BabelNet and Word Sense Disambiguation

Overview:

- Original BabelNet
- BabelNet 2.5 (‘today’)
- Extrinsic Evaluations (SemEval-2007 T#16, SemEval-2007 T#7)
- SemEval-2010 T#3, 2013

Next episode (preview):

- Babelfy (an online, unified graph-based approach to EL and WSD)

NANYANG TECHNOLOGICAL UNIVERSITY

BabelNet - Quick Introduction

BabelNet

- Large and wide coverage multilingual semantic network;
- Integrates lexicographic and encyclopaedic knowledge; (WordNet vs Wikipedia)
- Further enriched by Machine Translation;
- Coverage for 50 languages;
- +9 million entries;

How? With the automatic integration of:
WordNet, OMW (☺), Wikipedia, OmegaWiki, Wiktionary, and Wikidata + SMT of senses across languages

NANYANG TECHNOLOGICAL UNIVERSITY

Lexicographic & Encyclopaedic Knowledge

An “Encyclopaedic Dictionary” by merging:

Wordnet

- concepts = sets of synonyms (synsets, ss);
- POS marking and word polysemy (1 word, many ss);
- Synset definitions/glosses;
- Synset example sentences;
- Lexical and semantic relations (e.g. *is-a, is-part-of, antonym, in-domain-of, etc.*);

\+ **Gloss relations;**

Gloss relations

Given a Synset, S, and the set of disambiguated word in its gloss(S),

$$s_i \in \textit{gloss}(S) = \{s_1,\ldots,s_k\}, \quad i = 1,\ldots,k.$$

There is a relation between S and all the synsets contained in its disambiguated gloss;

NANYANG TECHNOLOGICAL UNIVERSITY

Lexicographic & Encyclopaedic Knowledge

An “Encyclopedic Dictionary” by merging:

Wordnet

- concepts = sets of synonyms (synsets, ss);
- POS marking and word polysemy (1 word, many ss);
- Synset definitions/glosses;
- Synset example sentences;
- Lexical and semantic relations (e.g. *is-a, is-part-of, antonym, in-domain-of, etc.*);

\+ **Gloss relations;**

Wikipedia

- 1 article/page = 1 concept;
- Title of article = lemma;
- (opt.) Title label to help disambiguate the lemma (e.g. ‘play (activity)’ vs. ‘play (theatre);
- partly structured text (e.g. gloss is provided in the 1st sentence, info boxes with summarised info);
- Article relations (e.g. *redirect pages, disambiguation pages, internal links, …*);

NANYANG TECHNOLOGICAL UNIVERSITY

Lexicographic & Encyclopaedic Knowledge

An “Encyclopedic Dictionary” by merging:

Wikipedia

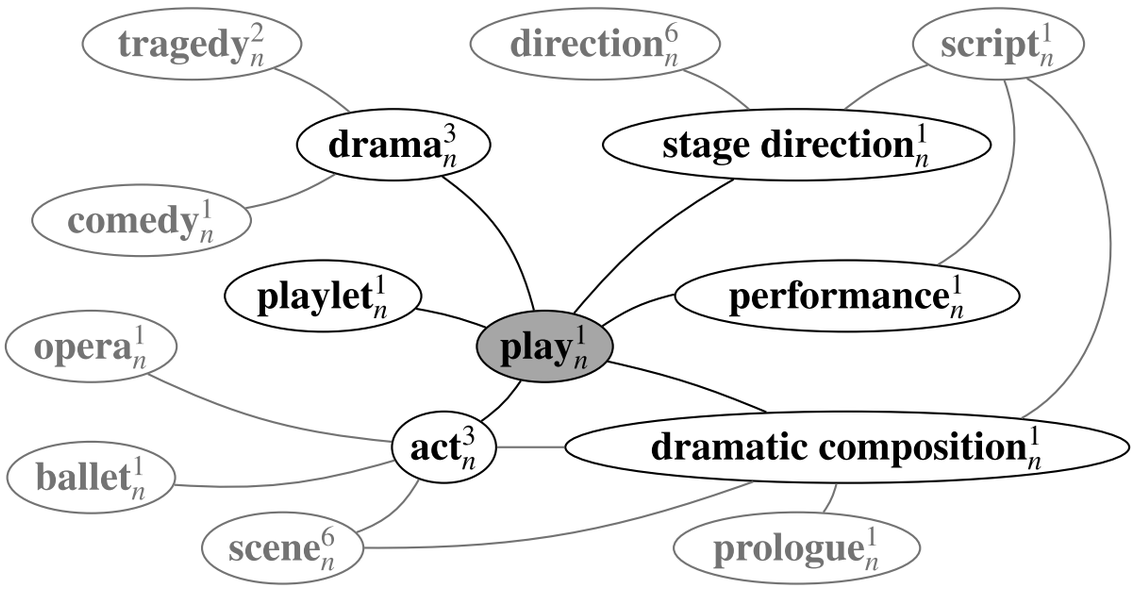
- 1 article/page = 1 concept;
- Title of article = lemma;
- (opt.) Title label to help disambiguate the lemma (e.g. ‘play (activity)’ vs. ‘play (theatre)’;
- partly structured text (e.g. gloss is provided in the 1st sentence, info boxes with summarised info);
- Article relations (e.g. *redirect pages, disambiguation pages, internal links, …*);

