鈣營養, Calcium, Calcium, Ασβέστιο, תי"ס, कैल्शियम, Calcio (elemento), カルシウム, Кальций, Calcio
- Calcium storage, Ca(2+), Calcium (biology), Calcium storages, Factor 4, Kalcium, Calium, Calcium ions, Ca2+, Ca storages, Ca (biology), Calcium (second messenger), Ca+2, Ca ion (physiology), Calcium supplements, Calcium compounds, Calcium physiology, Biological role of calcium, Ca2, Ca²⁺, Calcium ion physiology, Element 20, Ca, الكالسيوم, Calcium, 钙, 20號元素, 元素20, 第20號元素, Calcium, 鈣質, Calcium (nutriment), 7440-70-2, Oxydation du calcium, Ca++, Ca2+, Kalzium, קלציום, קלציום, कैलशियम, काल्सियम, कैल्सियम, Ca, Calcio (elemento chimico), Calcio (metallo), Fattore di coagulazione IV, Calcio (chimica), lone calcio, Fattore coagulante IV, カルシウムイオン, カルシューム, Ca, Ca2+, Factor de coagulación IV, Calcios
- calcium, كالسيوم, calcium, Kalzium, Calcium, ασβέστιο, תי"ס, काल्सियम, calcio, カルシウム, кальций, calcio
- calcium, كالسيوم, calcium, Kalzium, Calcium, ασβέστιο, תי"ס, चूर्णातु, calcio, カルシウム, кальций, calcio
- calcium, Calcium in biology, كالسيوم, 钙, 鈣營養, calcium, Calcium, Ασβέστιο, תי"ס, कैल्शियम, calcio, カルシウム, кальций, Кальций в живых организмах, calcio



Meaning: soccer¹ • ID: bn:00006547n • Type: Concept [\[details\]](#) [\[explore\]](#)

Senses:

-   soccer¹, association football¹
-   soccer, football,  ποδόσφαιρο,  כדורגל,  calcio, pallone, football,  アソシエーションフットボール, サッカー, フットボール, フートボール, 蹴球,  fútbol
-   Association football,  لعبة كرة القدم,  足球,  Football,  Fußball,  Ποδόσφαιρο,  כדורגל,  फुटबॉल,  Calcio (sport),  サッカー,  Футбол,  Fútbol
-   Associated football, Outdoor Soccer, Soccer history, Voetbal, Association soccer, Association footballer, Football (soccer), Association football (soccer), Soccer Scoring, Soccer team, Recreational soccer, Fútbol, Sawker, 蹴球, Football (real), Fitba, Soccer football, Asocciation football, Football association, Football (association), Association football, Soccer (football), Association Football Club, Socker, Association Football, Ass. football, International club football, Football(soccer), Football (Soccer), World football, Soccer (Football), Association Football (soccer), Senior football, Football (Association), Futébol, Football (original), Soccor, Soccer, Assocciation football, Football (Association rules), The Method of Scoring, Balompie, サッカー, Balompié, Football (soccer), Futebol, Metric football, Association footballers, Rooball,  اتحاد كرة القدم,  Football, 英式足球, 踢足球, 足球运动,  Équipe de football, Équipe de foot, Club de foot, Joueur de football, Football européen, Soccer, Le football, Footballeur, Contrôle (football), Tête (football), Futeballe, Association football, Football au Portugal, Contrôle de poitrine, Football Club, Football masculin, Le foot, Foot-ball, Footballeurs, Balle au pied, Footeux, Football association,  Fußballspieler, Fussball, Fußballspielen, Fussballspieler, Fuathboi, Seiteaus, Fußballer, Männerfußball, Fussballer, Herrenfußball,  Ποδοσφαιριστής,





BabelNet is a huge **multilingual inventory**
for both word senses and named entities!



