

Linking Thesauri and Glossaries

Case Study 0: linking a fake resource

Roberto Navigli

DIPARTIMENTO
DI INFORMATICA



SAPIENZA
UNIVERSITÀ DI ROMA

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

The Luxembourg BabelNet Workshop – Session 6

Session 6 – The Luxembourg BabelNet Workshop

[11:00-12:15, 3 March, 2016]

- Linking Glossaries and Thesauri to BabelNet and Babelfy
- Case study 0: linking a minimal fake resource

Ambiguity: a recurring issue

- Words in natural languages are ambiguous, though.
- In what meanings words are used in your thesaurus or glossary depends on its generality, its domain specificity, etc.

“The quick brown **fox** jumps over the lazy dog”



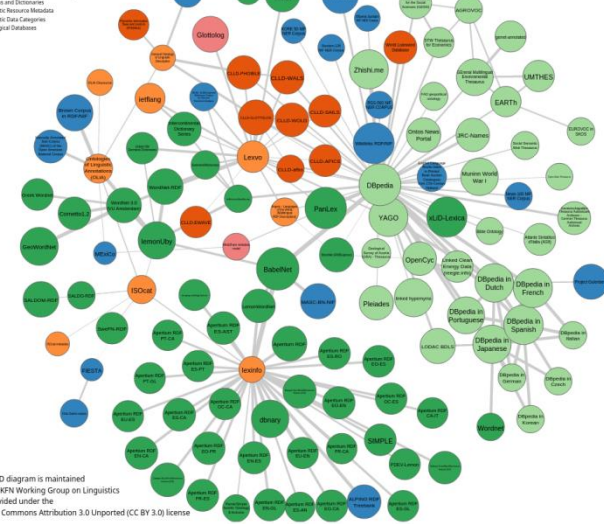
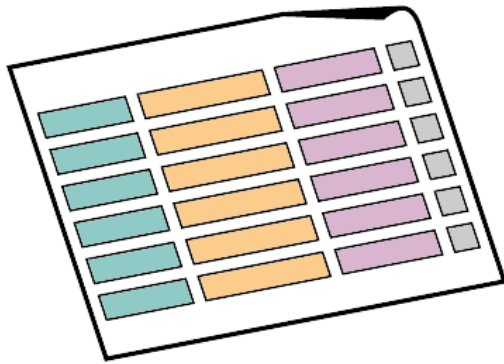
Ambiguity: a recurring issue

- Words in natural languages are ambiguous, though.
- In what meanings words are used in your thesaurus or glossary depends on its generality, its domain specificity, etc.

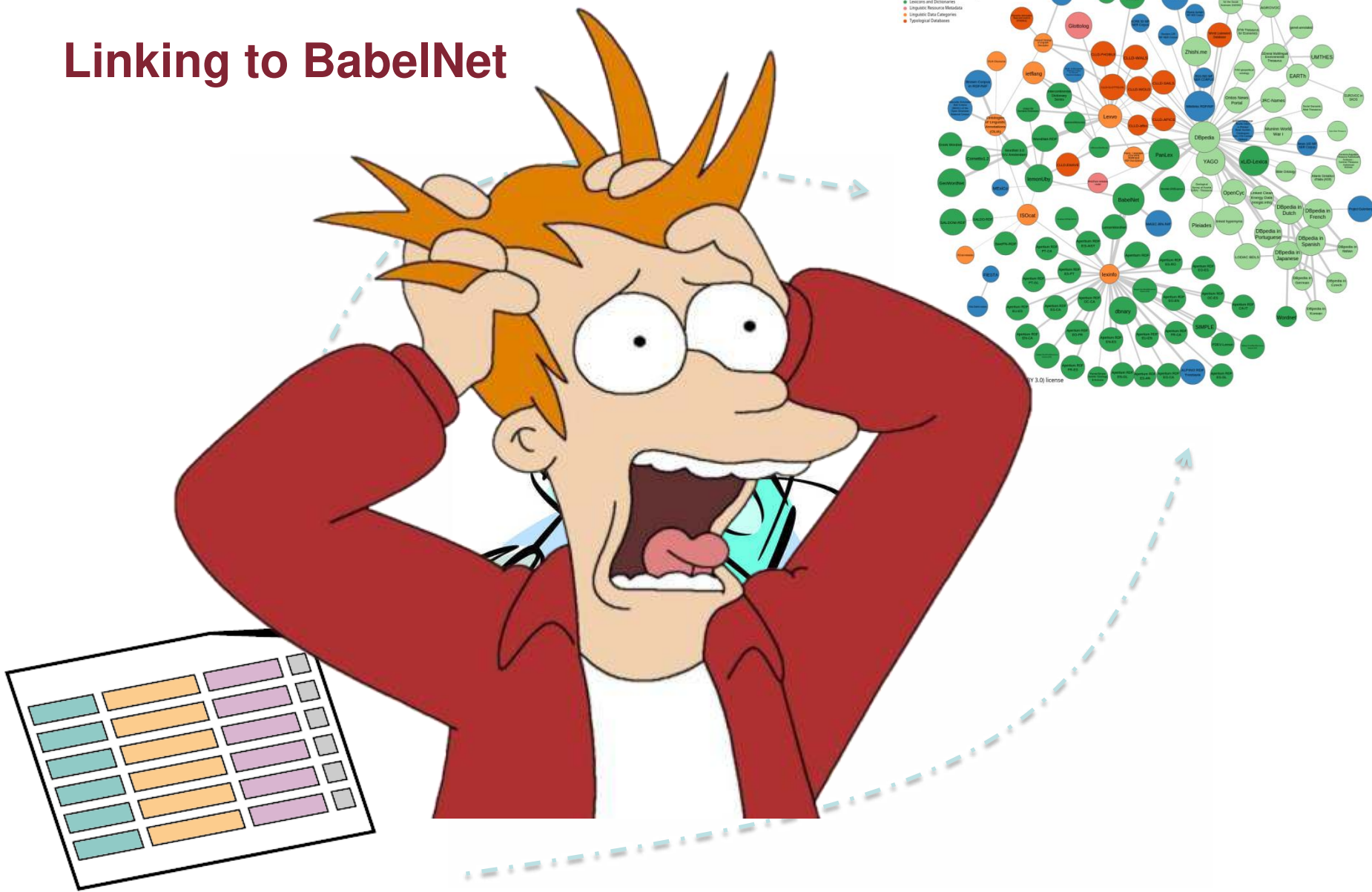
“The quick brown **fox** jumps over the lazy dog”