- Redirect pages ≈ synonymity relations;
- Disambiguation pages ≈ word polysemy;
- Inter-language links ≈ synset keys (cross lingual);
- Internal links ≈ related synsets;
- Categories ≈ related synsets

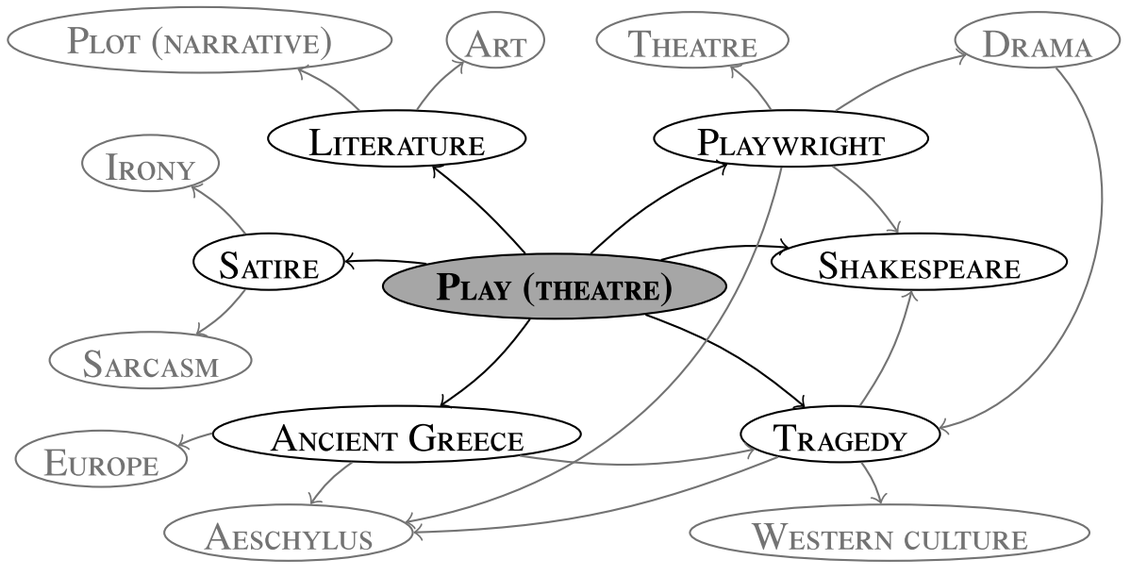
NANYANG TECHNOLOGICAL UNIVERSITY

Lexicographic & Encyclopaedic Knowledge

Wordnet



Wikipedia



Both can be viewed as graphs (w/ articles and synsets as nodes and relations and hyperlinks as edges).

It is evident that the two graphs complement each other.

* images obtained from [1]

NANYANG TECHNOLOGICAL UNIVERSITY

BabelNet - More Formally

- **Labeled directed graph** with a set of nodes V (concepts & named entities) and set of labeled edges $E \subseteq V \times R \times V$, that connect two nodes with a semantic relation from R, i.e., {is-a,part-of,…,ε}; (ε = unspecified semantic relation)
- Each node $v \in V$ contains a set of lexicalizations in multiple languages > referred to as **Babel synsets**;
- One unified resource in <u>three steps</u>:
  - Combine WordNet and Wikipedia;
  - Harvest multilingual lexicalizations;
  - Harvest relations between Babel synsets;

1. Combine WordNet and Wikipedia

Automatically acquiring a mapping between WordNet senses and Wikipages:

For $w \in \text{Senses}_{\text{Wiki}}$, (given by either its title or the main token)

$$\mu(w) = \begin{cases} s \in \text{Senses}_{\text{WN}}(w) \text{ if a link can be established,} \\ \varepsilon \text{ otherwise,} \end{cases}$$

- Treat mapping as a **disambiguation problem** - use disambiguation context to decide mapping;
- Mapping Algorithm - given w, **finds s that maximizes the probability of s providing** an adequate **corresponding concept for w;**
- Estimate the mapping conditional probability with two methods - simple bag-of-words (BoW), and graph based approach;

NANYANG TECHNOLOGICAL UNIVERSITY

1. Combine WordNet and Wikipedia

Pseudocode of the mapping algorithm:

for each

for each
  if
  then : $\mu(w) := w$

for each
  if $\mu(w) = \varepsilon$ **then :**

if $\mu(d) \neq \varepsilon$ and $\mu(d)$ is in a synset of w **then :**
$\mu(w)$:= sense of w in synset of $\mu(d)$; **break**

for each
  if
    if no tie occurs then :
$s \in SensesWN(w)$ $s \in SensesWN(w)$

return μ

1. Combine WordNet and Wikipedia

P(s, w) - The joint probability of a WordNet sense and Wikipage, or "the probability of a WordNet sense and Wikipage referring to the same concept":