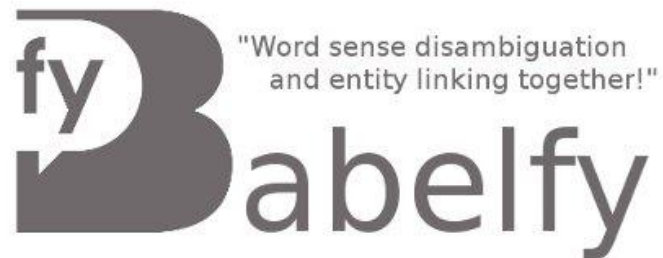
<http://babelfy.org>

Babelfy

- Based on **Personalized PageRank**, the state-of-the-art method for graph-based WSD.
 - However, it **cannot** be run for each new input on **huge graphs**.
- **Idea**: Precompute semantic signatures for the nodes!
- **Semantic signatures** are the most **relevant nodes** for a given node in the graph computed by using **random walk with restart**

Andrea Moro and Alessandro Raganato and Roberto Navigli. 2014. Entity Linking meets Word Sense Disambiguation: a Unified Approach. *Transactions of the Association for Computational Linguistics (TACL)*, 2.





Powered by BabelNet

Text to babelfy:

Enable partial matches: ☐ Language: English Babelfy!

When citing Babelfy, please refer to the following paper: Andrea Moro, Alessandro Raganato and Roberto Navigli. [Entity Linking meets Word Sense Disambiguation: A Unified Approach](#). Transactions of the Association for Computational Linguistics, 2, 2014.



Babelfy is an output of the [MultiJEDI](#) ERC Starting Grant No. 259234. Concept by [Andrea Moro](#) and [Roberto Navigli](#). **Babelfy** and its **API** are licensed under a [Creative Commons Attribution-Noncommercial-Share Alike 3.0 License](#).  For any commercial use, please contact us.





Powered by BabelNet

Text to babelify:

Conference 2014 is the premier international forum for the
scientists, industry specialists, and practitioners meet
scalable, user-friendly, and game changing solutions.

International Semantic Web

The International Semantic Web Conference is a series of academic conferences and the premier international forum, for the Semantic Web / Linked Data Community.

Semantic

the study of language

the Semantic Web

The Semantic Web is a collaborative movement led by the international standards body, the World Wide Web Consortium.



Linked Data

In computing, linked data describes a method of publishing structured data so that it can be used by other applications.

the future

the time yet to come

Community

The term community has two distinct commutative meanings: 1) Community usually refers to a social unit larger than a small village that shares common values.

international

concerning or belonging to all at least two or more nations;

meet

come together; "I'll probably see you at the meeting"; "How nice to see you again!"

Semantic

study of language meaning

network consisting of interlinking of Internet sites

tic Web

Community . Here , scientists , industry specialists , and practitioners meet to discuss the future of practical , scalable , user-friendly , and game changing solutions

practical

having or put to a practical purpose or use; "practical mathematics"; "practical applications of calculus"

solutions

the successful action of solving a problem; "the solution took three hours"

discuss

talk (something) with others; have a discussion; "we discussed our household"



Stars boycott Beverly Hills Hotel over Brunei sharia move



Demonstrators protest over Brunei's new harsh Islamic penal code at the Beverly Hills Hotel

Stars of Hollywood are boycotting one of the area's most famous hotels because of a harsh Islamic penal code introduced by Brunei's government.

Comedians Jay Leno and Ellen DeGeneres are among the celebrities supporting the boycott against the Beverly Hills Hotel and other hotels

Related Stories

[Brunei introduces tough Islamic law](#)

[Brunei profile](#)

fy Babelfy

"Word sense disambiguation and entity linking together!"

Stars
a widely known person; "he was a baseball celebrity"



boycott
refuse to sponsor; refuse to do business with

Ellen DeGeneres
Ellen Lee DeGeneres is an American stand-up comedian, television host and actress.



erly Hills Hotel over Brunei
c famous hotels because of a
and Ellen DeGeneres are am

sharia
the code of law derived from the Koran and from the teachings and example of Mohammed; "sharia is only applicable to Muslims"; "under Islamic law there is no separation of church and state"



Jay Leno
James Douglas Muir "Jay" Leno is an American stand-up comedian and television host.

hotels
a building where travelers can pay for lodging and meals and other services



brities
yly known person; "he baseball celebrity"

Beverly Hills Hotel over Brunei **shar**

one of the **area** 's

introduced by Bru

res are among the c

Beverly Hills Hotel ar

the Beverly Hills Hotel
The Beverly Hills Hotel, also called The Beverly Hills Hotel and Bungalows, is a hotel on Sunset Boulevard in Beverly Hills, California.