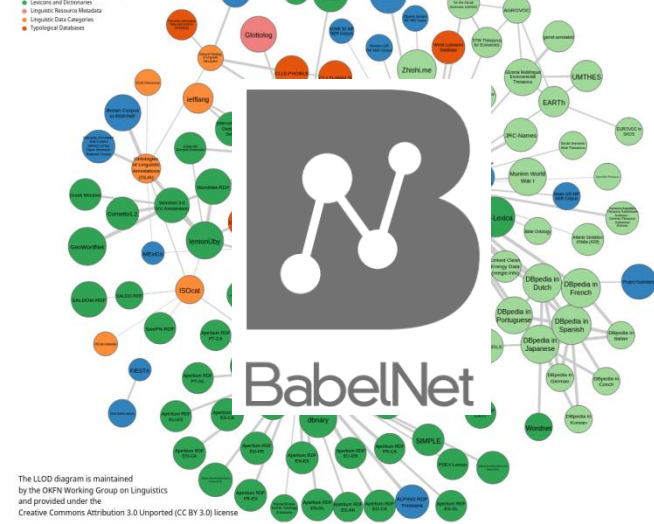
Linking to BabelNet



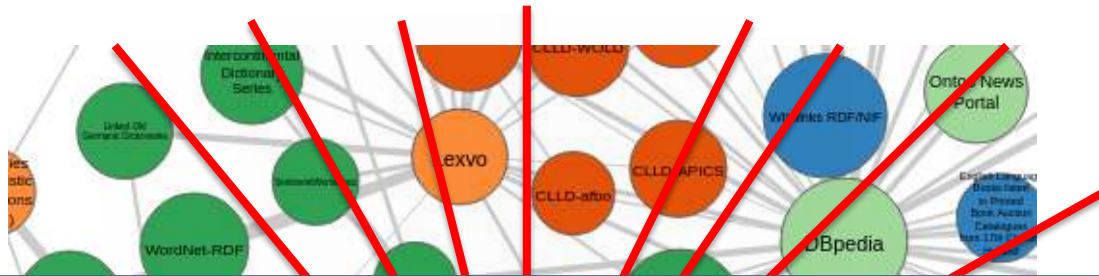
Linking to BabelNet



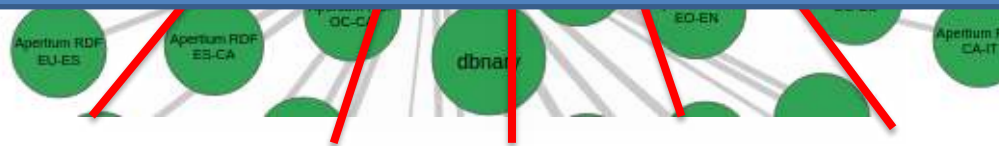
Linking to BabelNet



Linking to BabelNet



A key proxy
for
multilinguality!



Case study 0: Hello world (a minimal fake resource)

- Let's see how to perform a simple mapping between BabelNet and a minimal fake resource



Case study 0: Hello world (a minimal fake resource)

- Let's see how to perform a simple mapping between BabelNet and a minimal fake resource

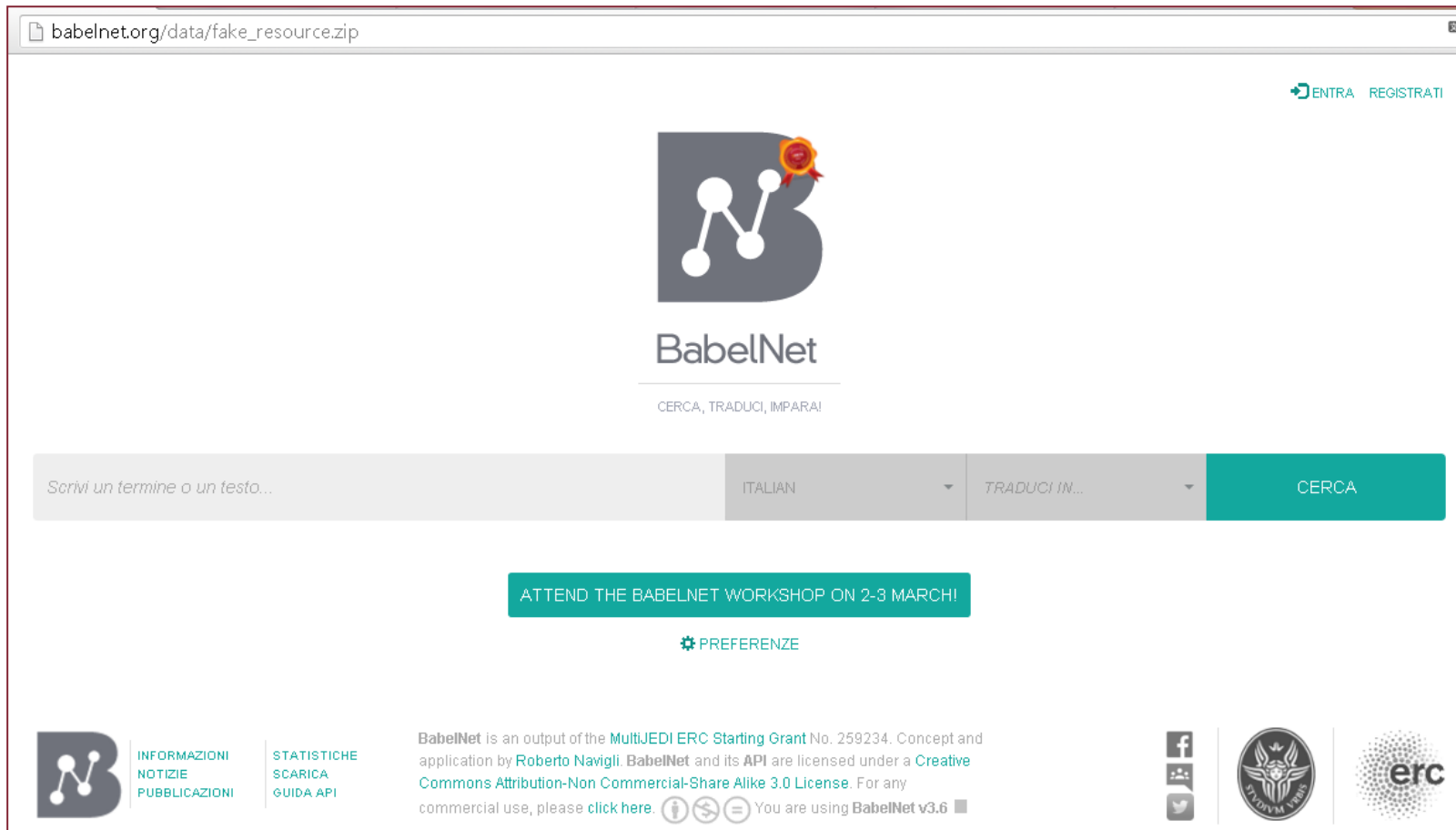


Case study 0: Hello world (a minimal fake resource)