- Similar to **WSD**
- The **disambiguation context** for each of the two concepts is the **set of words with some semantic relation to each** concept (from the corresponding resource).
  - labels, links, redirections and categories - WikiSenses (w)
  - synonymy, hypernymy/hyponymy and gloss - WNSenses (s)

Ctx(w) , Ctx(Play (theatre)) = {'theatre', 'literature', 'comedy', 'drama', 'character', … }

Ctx(s) , Ctx(play#01$_n$) = {'drama', 'composition', 'work', 'intend', 'actor', 'stage', … }

NANYANG TECHNOLOGICAL UNIVERSITY

1. Combine WordNet and Wikipedia

Back do the probability estimation:

$$p(s, w) = \frac{score(s, w)}{\sum_{\substack{s' \in Senses_{\mathrm{WN}}(w), \\ w' \in Senses_{\mathrm{Wiki}}(w)}} score(s', w')}$$

Two methods for computing score(s,w):

- **Bag-of-words method**

  score(s, w) = |Ctx(s) ⋂ Ctx(w)| + 1 (smoothing)

- **Graph-based method**

  Transforms Ctx(w) into a subgraph of the Wordnet containing all the word in context and all edges and intermediate senses along all paths of a maximal length L.

  The scoring function is then defined as:

$$score(s, w) = \sum_{cw \in Ctx(w)} \sum_{s' \in Senses_{\mathrm{WN}}(cw)} \sum_{p \in paths_{\mathrm{WN}}(s,s')} e^{-(length(p)-1)}$$

NANYANG TECHNOLOGICAL UNIVERSITY

2. Translating Babel synsets

So far, **Babel Synsets** are **S** ∪ **W**, and where **W** includes:

- w - WikiSense;
- the set of redirections to w;
- all inter-language links;
- the redirections to the inter-language links found in the Wikipedia of the target language.

Two issues:

- **Unlinked concepts** between Wiki and WN
- Even if linked, Wiki may **not provide any/all translations**

To guarantee **coverage for all languages**, this was also tackled with **automated processes**.

2. Translating Babel synsets

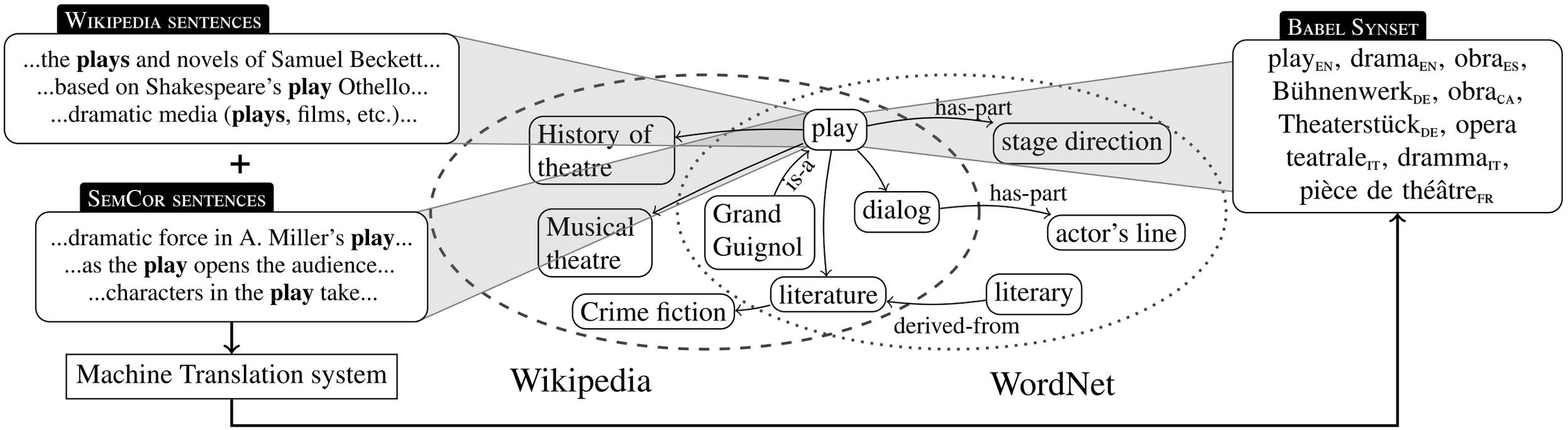
They wanted **full coverage for 6 languages! > SMT!**

For each polysemous WNsense and WikiSense, SemCor and Wikipedia were mined for sentences (respectively) - '**BabelCor**'.