NEWS

6 May 2014 Last updated at 11:08 GMT

[Home](#)
[UK](#)
[Africa](#)
[Asia](#)
[Europe](#)
[Latin America](#)
[Mid-East](#)
[US & Canada](#)
[Business](#)
[Health](#)
[Sci/Environment](#)
[Tech](#)
[Entertainment](#)
[Video](#)


▶ [Thai crisis explained in 60 seconds](#)
 Profile: Yingluck Shinawatra
 Thailand to hold fresh election
 Q&A: Thailand protests



Hope for boost on Ukraine elections

European foreign ministers seek to build support for this month's presidential election in Ukraine, where troops are battling pro-Russian separatists.

LIVE [Russia's Lavrov holds news briefing](#)

▶ [Ukraine soldiers killed in Sloviansk](#)

[Odessa images](#)

[How did Odessa's fire happen?](#)



Coca-Cola to remove drinks ingredient

Coca-Cola is to remove an ingredient which has been linked to negative health effects from some of its drinks brands by the end of this year.

[Can big food firms get us to eat more healthily?](#)

[Campaigners vow to cut sugar in food](#)

Features



Migrants

Trapped in the
workers, dre



Toppling

Does China's
down the for



Living out

Young Europ
hand-to-mou



Too old?

The 93-year-
India's electi



Haunted

Why South A
veterans say



31-0

The story be
football team



Muscular



fy Babelfy

"Word sense disambiguation
and entity linking together!"

Powered by BabelNet

Text to babelfy:

The world's largest beverage-maker, Coca-Cola, plans to remove some of its drinks brands by the end of this year, following

Brominated vegetable oil, or BVO, is found in Coca-Cola fruit and Powerade.

An element of the additive, also found in flame retardants,

drinks

any liquid suitable for drinking;
"may I take your beverage
order?"



Enable partial

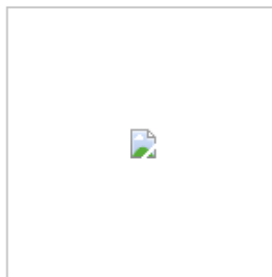
Fanta

Fanta is a global brand of fruit-flavored carbonated soft drinks from the Coca-Cola Company.

The world's largest

flame retardants

Flame retardants are chemicals used in thermoplastics, thermosets, textiles and coatings that inhibit or resist the spread of fire.



is by the end of this
table oil, or BVO, is
Fanta and Powerade.

Next Goal Wins for 'world's worst football team'

By Kev Geoghegan

Arts and entertainment reporter, BBC News



Goalkeeper Nicky Salapu is the only remaining player from the 31-0 hammering by Australia

In 2001, Australia set a world record for the largest ever victory in an international football match, netting 31 goals without a single breach of their own defence.

Related Stories

fy Babelfy

"Word sense disambiguation
and entity linking together!"

Powered by BabelNet

Text to babelfy:

In 2001, Australia set a world record for the largest ever victory in an international football match, netting 31 goals without a single breach of their own defence.

The game's top scorer, Archie
international match by claim

While Thompson is now the top

Enable

In 2001 , Australia
international football
their own defence

football

any of various games played
with a ball (round or oval) in
which two teams try to kick or
carry or propel the ball into
each other's goal



Archie Thompson

Archibald Gerald Thompson is a
New Zealand-born Australian
professional football player,
currently playing for Melbourne
Victory in the Hyundai
A-League, where he is 2nd in
the League all time scoring
record.