- Open your browser

Case study 0: Hello world (a minimal fake resource)

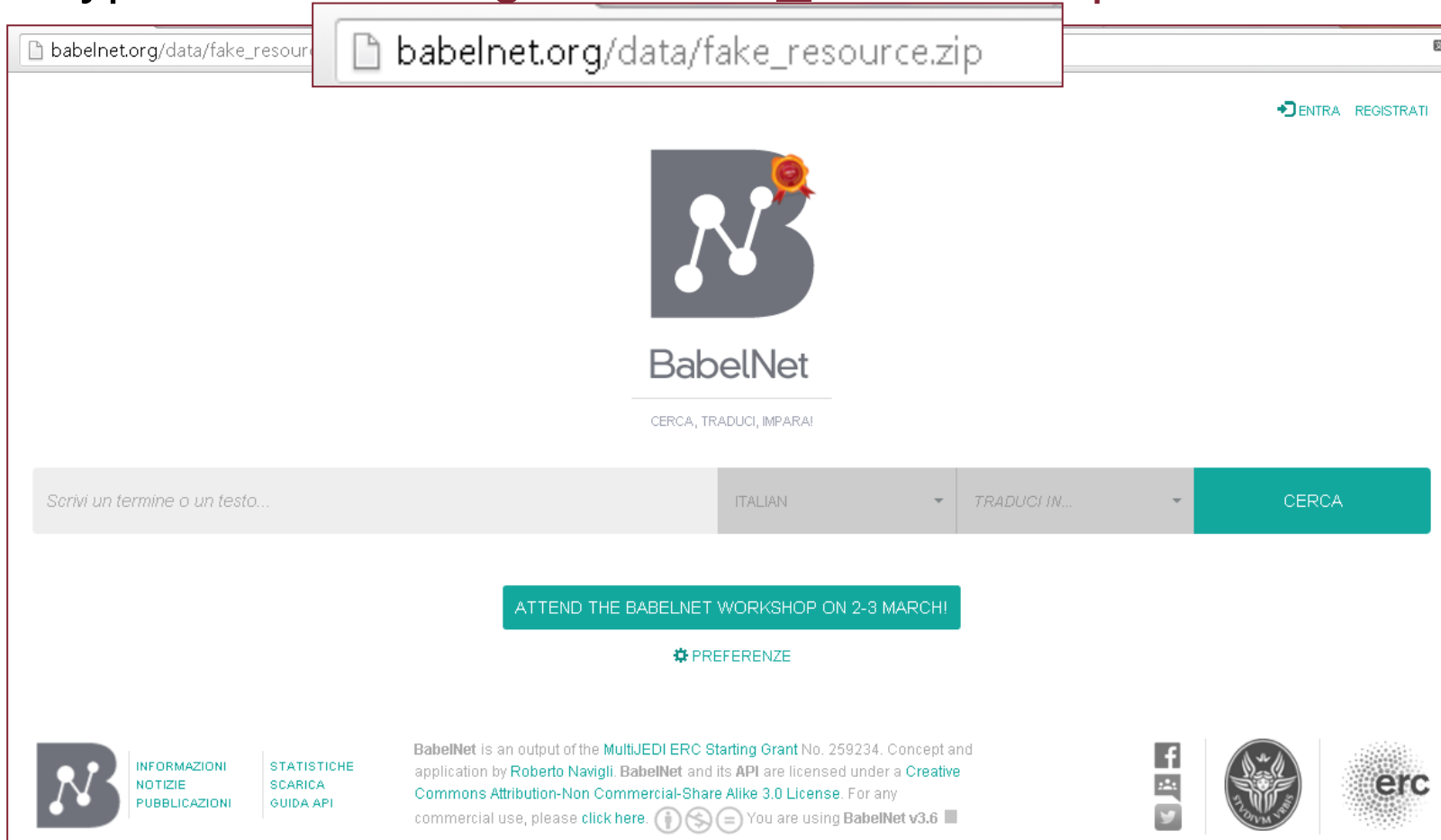
- Open your browser
- Type: babelnet.org/data/fake_resource.zip



The screenshot shows the BabelNet website interface. At the top, there is a navigation bar with "ENTRA" and "REGISTRATI" links. The main header features the BabelNet logo, which consists of a stylized 'B' with a network diagram inside, and the text "BabelNet" below it. Underneath the logo, the tagline "CERCA, TRADUCI, IMPARA!" is displayed. A search bar is prominently featured, containing the placeholder text "Scrivi un termine o un testo...". To the right of the search bar, there are dropdown menus for "ITALIAN" and "TRADUCI IN...", followed by a teal "CERCA" button. Below the search bar, a teal banner reads "ATTEND THE BABELNET WORKSHOP ON 2-3 MARCH!". A "PREFERENZE" link with a gear icon is located below the banner. The footer contains several elements: on the left, a BabelNet logo and links for "INFORMAZIONI", "NOTIZIE", "PUBBLICAZIONI", "STATISTICHE", "SCARICA", and "GUIDA API"; in the center, a paragraph of text stating "BabelNet is an output of the MultiJEDI ERC Starting Grant No. 259234. Concept and application by Roberto Navigli. BabelNet and its API are licensed under a Creative Commons Attribution-Non Commercial-Share Alike 3.0 License. For any commercial use, please click here." followed by icons for information, sharing, and a version indicator "You are using BabelNet v3.6"; on the right, social media icons for Facebook, YouTube, and Twitter, along with the logos for "STUDIVM IRES" and "erc".

