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
The instructors

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

- **Tiziano Flati**, PhD student, Department of Computer Science, Sapienza University of Rome
And, if you resist until the end...

you will...

...receive a prize!!!
A BabelNet t-shirt!!!

[model is not included]
Part 1:
Identifying multilingual concepts and entities in text
The driving force

- 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”
But how do we represent and encode semantics?

- **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 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...
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
Anatomy of BabelNet 2.5

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
New 2.5 version out!

- 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
<table>
<thead>
<tr>
<th>Meaning: newspaper</th>
<th>ID: bn:00057563n</th>
<th>Type: Concept</th>
</tr>
</thead>
</table>

**Senses:**

<table>
<thead>
<tr>
<th>English</th>
<th>Chinese</th>
<th>Italian</th>
<th>Japanese</th>
<th>Portuguese</th>
</tr>
</thead>
<tbody>
<tr>
<td>newspaper</td>
<td>报纸</td>
<td>giornale</td>
<td>新聞</td>
<td>diario</td>
</tr>
<tr>
<td>paper</td>
<td>新聞</td>
<td>新聞</td>
<td>新聞</td>
<td>diario</td>
</tr>
<tr>
<td>Newspaper</td>
<td>新聞</td>
<td>新聞</td>
<td>新聞</td>
<td>diario</td>
</tr>
<tr>
<td>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, 報刊</td>
<td>新聞</td>
<td>新聞</td>
<td>新聞</td>
<td>diario</td>
</tr>
<tr>
<td>Financement de la presse, Lokalblatt, Lokalpresse, Wochenschrift, Zeitungsartikel</td>
<td>新聞</td>
<td>新聞</td>
<td>新聞</td>
<td>diario</td>
</tr>
<tr>
<td>每日新聞, 每日新聞, 每日新聞</td>
<td>新聞</td>
<td>新聞</td>
<td>新聞</td>
<td>diario</td>
</tr>
<tr>
<td>gazette, newspaper</td>
<td>新聞</td>
<td>新聞</td>
<td>新聞</td>
<td>diario</td>
</tr>
<tr>
<td>newspaper</td>
<td>报纸</td>
<td>giornale</td>
<td>新聞</td>
<td>diario</td>
</tr>
</tbody>
</table>

WordNet+OpenMultilingualWordNet+ Wikipedia+...
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 تطبع 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 καρακτηρίζεται αποτελώντας έντυπη περιοδική έκδοση της οποίας η περιέχομενη ύλη αφορά κατά πλειονότητα ειδησεογραφία τρεχόντων γεγονότων της περιόδου στην οποία εκδίδεται.

فرماسير ينوي أن يعلم الناس، المجلات، المجلات، الفخافخ، "الخطيب"، "الخطيب الكبير". "الخطيب الكبير". "الخطيب الكبير". "الخطيب الكبير".

Pubblicazione periodica o anche aperiodica in forma cartacea

Un giornale è una pubblicazione periodica o anche aperiodica in forma cartacea.

Печатное периодическое издание

Газета — 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
- Zeitung
- Druckerzeugnis
- Druckerei
- 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 3.0 available from November 2014!!!
piano

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

piano

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 Pianos itself based on the incomplete auto-biography of the composer and pianist Egon Wellesz. The play tells the story of Wellesz's life and career, from his early days as a pianist in Austria to his later years in England.
New Babelfy interface available from November 2014!!!
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):
  
  \[ \text{bn-lemon:} \text{wikipediaPage on lemon:LexicalSense} \]

  Wikipedia categories (46M):
  
  \[ \text{bn-lemon:} \text{wikipediaCategory on skos:concept} \]

  DBpedia pages (4M):
  
  \[ \text{skos:exactMatch on skos:concept} \]

  DBpedia categories (15M):
  
  \[ \text{bn-lemon:} \text{dbpediaCategory on skos:concept} \]

• **Links towards lexical datasets:**

  – lemon-Wordnet (117k): \[ \text{skos:exactMatch on skos:concept} \]

  – lemon-Uby, OmegaWiki (en) (15k): \[ \text{skos:exactMatch on skos:concept} \]
• RDF dumps
  – format: n-triples
  – We produced dumps for both URIs and IRIs

• SPARQL endpoint
  – virtuoso universal server

• 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`
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`]
Some useful prefixes...

- 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/>
Exercise 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:

```sql
SELECT DISTINCT ?sense ?synset WHERE {
?entries a lemon:LexicalEntry .
?entries lemon:sense ?sense .
FILTER (str(?term)="home")
} LIMIT 10
```
Exercise 2: Retrieve the senses of a lemma for a certain language

• 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 .
FILTER (str(?term)="home")
} LIMIT 10
```
Exercise 3: Retrieve the translations of a given sense

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

```sparql
SELECT ?translation WHERE {
  ?entry a lemon:LexicalSense .
  FILTER (str(?entry)="http://babelnet.org/2.0/home_EN/s00044488n")
}
```
Exercise 4: Retrieve license information about a sense

• 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")
}
Exercise 5: Retrieve textual definitions in all languages

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