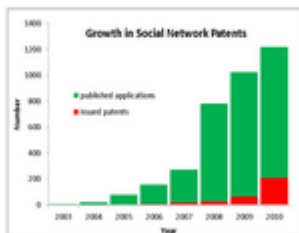


20/10/2014



social media

Social media employ web - and mobile-based technologies to support interactive dialogue and "introduce substantial and pervasive changes to communication between organizations, communities, and individuals."



Amazon si allea con Twitter : i consumatori

social media

servizio è

Amazon.com

Amazon.com, Inc. is an American multinational electronic commerce company with headquarters in Seattle, Washington, United States.

amazon.com



"Word sense disambiguation and entity linking together!"

Babelfy

Powered by BabelNet

Text to babelfy:

consumatori potranno comprare un prodotto con un
tezzata #AmazonCart negli Stati Uniti e #AmazonBas
il servizio è disponibile, il sistema unisce gli
nzione? Quando un cliente legge un tweet del suo
di un amico con un link con un prodotto su Amazon
sta un tweet seguito dall'hashtag #AmazonCart o #A
o su Amazon

consumatori

a person who uses goods or services



Twitter

Twitter is an online social networking service and microblogging service that enables its users to send and read text-based messages of up to 140 characters, known as "tweets".



n cinguettio senza lasciare il
, i due Paesi dove finora il
. Come funziona ? Quando un

Key fact!

- Annotating with BabelNet implies annotating with **WordNet and Wikipedia**
- (now **also** OmegaWiki, Open Multilingual WordNet, Wiktionary and WikiData!)



- **All current approaches exploit:**

- POS tagging
- Lemmatization

Noisy (>90% for English, but much less on morphologically rich languages).

- **How to improve?**

- Waiting for better POS taggers
- Character-based analysis of text

- **All current approaches exploit:**

- Knowledge of the input language
- Automatic language recognition

Noisy (>90% for English, but much less on resource poor languages). Moreover, text which consists of text in multiple languages will be wrongly analyzed for sure!

- **How to improve?**

- Waiting for better language recognition systems
- Unify the lexicalizations of different languages

- **Most of the current approaches exploit:**

- Named Entity Recognition
- Not overlapping text assumption

- **How to improve?**

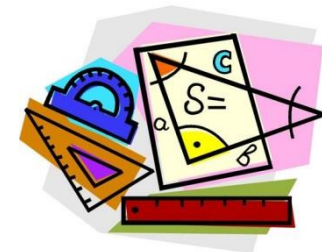
- Waiting for better NER system
- Overlap and match everything

Noisy (>80% for English, but much less on resource poor languages). Moreover, when assuming that entities and word senses should not overlap you lose information!