Case study 0: Hello world (a minimal fake resource)

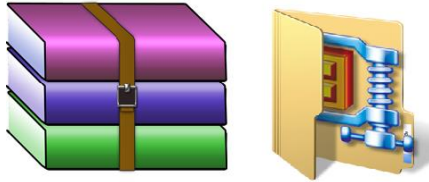
- Open your browser
- Type: `babelnet.org/data/fake_resource.zip`



The screenshot shows the BabelNet website interface. The browser's address bar is highlighted with a red box and contains the URL `babelnet.org/data/fake_resource.zip`. The website features the BabelNet logo, which is a stylized 'B' with a network diagram inside, and the text 'BabelNet' below it. Underneath the logo, it says 'CERCA, TRADUCI, IMPARA!'. A search bar is present with the placeholder text 'Scrivi un termine o un testo...'. To the right of the search bar, there are dropdown menus for 'ITALIAN' and 'TRADUCI IN...', followed by a teal 'CERCA' button. Below the search bar, there is a teal button that says 'ATTEND THE BABELNET WORKSHOP ON 2-3 MARCH!' and a 'PREFERENZE' link with a gear icon. The footer contains the BabelNet logo, navigation links for 'INFORMAZIONI', 'NOTIZIE', 'PUBBLICAZIONI', 'STATISTICHE', 'SCARICA', and 'GUIDA API', a paragraph of text about the MultiJEDI ERC Starting Grant, and social media icons for Facebook, Twitter, and YouTube. The footer also includes the logos for 'STUDIVM IERS' and 'erc'.

Case study 0: Hello world (a minimal fake resource)

- Open your browser
- Type: babelnet.org/data/fake_resource.zip
- Unpack your zip file



Case study 0: Hello world (a minimal fake resource)

1	Exhaust	VERB	To draw or let out wholly; to drain off completely; as, to exhaust the water of a well; the moisture of the earth is exhausted by evapora
2	Orb	NOUN	A period of time marked off by the revolution of a heavenly body.
3	Conduct	NOUN	Plot; action; construction; manner of development.
4	Cop	NOUN	A tube or quill upon which silk is wound.
		NOUN	The female of the sheep, and of sheeplike animals.
		VERB	To pierce or bore with a drill, or a with a drill; to perforate; as, to dr
		NO	rock; to drill a piece of metal.
8	Play	VE	play, as a ho
9	Superlative	NOUN	The superlative also, a form or word by which the superlative degree is expressed; as, strongest
10	Evidence	NOUN	That which is legally submitted to competent tribunal, as a means of ascertaining the truth of any alleged matter of fact under i
11	Hopper	NOUN	See Grasshopper, and Frog hopper, Grape hopper, Leaf hopper, Tree hopper, under Frog, Grape, Leaf, and Tree.
12	Cherub	NOUN	A symbolical winged figure of unknown form used in connection with the mercy seat of the Jewish Ark and Temple.
13	Breach	NOUN	A breaking up of amicable relations; rupture.
14	Manis	NOUN	A genus of edentates, covered with large, hard, triangular scales, with sharp edges that overlap each other like tiles on a roof. They in
15	Currency	NOUN	The state or quality of being current; general acceptance or reception; a passing from person to person, or from hand to hand; ci
16	Prosecute	VERB	To follow or pursue with a view to reach, execute, or accomplish; to endeavor to obtain or complete; to carry on; to continue; as
17	Indifference	NOUN	The quality or state of being indifferent, or not making a difference; want of sufficient importance to constitute a difference;
18	Solution	NOUN	The act or process by which a body (whether solid, liquid, or gaseous) is absorbed into a liquid, and, remaining or becoming flui
19	Mercury	NOUN	Sprightly or mercurial quality; spirit; mutability; fickleness.
20	Tor	NOUN	A tower; a turret.
21	Hog	NOUN	A rough, flat scrubbing broom for scrubbing a ship's bottom under water.
22	Pica	NOUN	A service-book. See Pie.
23	Mot	NOUN	A pithy or witty saying; a witticism.
24	Zone	NOUN	An area or part of a region characterized by uniform or similar animal and plant life; a life zone; as, Littoral zone, Austral zone, etc.
25	Marshal	NOUN	The chief officer of arms, whose duty it was, in ancient times, to regulate combats in the lists.
26	Harrow	NOUN	An implement of agriculture, usually formed of pieces of timber or metal crossing each other, and set with iron or wooden teeth. It is dr
27	Requisition	NOUN	A demand by the invader upon the people of an invaded country for supplies, as of provision, forage, transportation, etc.
28	Damnation	NOUN	A sin deserving of everlasting punishment.
29	Atmosphere	NOUN	The pressure or weight of the air at the sea level, on a unit of surface, or about 14.7 lbs. to the sq. inch.
30	Dictionary	NOUN	Hence, a book containing the words belonging to any system or province of knowledge, arranged alphabetically; as, a dictionary of
31	Sculpture	NOUN	The art of carving, cutting, or hewing wood, stone, metal, etc., into statues, ornaments, etc., or into figures, as of men, or ot
32	Ciliary	ADJECTIVE	Pertaining to or connected with the cilia in animal or vegetable organisms; as, ciliary motion.
33	Side	NOUN	Any outer portion of a thing considered apart from, and yet in relation to, the rest; as, the upper side of a sphere; also, any part or p
34	Sympathy	NOUN	An agreement of affections or inclinations, or a conformity of natural temperament, which causes persons to be pleased, or in acc
35	Wash	NOUN	Substances collected and deposited by the action of water; as, the wash of a sewer, of a river, etc.
36	Root	NOUN	The underground portion of a plant, whether a true root or a tuber, a bulb or rootstock, as in the potato, the onion, or the sweet flag.
37	Justification	NOUN	The act of justifying or the state of being justified; a showing or proving to be just or conformable to law, justice, right, or
38	Pacer	NOUN	One who, or that which, paces; especially, a horse that paces.
39	Bile	NOUN	Bitterness of feeling; cholera; anger; ill humor; as, to stir one's bile.
40	Jacob	NOUN	A Hebrew patriarch (son of Isaac, and ancestor of the Jews), who in a vision saw a ladder reaching up to heaven (Gen. xxviii. 12);
41	Repeater	NOUN	A watch with a striking apparatus which, upon pressure of a spring, will indicate the time, usually in hours and quarters.
42	Bevy	NOUN	A company; an assembly or collection of persons, especially of ladies.
43	Crown	NOUN	The person entitled to wear a regal or imperial crown; the sovereign;
44	Miner	NOUN	One who mines; a digger for metals, etc.; one engaged in the business of getting ore, coal, or precious stones, out of the earth; one who
45	Chivalry	NOUN	Exploit.
46	Tinker	NOUN	A young mackerel about two years old.
47	Charge	NOUN	Thirty-six pigs of lead, each pig weighing about seventy pounds;
48	Virtuous	ADJECTIVE	Having moral excellence; characterized by morality; upright; righteous; pure; as, a virtuous action.
49	Visibility	NOUN	The quality or state of being visible.
50	Ward	NOUN	To defend by walls, fortifications, etc.
51	Tourniquet	NOUN	An instrument for arresting hemorrhage. It consists essentially of a pad or compress upon which pressure is made by a band which

ID

Lemma

POS (Part Of Speech)

Glosses

Case study 0: Hello world (a minimal fake resource)

ARE YOU READY?