- min. 3 sentences/sense (for precision);
- max. 10 sentences/sense (for time saving);
- **excluded** WikiSenses recognised as **Named Entities** - assumed they are kept the same across languages (didn't account for transliterations);

NEs simple heuristic: titles which contained at least two tokens starting with an uppercase letter were NEs - 94% on a validation sample of 100 pages. (e.g. William Shakespeare)

2. Translating Babel synsets



Applied SOA SMT, and identified top-scored translations as fit lexical entries for Babel Synsets. (Google Translate)

- monosemous senses were translated contextless;

As a result, **translated 324,137 WikiSenses** (reduced from over 3 million).

* image obtained from [1]

NANYANG TECHNOLOGICAL UNIVERSITY

3. Harvesting semantic relations

- All lexical and semantic relations from WordNet (+ gloss relations) are inherited by BabelNet
- All hyperlink relations from Wikipedia are collected and assigned an unspecified semantic relation ε.
  - including relations from other languages

Weighted edges:

- WN edges based on the Dice coefficient $\left(\frac{2\times|S\cap S'|}{|S|+|S'|}\right)$
  - overlap between synonyms + gloss's content words
- WikiPages uses a co-occurrence based method also applied to a Dice coefficient $\left(\frac{2\times f_{w,w'}}{f_w+f_{w'}}\right)$
  - co-occurrence context of 40 words, by the total number of hyperlinks

NANYANG TECHNOLOGICAL UNIVERSITY

In Vitro Evaluation - Mapping

- Gold Standard set of 1000 WikiPages hand linked to WNSenses (w/ inter-annotator agreement of 0.9);
- Evaluation by replicating the BoW and the graph-based methods to estimate mapping probabilities;
- Explored different disambiguation contexts for WN;
- Disambiguation context for Wikipages is ‘everything’ they have

NANYANG TECHNOLOGICAL UNIVERSITY

In Vitro Evaluation - Mapping

- Varied the maximum depth of graph search (exploring bigger portions of WordNet when building the disambiguation graphs)

- Baselines: Most Frequent Sense + Random Sense

- Removed the 100 most frequent linked sense labels Wikipages from the WordNet-Wikipedia intersection (avoid mapping WordNet senses to pages belonging to domains which are typically found in Wikipedia only - as ‘bands’, ‘movies’, etc.);

NANYANG TECHNOLOGICAL UNIVERSITY

In Vitro Evaluation - Mapping

	Mapping method		P	R	F_1	A
BoW	taxonomic		**89.7**	47.8	62.3	72.6
	gloss		87.6	51.8	65.1	74.0
	taxonomic + gloss		87.5	65.6	75.0	80.9
Graph	taxonomic relations					
	max depth	@ 2	87.2	60.8	71.6	77.9
		@ 3	81.6	65.0	72.4	78.7
		@ 4	78.3	69.5	73.6	79.4
	gloss relations					
	max depth	@ 2	80.5	60.6	69.1	77.0
		@ 3	77.5	65.2	70.9	78.2
		@ 4	72.4	67.1	69.6	78.0
	taxonomic + gloss relations					
	max depth	@ 2	81.2	74.6	**77.7**	**82.7**
		@ 3	72.8	**77.4**	75.1	80.1
		@ 4	64.3	76.2	69.8	75.0
	MFS baseline		25.4	49.2	33.5	25.4
	Random baseline		24.2	46.9	31.9	24.2

- MFS and Random baselines are virtually the same thing!
- Richer disambiguation context helps,
- Graph based methods give a much higher recall
- Depth > 3 seems to hurt F_1 (noisy gloss-derived relations)

* table obtained from [1]

NANYANG TECHNOLOGICAL UNIVERSITY

In Vitro Evaluation - Translation

Number of WordNet synsets and senses translated using SemCor and/or Wikipedia, as monosemous words, and their union.

	SemCor	Wikipedia	SemCor ∪ Wikipedia	monosemous	all
# synsets	3,901	31,308	33,359	62,259	68,554
# senses	6,852	35,372	40,504	101,853	115,606

On source of translation:

- Translations came from **multiple sources**;
- **SemCor was too small** to provide a substantial number of translations; (remember that a min. of 3 sentences was required)
- Combined translations from all sources (+ large contribution of sourceless monosemous translations);
- Translate a substantial portion of WordNet:
  - **83.4% of** 82,115 nominal **synsets**
  - **79.0% of** 146,312 nominal **senses**

* table obtained from [1]

NANYANG TECHNOLOGICAL UNIVERSITY

In Vitro Evaluation - Translation

Coverage against gold-standard Wordnets:

- **5** golden standard Wordnets; (Spanish, Catalan, French, German, Italian)
- All linked to PWN which is contained in BabelNet - so it's easy to quantify their overlap;
- **Synset Coverage** is measured by the % of synsets of the gold-standard WN that shares at least one term with BabelNet;
- **Word Coverage** is measured by the ration of word senses in the gold-standard WN that overlap with each Babel Synset;
- **Extra Synset Coverage** and **Extra Word Coverage** measures the PWN synsets and word ratio that are not covered by gold-standard WNs, but that are covered by BabelNet;

NANYANG TECHNOLOGICAL UNIVERSITY

In Vitro Evaluation - Translation

Size of the gold-standard wordnets.