Hands-on Session: Babelfy

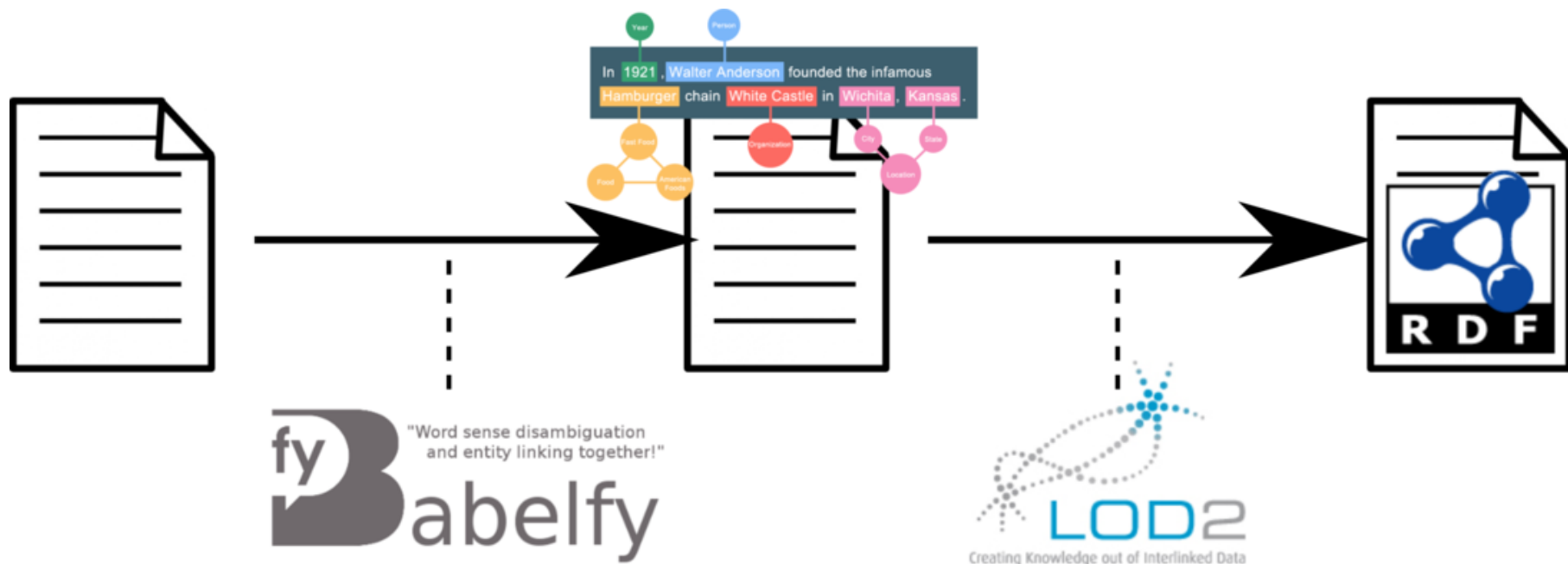


- Go to babelfy.org
- **Type** in or copy/paste your favourite text in your favourite language in the text area
- **Select** the text language
- **Click** on «Babelfy!»
- **Understand** the difference between green and yellow balloons



Part 3:

Producing multilingual linked data



Free text



Text annotated
with BabelFy



Annotated text
in RDF

From the previous
talk you should
already know
everything about
NIF, don't you?



But just in case...

- The NLP Interchange Format (NIF) is an RDF/OWL-based format that aims to achieve **interoperability** between Natural Language Processing (NLP) tools, language resources and annotations. NIF consists of specifications, ontologies and software

- **Reuse** of existing standards (such as RDF, OWL 2, the PROV Ontology, etc.)
- NIF identifiers are used in the **Internationalization Tag Set (ITS) Version 2.0**
- Royalty-free and published under an **open license**.
- Driven by its open community project **NLP2RDF**
- **good uptake** by industry, open-source projects and developers.

Wiki-Link Corpus in RDF

[corpus URI](#)

NIF Conversion Demo Service (★ important)

[demo URI](#)

NIF Combinator Demo

[demo URI](#)

NIF 2.0 Core Ontology - Terminological Model (★ important)

[ontology URI](#)

NIF Validator TestCase Suite

[ontology URI](#)

NIF 2.0 Core Ontology - Validation Model

[ontology URI](#)

NIF 2.0 Core Ontology - Inference Model

[ontology URI](#)

RLOG - an RDF Logging Ontology

[ontology URI](#)

STC - Simple Test Case Ontology

[ontology URI](#)

Promotional poster for Meta Forum (★ important)

[poster URI](#)

NIF 2.0 Validator Tool (★ important)

[software URI](#)

NIF 2.0 Stanford Core Implementation (★ important)

[software URI](#)

NIF 2.0 Stanbol Profile Specification

[specification URI](#)

Public API Specification (★ important)

[specification URI](#)

```
public static void run() throws IOException, URISyntaxException, BabelifyKeyNotValidOrLimitReached
{
    // Set the language
    Language lang = Babelify2NIFPropertyHandler.getInstance().getLanguage();

    // Set the text
    String inputText = Babelify2NIFPropertyHandler.getInstance().getText();

    // Select the overlapping annotations algorithm
    OverlappingAnnotationsAlgorithm algorithm = Babelify2NIFPropertyHandler.getInstance().getAlgorithm();

    // Get the key for Babelify
    String key = Babelify2NIFPropertyHandler.getInstance().getKey();

    // Set the RDF format
    String format = Babelify2NIFPropertyHandler.getInstance().getFormat();
    OutputStream outputstream = Babelify2NIFPropertyHandler.getInstance().getOutputStream();

    // Get the annotated output
    BabelifyWrapper bfyWrapper = BabelifyWrapper.getInstance(key);
    List<BabelifySentence> annotatedText = bfyWrapper.getAnnotatedText(inputText, lang, algorithm);

    // Convert Babelify's output into RDF-NIF
    Babelify2NIF babelify2nif = Babelify2NIF.getInstance();
    OntModel ontModel = babelify2nif.convert(annotatedText);

    // Print out the RDF file
    ontModel.write(outputstream, Format.toJena(format));
}
```