Case study 0: Hello world (a minimal fake resource)



Case study 0: Hello world (a minimal fake resource)



Browser User

Java Programmer

Case study 0: Hello world (a minimal fake resource)

- First, let's see how to perform a semi-automatic mapping (for **Browser Users**)



- Then we will create a simple program for automatically mapping via the BabelNet and Babelify Java APIs (for **Java Programmers**)

Case study 0: Hello world (a minimal fake resource)

- First, let's see how to perform a semi-automatic mapping (for **Browser Users**)



Case study 0: Hello world (a minimal fake resource)



- Steps to follow to perform a **semi-automatic mapping**

1084	Cow	NOUN	A chimney cap; a cowl
1085	Press	NOUN	An apparatus or machine by which any substance or body is pressed,
1086	Lap	NOUN	To lay or place over anything so as to partly or wholly cover it;
1087	Lotus	NOUN	The lote, or nettle tree. See Lote.
1088	Angle	NOUN	The figure made by two lines which meet.
1089	Cyprian	NOUN	A lewd woman; a harlot.
1090	Uncertainty	NOUN	That which is uncertain; something unknown.
1091	Base	NOUN	A place or tract of country, protected by fortifications, or by na

- Copy a **lemma** defined in our resource to [Babelify.org](https://babelify.org)



Text to babelify...

Enable partial matches:

ENGLISH

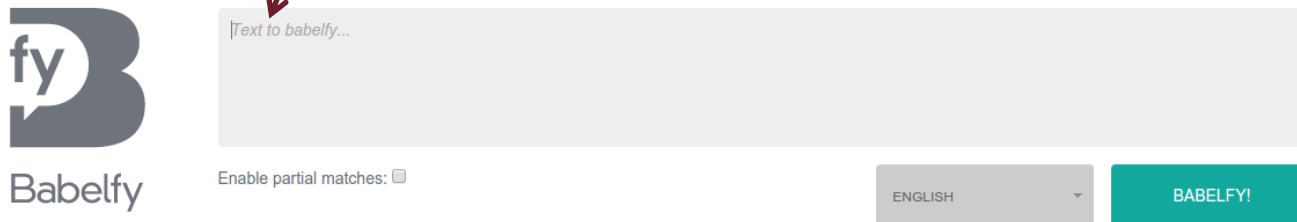
BABELFY!

Case study 0: Hello world (a minimal fake resource)



1088	Angle	NOUN	The figure made by two lines which meet.
1089	Cypria	NOUN	A lewd woman; a harlot.
1090	Uncertainty	NOUN	That which is uncertain; something unknown.
1091	Base	NOUN	A place or tract of country, protected by fortifications, or by na

- Copy a **lemma** defined in our resource to [Babelfy.org](https://babelfy.org)



Case study 0: Hello world (a minimal fake resource)



- Also add the **Glosses** available for that lemma

1084	Cow	NOUN	A chimney cap; a cowl
1085	Press	NOUN	An apparatus or machine by which any substance or body is pressed,
1086	Lap	NOUN	To lay or place over anything so as to partly or wholly cover it;
1087	Lotus	NOUN	The lute, or nettle tree. See Lote.
1088	Angle	NOUN	The figure made by two lines which meet.
1089	Cyprian	NOUN	A lewd woman; a harlot.
1090	Uncertainty	NOUN	That which is uncertain; something unknown.
1091	Base	NOUN	A place or tract of country, protected by fortifications, or by na

The screenshot shows the Babelify search interface. On the left is the Babelify logo. In the center is a search bar containing the text "Angle.". Below the search bar is a checkbox labeled "Enable partial matches:". On the right side of the interface are two buttons: "ENGLISH" and "BABELFY!". A red arrow points from the "Angle" entry in the table above to the search bar.

Case study 0: Hello world (a minimal fake resource)



- Also add the **Glosses** available for that lemma

1084 Cow NOUN A chimney cap; a cowl

The screenshot shows the BabelNet interface. On the left is the 'fy B Babelfy' logo. The search input field contains the text 'Angle, The figure made by. two lines which meet.'. Below the input field is a checkbox labeled 'Enable partial matches:'. To the right of the input field is a dropdown menu set to 'ENGLISH' and a teal button labeled 'BABELFY!'. A red box highlights the search input field and the 'BABELFY!' button. A red arrow points from the word 'Glosses' in the list above to the search input field.

The screenshot shows the BabelNet interface. On the left is the 'fy B Babelfy' logo. The search input field contains the text 'Angle.'. Below the input field is a checkbox labeled 'Enable partial matches:'. To the right of the input field is a dropdown menu set to 'ENGLISH' and a teal button labeled 'BABELFY!'. A red box highlights the search input field and the 'BABELFY!' button. A red arrow points from the search input field in the top screenshot to the search input field in this screenshot.

Click!!