```sparql
?url a skos:Concept .
OPTIONAL {
}
FILTER (str(?url)="http://babelnet.org/2.0/s00000356n")
}
```
Exercise 6: Retrieve a synset’s hypernyms

• 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
Word Sense Disambiguation in a Nutshell

**strikers** (target word)

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

WSD system

knowledge

sense of target word
Entity Linking in a Nutshell

Thomas (target mention)

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

EL system

knowledge

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**
The main difference between WSD and EL is the kind of inventory used.
But BabelNet can be used as a multilingual inventory for both:

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
Calcio/Kick in BabelNet 2.5

Meaning: kick¹

ID: bn:00012171n
Type: Concept

Senses:

kick¹, boot⁷, kicking²

Kick, coup de pied, calcio, pedata, scalciata, Patada, Patadas, Patear

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.
Calcio/Soccer in BabelNet 2.5

Meaning: soccer

Senses:
- soccer, association football
- Association football, calcio, pallone, football
BabelNet is a huge multilingual inventory for both word senses and named entities!
So what?

http://babelfy.org
Babelfy: A Joint approach to WSD and EL [Moro et al., TACL 2014]

• 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**


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.
Babelfying ISWC!

"Word sense disambiguation and entity linking together!"

Babelfy

Linked Data
The study of language and meaning.

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.

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

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!"

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

Roberto Navigli e Tiziano Flati – La Sapienza University of Rome
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
Stars  
a widely known person; "he was a baseball celebrity"

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

Beverly Hills Hotel  
over Brunei's famous hotels because of a boycott and Ellen DeGeneres are among the

boycott  
refuse to sponsor; refuse to do business with

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

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.

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"
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
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, is suitable for drinking.

**Fanta**
Fanta is a global brand of fruit-flavored carbonated soft drinks from the Coca-Cola Company.
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.
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 Thompson, international match by claiming the title.

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.

**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.
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."

Babelfy
"Word sense disambiguation and entity linking together!"

Text to babelfy:

consumatori
a person who uses goods or services

Amazon
Amazon.com, Inc. is an American multinational electronic commerce company with headquarters in Seattle, Washington, United States.
Annotating with BabelNet: all in one!

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

• How to improve?
  • Waiting for better POS taggers
  • Character-based analysis of text

Noisy (>90% for English, but much less on morphologically rich languages).
• **All current approaches exploit:**
  - Knowledge of the input language
  - Automatic language recognition

• **How to improve?**
  - Waiting for better language recognition systems
  - Unify the lexicalizations of different languages

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!
• **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
Exercise 1

• 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
Babelfy 2 NIF (hackathon at SEMANTiCS 2014)

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.
<table>
<thead>
<tr>
<th>Wiki-Link Corpus in RDF</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NIF Conversion Demo Service ( ★ important)</td>
<td>demo URI</td>
</tr>
<tr>
<td>NIF Combinator Demo</td>
<td>demo URI</td>
</tr>
<tr>
<td>NIF 2.0 Core Ontology - Terminological Model ( ★ important)</td>
<td>ontology URI</td>
</tr>
<tr>
<td>NIF Validator TestCase Suite</td>
<td>ontology URI</td>
</tr>
<tr>
<td>NIF 2.0 Core Ontology - Validation Model</td>
<td>ontology URI</td>
</tr>
<tr>
<td>NIF 2.0 Core Ontology - Inference Model</td>
<td>ontology URI</td>
</tr>
<tr>
<td>RLOG - an RDF Logging Ontology</td>
<td>ontology URI</td>
</tr>
<tr>
<td>STC - Simple Test Case Ontology</td>
<td>ontology URI</td>
</tr>
<tr>
<td>Promotional poster for Meta Forum ( ★ important)</td>
<td>poster URI</td>
</tr>
<tr>
<td>NIF 2.0 Validator Tool ( ★ important)</td>
<td>software URI</td>
</tr>
<tr>
<td>NIF 2.0 Stanford Core Implementation ( ★ important)</td>
<td>software URI</td>
</tr>
<tr>
<td>NIF 2.0 Stanbol Profile Specification</td>
<td>specification URI</td>
</tr>
<tr>
<td>Public API Specification ( ★ important)</td>
<td>specification URI</td>
</tr>
</tbody>
</table>
```java
public static void run() throws IOException, URISyntaxException, BabelfyKeyNotValidOrLimitReached {

    // Set the language
    Language lang = Babelfy2NIFPropertyHandler.getInstance().getLanguage();

    // Set the text
    String inputText = Babelfy2NIFPropertyHandler.getInstance().getText();

    // Select the overlapping annotations algorithm
    OverlappingAnnotationsAlgorithm algorithm = Babelfy2NIFPropertyHandler.getInstance().getAlgorithm();

    // Get the key for Babelfy
    String key = Babelfy2NIFPropertyHandler.getInstance().getKey();

    // Set the RDF format
    String format = Babelfy2NIFPropertyHandler.getInstance().getFormat();
    OutputStream outputStream = Babelfy2NIFPropertyHandler.getInstance().getOutputStream();

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

    // Convert Babelfy's output into RDF-NIF
    Babelfy2NIF babelfy2nif = Babelfy2NIF.getInstance();
    OntModel ontModel = babelfy2nif.convert(annotatedText);

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

Text is modelled by fragments

Fragments are identified by left and right indices

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

1. Open directory babelfy

2. Open file \textit{babelfy2nif.properties} under \texttt{config/} directory
   - Select your language (\texttt{language} variable)
   - Type in your favourite text (\texttt{text} variable)
   - Select the algorithm for handling overlapping annotations (\texttt{algorithm} variable)
   - Select the output format (\texttt{rdf\_format} variable)
   - Select the output stream (\texttt{output\_stream} variable)

3. Launch “\texttt{sh run\_babelfy2nif-demo.sh}”
Exercise 1


• 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
In conclusion

Zero is equal to zero
To summarize

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**
Acknowledgements

• 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?