Set the language

Set the text

Obtain the annotations

Convert into RDF

Toy sentence:

“the semantic web is a
collaborative movement led by
the international standards body
world wide web consortium”

“the semantic web is a collaborative movement led by the international standards body world wide web consortium”

```
<http://lcl.uniroma1.it/babelfy2nif#char=0,488>
```

```
a      nif:Context ,
nif:beginIndex "0" ;
nif:endIndex   "488" ;
```

nif:Context defines the overall text

```
nif:isString "the semantic web is a collaborative movement led
by the int... -lrb- w3c -
rrb- .
encouraging the inclusion...
web aims at converting the current web
semi-structured documents into a `` web
builds on the w3c 's resource description framework
```

Text is modelled by fragments

Fragments are identified by left and right indices

```
<http://lcl.uniroma1.it/babelfy2nif#char=0,16>
```

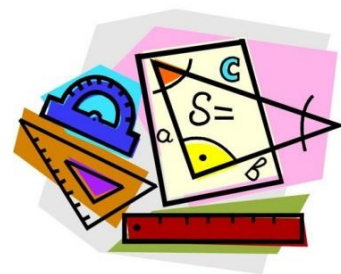
```
a      nif:Word , nif:RFC5147String ;
nif:anchorOf "the
nif:beginIndex
nif:endIndex
nif:oliaCategory olia:Noun , olia:CommonNoun ;
nif:oliaLink      <http://purl.org/olia/penn.owl#NN> ;
nif:referenceContext <http://lcl.uniroma1.it/babelfy2nif#char=0,488> ;
itsrdf:taIdentRef   bn:202276858n .
```

The BabelNet synset
(i.e., the annotation of the fragment)