BabelNet

[Type a term or a text...] ENGLISH TRANSLATE INTO... SEARCH

PREFERENCES

- English
- Arabic
- Chinese
- French
- Greek
- Hebrew
- Hindi
- Italian
- Japanese
- Russian
- Spanish

- Dictionary
- Images
- Translations
- Sources
- Categories
- Compounds
- Other forms
- External Links

• **bn:00004117n** • NOUN • Concept • Categories: Angle

EN angle • **Adjacent angle** • **Complementary angles** • **Supplementary angles** • **Vertical angles**

The space between two **lines** or **planes** that **intersect**; the inclination of one **line** to another; measured in degrees or **radians** [More definitions](#)

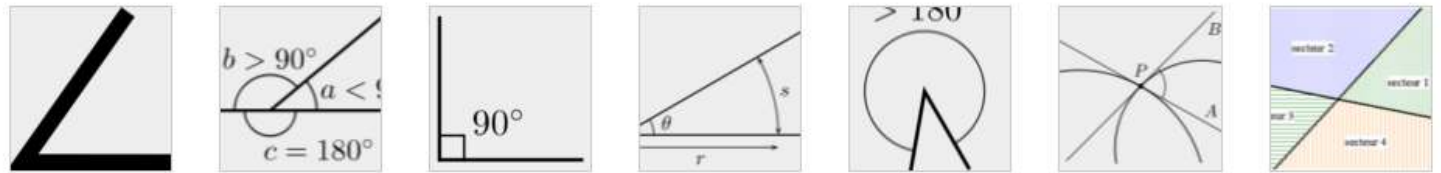
The angle between lines A and B

IS A: [space](#) • [physical quantity](#) • [geometry](#)

EXPLORE NETWORK

RELATED:

- plane angle
- vertical angle
- reflex angle
- ray
- milliradian
- straight line
- exterior angle
- Sum of angles of a tria...
- Euclidean geometry
- Solution of triangles
- tan





BabelNet

[Type a term or a text...] ENGLISH TRANSLATE INTO... SEARCH

- English
- Arabic
- Chinese
- French
- Greek
- Hebrew
- Hindi
- Italian
- Japanese
- Russian
- Spanish

- Dictionary
- Images
- Translations
- Sources
- Categories
- Compounds
- Other forms
- External Links

bn:00004117n • NOUN • Concept • Categories: Angle

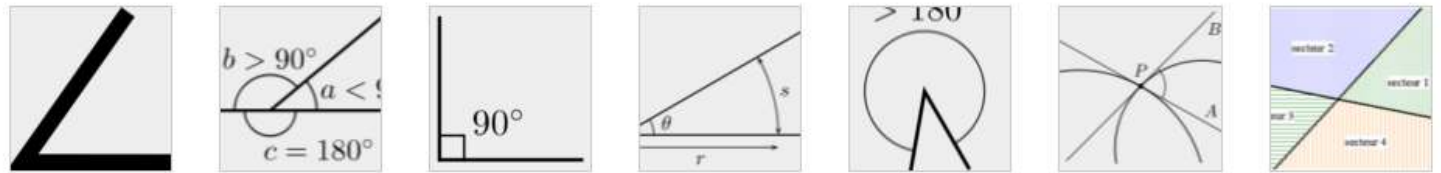
The space between two lines or planes that intersect; the inclination of one line to another; measured in degrees or radians

The angle between lines A and B

IS A: space • physical quantity • geometry

1088 Angle NOUN The figure made by two lines which meet.

- RELATED:
- plane angle
 - vertical angle
 - reflex angle
 - ray
 - milliradian
 - straight line
 - exterior angle
 - Sum of angles of a tria...
 - Euclidean geometry
 - Solution of triangles
 - tan





BabelNet

[Type a term or a text...] ENGLISH TRANSLATE INTO... SEARCH

- English
- Arabic
- Chinese
- French
- Greek
- Hebrew
- Hindi
- Italian
- Japanese
- Russian
- Spanish

- Dictionary
- Images
- Translations
- Sources
- Categories
- Compounds
- Other forms
- External Links

bn:00004117n • NOUN • Concept • Categories: Angle

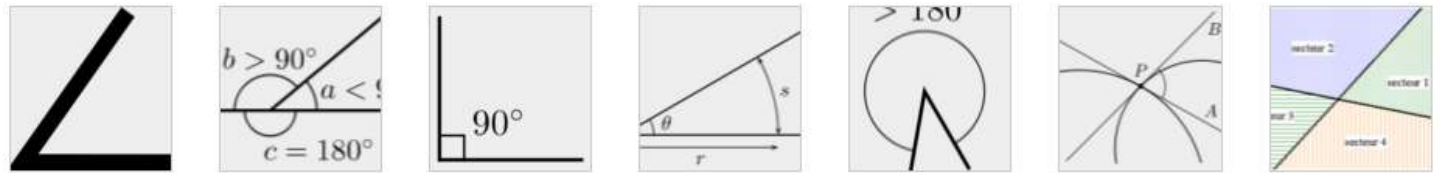
The space between two lines or planes that intersect; the inclination of one line to another; measured in degrees or radians

The angle between lines A and B

IS A: space • physical quantity • geometry

1088 Angle NOUN The figure made by two lines which meet.

- RELATED:
- plane angle
 - vertical angle
 - reflex angle
 - ray
 - milliradian
 - straight line
 - exterior angle
 - Sum of angles of a tria...
 - Euclidean geometry
 - Solution of triangles
 - tan



Case study 0: Hello world (a minimal fake resource)



- Babelfy disambiguates "Angle" with the synset `bn:00004117n`
- We add this answer to the annotation form created for this session
- Type this URI on your browser:
<http://goo.gl/forms/gNm7wRPJJX>
- OR go to the workshop site (<http://babelnet.org/lux>) session 6 and click on "form".
- Answer the questions! :D

Case study 0: Hello world (a minimal fake resource)

The Luxembourg BabelNet Workshop - Session 6

Result of mapping using Babelify on Browser

*Campo obbligatorio

ID FAKE_RESOURCE *

La tua risposta

ID BABELNET (e.g bn:00000001n or null) *

La tua risposta

IS CORRECT? *

Yes!

No!

INVIA

Non inviare mai le password tramite Moduli Google.

1088

Angle NOUN



bn:00004117n *



Send

Case study 0: Hello world (a minimal fake resource)

The Luxembourg BabelNet Workshop

- Sess

Result of mappi

*Campo obbliga

ID FAKE_R

La tua rispos

ID BABELN

La tua rispos

IS CORREC

Yes!

No!

INVIA

Non inviare mai le password tramite Moduli Google.