	Catalan	French	German	Italian	Spanish
Word senses	64,171	44,265	15,762	57,255	83,114
Synsets	40,466	31,742	9,877	32,156	55,365

Coverage against gold-standard wordnets (percentages).

	WordCov (Senses)				SynsetCov (Synsets)			
Resource	Wiki		WordNet	BabelNet	Wiki		WordNet	BabelNet
Method	Links	Transl.	Transl.	All	Links	Transl.	Transl.	All
Catalan	20.3	46.9	25.0	**64.0**	25.2	54.1	29.6	**73.3**
French	70.0	69.6	16.3	**86.0**	72.4	79.6	19.4	**92.9**
German	39.6	42.6	21.0	**57.6**	50.7	58.2	28.6	**73.4**
Italian	28.1	39.9	19.7	**52.9**	40.0	58.0	28.7	**73.7**
Spanish	34.4	47.9	25.2	**66.4**	40.7	56.1	30.0	**76.6**

Extra coverage against gold-standard wordnets (percentages).

	WordExtraCov (Senses)				SynsetExtraCov (Synsets)			
Resource	Wiki		WordNet	BabelNet	Wiki		WordNet	D..BabelNet
Method	Links	Transl.	Transl.	All	Links	Transl.	Transl.	All
Catalan	100	204	71	**340**	35	105	42	**142**
French	255	223	92	**514**	63	102	67	**159**
German	1349	940	367	**2298**	506	668	303	**902**
Italian	160	234	83	**419**	87	153	68	**213**
Spanish	214	158	56	**384**	48	74	30	**102**

* tables obtained from [1]

NANYANG TECHNOLOGICAL UNIVERSITY

In Vitro Evaluation - Translation

On the precision of the extra coverage:

- Manual validation of 3000 random Babel Synsets:
  (check whether lexical entries fit the Babel Synset glosses translations, SemCor translations and wiki sentences translations)
  - 600 synsets x 5 languages (exc. English);
  - 200 synsets present only in PWN;
  - 200 synsets derived from Wikipedia only;
  - 200 synsets derived from intersection of PWN and Wiki;

Precision of BabelNet on synonyms in WordNet (WN), Wikipedia (Wiki) and their intersection (WN ∩ Wiki): percentage and total number of words (in parentheses) are reported.

Resource	Catalan	French	German	Italian	Spanish
WordNet	75.58 (258)	67.16 (268)	73.76 (282)	72.32 (271)	69.45 (275)
Wiki	92.71 (398)	96.44 (758)	97.74 (709)	99.09 (552)	92.46 (703)
WordNet ∩ Wiki	82.98 (517)	77.43 (709)	78.37 (777)	80.83 (574)	78.53 (643)

* table obtained from [1]

Original BabelNet Stats

Number of monosemous and polysemous words by part of speech (verbs, adjectives and adverbs are the same as in WordNet 3.0).

POS	Monosemous words	Polysemous words
Noun	22,763,265	1,134,857
Verb	6,277	5,252
Adjective	1,503	4,976
Adverb	3,748	733
Total	22,789,793	1,145,818

Number of lemmas, synsets and word senses in the 6 languages currently covered by BabelNet.

Language	Lemmas	Synsets	Word senses
English	5,938,324	3,032,406	6,550,579
Catalan	3,518,079	2,214,781	3,777,700
French	3,754,079	2,285,458	4,091,456
German	3,602,447	2,270,159	3,910,485
Italian	3,498,948	2,268,188	3,773,384
Spanish	3,623,734	2,252,632	3,941,039
Total	23,935,611	3,032,406	26,044,643

* tables obtained from [1]

NANYANG TECHNOLOGICAL UNIVERSITY

Original BabelNet Stats

Number of lexico-semantic relations harvested from WordNet, WordNet glosses and the 6 wikipedias.

	English	Catalan	French	German	Italian	Spanish	Total
WordNet	364,552	–	–	–	–	–	364,552
WordNet glosses	617,785	–	–	–	–	–	617,785
Wikipedia	50,104,884	978,006	5,613,873	5,940,612	3,602,395	3,411,612	69,651,382
Total	51,087,221	978,006	5,613,873	5,940,612	3,602,395	3,411,612	70,633,719

Glosses for the Babel synset referring to the concept of play as 'dramatic work'.