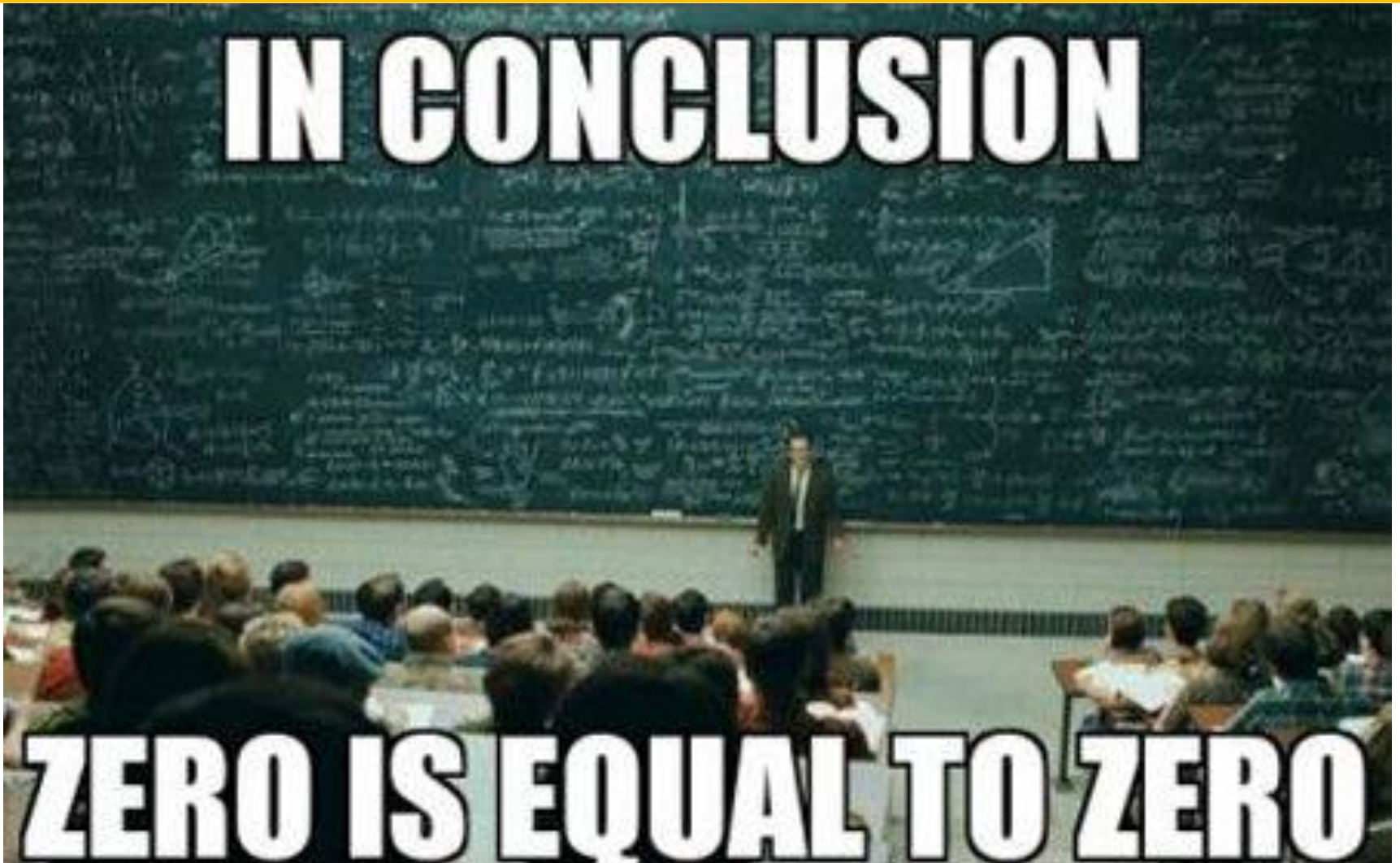
Hands-on Session: Babelfy 2 NIF



1. Open directory **babelfy**
2. Open file ***babelfy2nif.properties*** under ***config/*** directory
 - Select your language (*language* variable)
 - Type in your favourite text (*text* variable)
 - Select the algorithm for handling overlapping annotations (*algorithm* variable)
 - Select the output format (*rdf_format* variable)
 - Select the output stream (*output_stream* variable)
3. **Launch** “sh run_babelfy2nif-demo.sh”

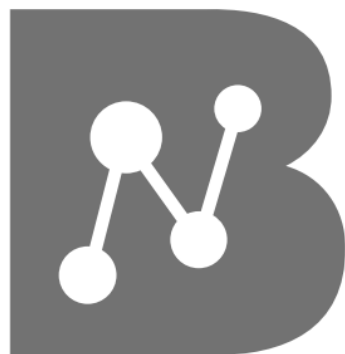


- Take the first paragraph of the English **ISWC Wikipedia page**
http://en.wikipedia.org/wiki/International_Semantic_Web_Conference
- Feed it in the property file and produce 2 outputs:
 1. With LONGEST_ANNOTATION_GREEDY_ALGORITHM algorithm and NTRIPLE rdf format
 2. With FIRST_COME_FIRST_SERVED_ALGORITHM algorithm and TURTLE rdf format



We have taken you through a tour of:

- A very large multilingual semantic network: **BabelNet**
- A state-of-the-art WSD and EL system: **Babelfy**



BabelNet



- European Research Council and the EU Commission for funding our research
- **Maud Ehrmann and Andrea Moro** for their help with slides



NOW...

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



SAPIENZA
UNIVERSITÀ DI ROMA

Linguistic Computing Laboratory

<http://lcl.uniroma1.it>

<http://babelnet.org>

<http://babelfy.org>

Google group: babelnet-group