I'D LIKE TO SEE YOU TRY



Send

Case study 0: Hello world (a minimal fake resource)



- Try yourself to map the fake resource to BabelNet
- Choose some random entries
- Apply the above steps
- Fill in the form for each mapped entry (type **null** if the target word cannot be disambiguated)

Case study 0: Hello world (a minimal fake resource)

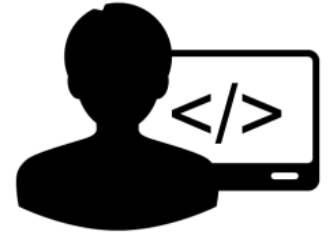


Case study 0: Hello world (a minimal fake resource)



- We will create a simple program for automatically mapping through the BabelNet and Babelfy Java API (for **Java Programmers**)

Case study 0: Hello world (a minimal fake resource)

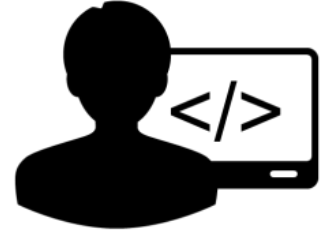


- Create a project containing
 - The BabelNet API jar
 - The Babelfy API jar
- Create a simple mapping class (Mapping) that:
 - Reads the fake resource (FR) line by line
 - Retrieves all the synsets for the target lemma with the BabelNet API
 - Disambiguates the lemma and its definitions using Babelfy
 - Checks if the target lemma has been disambiguated
 - Returns the synset provided by Babelfy for that lemma, if available; else, returns null
 - If you are also using the BabelNet API, you can retrieve more information about



Case study 0: Hello world (a minimal fake resource)

- Create a project containing
 - The BabelNet API jar
 - The Babelfy API jar



Like we did yesterday during the Babelfy tech session

Case study 0: Hello world (a minimal fake resource)



- Create a simple mapping class that:
 - Reads the fake resource (FR) line by line



Case study 0: Hello world (a minimal fake resource)



- Create a simple fake resource
 - Reads the fake resource

```
private void mapping()
{
    try {
        BabelNet bn = BabelNet.getInstance();
        BufferedReader reader = Files.getBufferedReader(PATH+"fake_resource.txt");
        BufferedWriter bufferedWriter = Files.getBufferedWriter(PATH+"mapping_fake_resource.txt");
        String line = null;
        //read file
        while ((line = reader.readLine()) != null) {
            String[] fields = line.split("\t");
            String id = fields[0];
            String lemma = fields[1];
            BabelPOS pos = BabelPOS.valueOf(fields[2]);
            String glosses = fields[3];

            HashSet<BabelSynsetID> candidates = new HashSet<BabelSynsetID>();
            //
            for(BabelSynset by : bn.getSynsets(lemma, Language.EN, pos))
            {
                candidates.add(by.getId());
            }
            //
            BabelSynsetID idMapping = applyBabelfy(lemma+", "+glosses, candidates);
            if(idMapping!=null)
            {
                BabelSynset babelSynset = idMapping.toBabelSynset();
                bufferedWriter.append(id+"\t"+lemma+"\t"+pos+"\t"+idMapping+"\n"
                    + "BABELNET:\t"+babelSynset.getMainGloss(Language.EN)+"\n"
                    + "FAKE_RESOURCE:\t"+glosses+"\n\n");
            }
        }
    }
    bufferedWriter.flush();
    bufferedWriter.close();
    reader.close();
} catch (FileNotFoundException e) {
    e.printStackTrace();
} catch (IOException e) {
    e.printStackTrace();
}
}
```

• Just neglect the greyed parts for the moment

• Code:

- Open the file
- Parse each line
- Close the file



Case study 0: Hello world (a minimal fake resource)



- Create a simple mapping class that:
 - Retrieves all synsets in BabelNet for the target lemma to be mapped

Case study 0: Hello world (a minimal fake resource)



- Create a simple
– Retrieves all s
mapped

• Just neglect the greyed
parts for the moment

• Code:

- Obtain the BabelNet
object
- Retrieve synsets that
contain the candidate
lemma in English
- Add their IDs to a set
called "candidates"

```
private void mapping()
{
    try {
        BabelNet bn = BabelNet.getInstance();
        BufferedReader reader = Files.getBufferedReader(PATH+"fake_resource.txt");
        BufferedWriter bufferedWriter = Files.getBufferedWriter(PATH+"mapping_fake_resource.txt");
        String line = null;
        //read file
        while ((line = reader.readLine()) != null) {
            String[] fields = line.split("\t");
            String id = fields[0];
            String lemma = fields[1];
            BabelPOS pos = BabelPOS.valueOf(fields[2]);
            String glosses = fields[3];
            HashSet<BabelSynsetID> candidates = new HashSet<BabelSynsetID>();
            //
            for(BabelSynset by : bn.getSynsets(lemma, Language.EN, pos))
            {
                candidates.add(by.getId());
            }
            //
            BabelSynsetID idMapping = applyBabelfy(lemma+", "+glosses, candidates);
            if(idMapping!=null)
            {
                BabelSynset babelSynset = idMapping.toBabelSynset();
                bufferedWriter.append(id+"\t"+lemma+"\t"+pos+"\t"+idMapping+"\n"
                    + "BABELNET:\t"+babelSynset.getMainGloss(Language.EN)+"\n"
                    + "FAKE_RESOURCE:\t"+glosses+"\n\n");
            }
        }
        bufferedWriter.flush();
        bufferedWriter.close();
        reader.close();
    } catch (FileNotFoundException e) {
        e.printStackTrace();
    } catch (IOException e) {
        e.printStackTrace();
    }
}
```



Case study 0: Hello world (a minimal fake resource)



- Create a simple mapping class that:
 - Disambiguates the target lemma and its definitions using Babelfy

Case study 0: Hello world (a minimal fake resource)



- Create a simple disambiguator
– Disambiguates Babelfy

```
private void mapping()
{
    try {
        BabelNet bn = BabelNet.getInstance();
        BufferedReader reader = Files.getBufferedReader(PATH+"fake_resource.txt");
        BufferedWriter bufferedWriter = Files.getBufferedWriter(PATH+"mapping_fake_resource.txt");
        String line = null;
        //read file
        while ((line = reader.readLine()) != null) {
            String[] fields = line.split("\t");
            String id = fields[0];
            String lemma = fields[1];
            BabelPOS pos = BabelPOS.valueOf(fields[2]);
            String glosses = fields[3];
            HashSet<BabelSynsetID> candidates = new HashSet<BabelSynsetID>();
            //
            for(BabelSynset by : bn.getSynsets(lemma, Language.EN, pos))
            {
                candidates.add(by.getId());
            }
            //
            BabelSynsetID idMapping = applyBabelfy(lemma+", "+glosses, candidates);
            //
            {
                BabelSynset babelSynset = idMapping.toBabelSynset();
                bufferedWriter.append(id+"\t"+lemma+"\t"+pos+"\t"+idMapping+"\n"
                    + "BABELNET:\t"+babelSynset.getMainGloss(Language.EN)+"\n"
                    + "FAKE_RESOURCE:\t"+glosses+"\n\n");
            }
        }
        bufferedWriter.flush();
        bufferedWriter.close();
        reader.close();
    } catch (FileNotFoundException e) {
        e.printStackTrace();
    } catch (IOException e) {
        e.printStackTrace();
    }
}
```