English	WordNet	A dramatic work intended for performance by actors on a stage.
	Wikipedia	A play is a form of literature written by a playwright, usually consisting of scripted dialogue between characters, intended for theatrical performance rather than just reading.
Catalan		El drama en termes generals és una obra literària o una situació de la vida real que resulta complexa i difícil però amb un final favorable o feliç.
French		Le drame (du latin drama, emprunté au grec ancien δρᾶμα/*drâma*, qui signifie action (théâtrale), pièce de théâtre) désigne étymologiquement toute action scénique.
German		Drama (altgriechisch δρᾶμα dráma 'Handlung') ist ein Oberbegriff für Texte mit verteilten Rollen.
Italian		Un dramma, dal greco "drama" (azione, storia; da δρᾶν, *fare*), è una forma letteraria che include parti scritte per essere interpretate da attori.
Spanish		Drama (del griego δρᾶμα, hacer o actuar) es la forma de presentación de acciones a través de su representación por actores.

+ BabelCor - Sense-tagged corpus with almost 2 million sentences (46,155 from SemCor and 1,940,402 from Wikipedia) - 330,993 annotated senses

* tables obtained from [1]

NANYANG TECHNOLOGICAL UNIVERSITY

New BabelNet 2.5

Integrates data from:

- WordNet3.0
- Wikipedia, Wiktionary, Wikidata
- OmegaWiki
- **Open Multilingual WordNet ☺**
- DBpedia

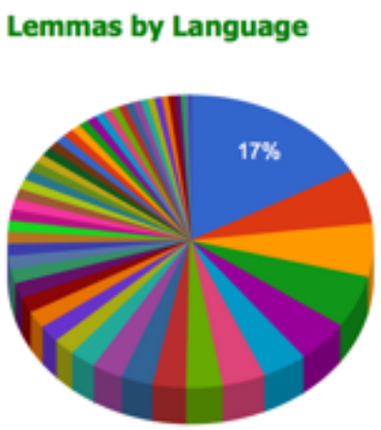
Now also includes:

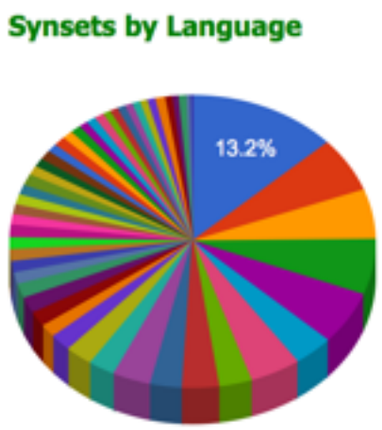
- **Translations for all open-class POS;**
- **Links to Categories;**
- **Images;**
- **etc.;**

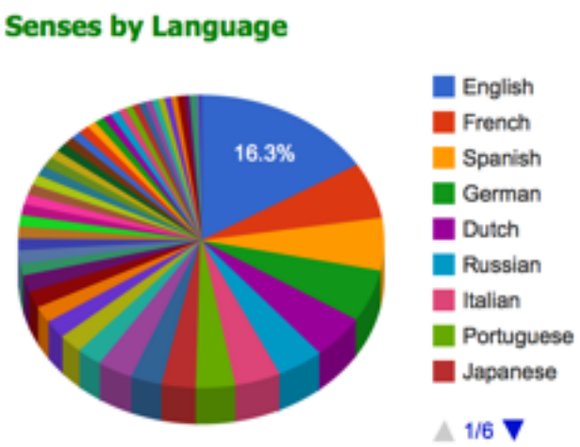
NANYANG TECHNOLOGICAL UNIVERSITY

New BabelNet 2.5 Stats

Number of languages:	50
Total number of Babel synsets:	9,348,287
Total number of Babel senses:	67,873,191
Total number of concepts:	3,684,512
Total number of Named Entities:	5,663,775
Total number of lexico-semantic relations:	262,687,848
Total number of glosses (textual definitions):	21,771,854
Total number of images:	7,764,270
Total number of RDF triples:	1,138,337,378







* data and charts obtained from http://babelnet.org/stats

Let’s go online for a second…

We’ll continue shortly after with…

Extrinsic Evaluation

SemEval-2007 - Task#16

SemEval-2007 T#16 - Evaluating wide-coverage Knowledge Resources (KBEval)

Knowledge bases were assessed by first generating so-called topic signatures + monolingual WSD;

Task:

1. Given a concept, generate a topic signature (e.g. word vector)
2. Unsupervised monolingual WSD:
   - given a word in context, compute the topic signature for each word in context,
   - compute a simple overlap score (with test sentence),
   - word sense with max score is selected;

NANYANG TECHNOLOGICAL UNIVERSITY

SemEval-2007 - Task#16

Test data:

- Two sets from previous Senseval and SemEval tasks;
- Sense annotated with PWN senses;

BabelNet-1 & BabelNet-2 Systems

1. Collect all synsets where word appears as synonym in a WN ‘Babel enriched’ synset;
2. Topic signature is all the english lexicalizations reachable by a distance of 1 and 2 (respectively);
3. Output the PWN synset associated with the winning BabelNet Synset;

NANYANG TECHNOLOGICAL UNIVERSITY

SemEval-2007 - Task#16

Results on the SemEval-2007 task 16: Evaluation of wide coverage knowledge resources.