Just neglect the greyed parts for the moment

Code:

- Retrieve the lemma and its definitions and pass them to the disambiguation method "applyBabelfy" together with the candidate synsets for the target lemma

Case study 0: Hello world (a minimal fake resource)



- Create a simple mapping class that:
 - Disambiguates the target lemma and its definitions using Babelfy

```
//  
BabelSynsetID idMapping = applyBabelfy(lemma+", "+glosses, candidates);  
if(idMapping==null)
```

Code:

- Obtain the Babelfy object
- Disambiguate the lemma with its context
- Retrieve all the disambiguated concepts
- Check if there is an intersection with the candidate synsets in input
- If only a single concept is in the intersection, we return it; else, we return null

```
private BabelSynsetID applyBabelfy(String glosses, HashSet<BabelSynsetID> ids)  
{  
    Babelfy bfy = new Babelfy();  
    //  
    List<SemanticAnnotation> bfyAnnotations = bfy.babelfy( glosses, Language.EN);  
    //|  
    HashSet<BabelSynsetID> synsetsInGloss = new HashSet<BabelSynsetID>();  
    for (SemanticAnnotation annotation : bfyAnnotations)  
    {  
        synsetsInGloss.add(new BabelSynsetID(annotation.getBabelSynsetID()));  
    }  
    SetView<BabelSynsetID> intersection = Sets.intersection(ids, synsetsInGloss);  
    if(intersection.isEmpty())return null;  
    if(intersection.size() > 1)return null;  
    else return intersection.iterator().next();  
}
```



Case study 0: Hello world (a minimal fake resource)



- Create a simple mapping class that:
 - Returns the Babel synset corresponding to the input entry

Case study 0: Hello world (a minimal fake resource)



- Create a simple fake resource
 - Returns the BabelNet instance

```
private void mapping()
{
    try {
        BabelNet bn = BabelNet.getInstance();
        BufferedReader reader = Files.getBufferedReader(PATH+"fake_resource.txt");
        BufferedWriter bufferedWriter = Files.getBufferedWriter(PATH+"mapping_fake_resource.txt");
        String line = null;
        //read file
        while ((line = reader.readLine()) != null) {
            String[] fields = line.split("\t");
            String id = fields[0];
            String lemma = fields[1];
            BabelPOS pos = BabelPOS.valueOf(fields[2]);
            String glosses = fields[3];
            HashSet<BabelSynsetID> candidates = new HashSet<BabelSynsetID>();
            //
            for(BabelSynset by : bn.getSynsets(lemma, Language.EN, pos))
            {
                candidates.add(by.getId());
            }
            //
            BabelSynsetID idMapping = applyBabelfy(lemma+", "+glosses, candidates);
            if(idMapping!=null)
            {
                BabelSynset babelSynset = idMapping.toBabelSynset();
                bufferedWriter.append(id+"\t"+lemma+"\t"+pos+"\t"+idMapping+"\n"
                    + "BABELNET:\t"+babelSynset.getMainGloss(Language.EN)+"\n"
                    + "FAKE_RESOURCE:\t"+glosses+"\n\n");
            }
        }
        bufferedWriter.flush();
        bufferedWriter.close();
        reader.close();
    } catch (FileNotFoundException e) {
        e.printStackTrace();
    } catch (IOException e) {
        e.printStackTrace();
    }
}
```

Codice:

- Create the output file
- Check if a valid candidate exists
- Write the output file



Case study 0: Hello world (a minimal fake resource)



- Run the class:

```
public static void main(String[] args) throws IOException {  
    Mapping test = new Mapping();  
    test.mapping();  
}
```

8806> Imperial> NOUN> bn:00046103n
BABELNET:> A piece of luggage carried on top of a coach
FAKE_RESOURCE:> A luggage case on the top of a coach.

8807> Spender> NOUN> bn:00069538n
BABELNET:> Someone who spends money prodigally
FAKE_RESOURCE:> One who spends; esp., one who spends lavishly; a prodigal; a spendthrift.

8808> Issue> NOUN> bn:00047692n
BABELNET:> The immediate descendants of a person
FAKE_RESOURCE:> Progeny; a child or children; offspring. In law, sometimes, in a general sense, all persons descended from a common ancestor; all lineal descendants.

8809> Slip> NOUN> bn:00010074n
BABELNET:> A place where a craft can be made fast
FAKE_RESOURCE:> To move or fly (out of place); to shoot;

8810> Nick> VERB> bn:00091058v
BABELNET:> Cut slightly, with a razor
FAKE_RESOURCE:> To make a nick or nicks in; to notch; to keep count of or upon by nicks; as, to nick a stick, tally, etc.

8811> Bureau> NOUN> bn:00001961n
BABELNET:> An administrative unit of government
FAKE_RESOURCE:> Hence: A department of public business requiring a force of clerks; the body of officials in a department who labor under the direction of a chief.

8812> Hornpipe> NOUN> bn:00044804n
BABELNET:> An ancient (now obsolete) single-reed woodwind; usually made of bone
FAKE_RESOURCE:> An instrument of music formerly popular in Wales, consisting of a wooden pipe, with holes at intervals. It was so called because the bell at the open

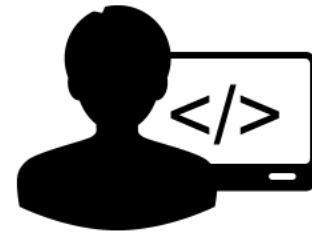
8813> Lose> VERB> bn:00090485v
BABELNET:> Fail to keep or to maintain; cease to have, either physically or in an abstract sense

Case study 0: Hello world (a minimal fake resource)

- In summary, we have just seen how to map a simple resource by using a manual and an automatic procedure, both based on Babelify
- In the next sessions, we will see more complex strategies to map EU resources!



Browser User



Java Programmer