(a) Senseval-3 English Lexical Sample task:

Knowledge base	P	R	F_1	Avg. size
TRAIN	65.1	65.1	65.1	450
TRAIN-MFS	54.5	54.5	54.5	–
WN-MFS	53.0	53.0	53.0	–
SEMCOR-MFS	49.0	49.1	49.0	–
TSSEM	52.5	52.4	52.4	103
BabelNet-1	44.3	44.3	44.3	119
BabelNet-2	35.0	35.0	35.0	2,128
KnowNet-20	44.1	44.1	44.1	610
RANDOM	19.1	19.1	19.1	–

(b) SemEval-2007 English Lexical Sample (task 17):

Knowledge base	P	R	F_1	Avg. size
TRAIN	87.6	87.6	87.6	450
TRAIN-MFS	81.2	81.2	81.2	–
WN-MFS	66.2	59.9	62.9	–
SEMCOR-MFS	42.4	38.4	40.3	–
WN + XWN + KN-20	53.0	53.0	53.0	627
BabelNet-1	52.2	46.3	49.1	130
BabelNet-2	56.9	53.1	54.9	2, 352
KnowNet-20	49.5	46.1	47.7	561
RANDOM	19.1	19.1	19.1	–

F1 = 49.9 and 43.3 on SemEval-2007 for WordNet and Wikipedia-only relations at distance 2, respectively

* tables obtained from [1]

NANYANG TECHNOLOGICAL UNIVERSITY

SemEval-2007 - Task#7

SemEval-2007 T#7 - Coarse-grained all-words WSD task

Is granularity of WN senses an obstacle for WSD?

A coarse sense inventory is obtained semi-automatically by clustering WN senses via a mapping to the Oxford Dictionary of English;

Task:

1. ~6,000 words ‘coarse + grained’ sense tagged,
2. Participants have access to a lemma and a POS for each content word;
3. They have to output a coarse sense (optionally a cluster replaces the fine-grained sense choice) for each word.

NANYANG TECHNOLOGICAL UNIVERSITY

SemEval-2007 - Task#7

BabelNet in SemEval-2007 T#7:

Hypothesis: the meanings of Wikipages are intuitively coarser than those in WordNet, so it should be better at coarser WSD.

1. Edge filtering: filtered paths connecting different senses of the same word + removed edges from the graph whose weight is below a certain threshold;
2. Defined a general framework for transforming an input context into a graph; (same as for estimating mapping probabilities)
3. Applied 4 algorithms for graph-based lexico-semantic disambiguation;
4. Best results with MFS is assigned when no sense assignment it attempted; (weakly supervised)

NANYANG TECHNOLOGICAL UNIVERSITY

SemEval-2007 - Task#7

Performance on SemEval-2007 coarse-grained all-words WSD with MFS as a back-off strategy when no sense assignment is attempted. The differences between the results in bold in each column of the table are not statistically significant at $p < 0.05$ based on a χ^2 test.

Resource	Algorithm	Nouns only P/R/F_1	All words P/R/F_1
WordNet	Degree	80.1	79.7
	PLength	80.3	79.8
	SProbability	79.5	79.3
	PageRank	79.7	79.4
BabelNet	Degree	**84.7**	**82.3**
	PLength	**85.4**	**82.7**
	SProbability	**84.6**	**82.1**
	PageRank	82.1	80.1
	SUSSX-FR	81.1	77.0
	TreeMatch	N/A	73.6
	NUS-PT	82.3	**82.5**
	SSI	**84.1**	**83.2**
	MFS BL	77.4	78.9
	Random BL	63.5	62.7

* Best results were found for a maximum depth of 3, and a minimum edge weight of 0.01

BabelNet beats the MFS baseline on nouns (a notably difficult competitor for unsupervised and knowledge-rich systems) - even without back-off strategy.

* table obtained from [1]

SemEval-2010 - Task#3

SemEval-2010 T#3: Cross-lingual WSD

In this task, lexical disambiguation is operationalized as a word translation task;

Task:

1. Given a predefined predefined sense inventory in a MLCorpus (where all necessary sense distinctions are available for every language);
2. 20 target words x 50 sentences each (1,000 test instances), for each word in context, participants disambiguate the target word by translating it into a different language;
3. The meaning preservingness of the translations are evaluated from a list of weighted/ranked, gold-standard translations;

NANYANG TECHNOLOGICAL UNIVERSITY

SemEval-2010 - Task#3

BabelNet in SemEval-2010 T#3:

Same ‘turn input into graph’ framework from an input + context (max. depth of 3) + 4 algorithms for monolingual WSD;

- **Standard Setting:** return MFTranslation from wining synset (as ordered by frequency of alignment in Europarl); back-off to MFAlignment in Europarl - if no sense assigned

For better insight: remove from selected synset lexicalizations not seen in the gold-standard corpus; then MFTranslation; back-off to MFAlignment in Europarl; (+Oracle Transl.)

Upper bounds:

- BabelNet: return all gold-standard translations found in the known BabelSS of the test instance;
- Task: since evaluation metrics are not in %, this gives the highest ranking translations chosen by humans;

NANYANG TECHNOLOGICAL UNIVERSITY

SemEval-2010 - Task#3

Results on the SemEval-2010 task 3: Cross-lingual Word Sense Disambiguation.

	French		German		Italian		Spanish	
	P	R	P	R	P	R	P	R
Baseline	21.25	21.25	13.16	13.16	15.18	15.18	19.74	19.74
UvT-v	–	–	–	–	–	–	23.39	23.39
UvT-g	–	–	–	–	–	–	19.83	19.64
T3-COLEUR	21.97	21.75	13.18	13.05	14.82	14.67	19.83	19.64
Degree	22.94	22.94	17.15	17.15	18.03	18.03	22.48	22.48
+ ORACLE TRANSLATIONS	25.82	25.82	20.16	20.16	21.13	21.13	25.26	25.26
PLength	23.42	23.42	17.72	17.72	18.19	18.19	22.76	22.76
+ ORACLE TRANSLATIONS	25.87	25.87	20.42	20.42	21.47	21.47	25.76	25.76
SProbability	23.27	23.27	17.61	17.61	18.14	18.14	22.69	22.69
+ ORACLE TRANSLATIONS	25.85	25.85	20.50	20.50	21.74	21.74	25.48	25.48
PageRank	22.62	22.62	16.98	16.98	16.76	16.76	21.11	21.11
+ ORACLE TRANSLATIONS	26.00	26.00	20.85	20.85	21.71	21.71	26.19	26.19
BabelNet upper bound	30.21	30.21	25.39	25.39	27.67	27.67	30.73	30.73
Task upper bound	39.44	100.00	34.36	100.00	40.00	100.00	39.54	100.00

Precision: number of correct system translations over the total number of translations returned by the system, each weighted by their score. (higher ranked would receive higher points)

Recall: number of correct translations given by the system **over the total number of items in the test** set, each weighted by their score.

* tables obtained from [1]

SemEval-2013 - Task#12

Multilingual WSD using BabelNet Senses

Task Preparation:

1. 13 articles were selected from WSMT, all existed in 4 langs. (English, French, German and Spanish) + translation in Italian;
2. Due to the automatic integration, all the mappings of 8306 synsets (for 978 lemmas appearing in the corpus), were manually checked (delete, add mapping or merge);
3. Manual correction of POS, NE and MWE tagging;
4. Manual sense annotation for English, projected across to other languages;
5. Evaluated for precision and recall on BabelNet, Wordnet, (against WN MFS) and Wikipedia senses (against pseudo MFS for WikiSenses - WN frequency or lexical ordering).

NANYANG TECHNOLOGICAL UNIVERSITY

SemEval-2013 - Task#12

Cross Language Sense Projection:

1. English dataset was manually annotated; (1+ senses allowed)
2. Other datasets were sentence aligned, and lemmas compared to that lang's lexical entries in used english senses;
   - if a match occurred, that english sense would be projected;
   - labelled 50%-70% of non-english datasets;
3. Manually completed, corrected and later reviewed;
   - only 22-37% needed correction; (simple but efficient)

Sense projection statistics

Language	Projected instances	Valid projections	Invalid projections
French	1016	791	225
German	592	373	219
Italian	1029	774	255
Spanish	911	669	242

* table obtained from [2]

SemEval-2013 - Task#12

Statistics for sense annotated data

Language	Instances	Single-words	Multiword expressions	Named Entities	Mean senses per instance	Mean senses per lemma
			BabelNet			
English	1931	1604	127	200	1.02	1.09
French	1656	1389	89	176	1.05	1.15
German	1467	1267	21	176	1.00	1.05
Italian	1706	1454	211	41	1.22	1.27
Spanish	1481	1103	129	249	1.15	1.19
			Wikipedia			
English	1242	945	102	195	1.15	1.16
French	1039	790	72	175	1.18	1.14
German	1156	957	21	176	1.07	1.08
Italian	977	869	85	41	1.20	1.18
Spanish	1103	758	107	248	1.11	1.10
			WordNet			
English	1644	1502	85	57	1.01	1.10

* table obtained from [2]

NANYANG TECHNOLOGICAL UNIVERSITY

SemEval-2013 - Task#12

Participants:

- 7 systems participated; (6 for BabelNet; 4 to WN; 3 to Wiki)
- All of them used graph-based approaches for WSD;
  - **DAEBAK!** (1 BabelNet) - ±5 sentence window around the target word; sense selection based on measuring connectivity to the synsets of neighboring lemmas; MFS as back off;
  - **GETALP** (2x BabelNet, 1 WN) - all based on the ant-colony algorithm (tuned differently); BN1 optimizes from the trial data; BN2 and WN1 are completely unsupervised;
  - **UMCC-DLSI** (3 x BabelNet, 1 WN, 1 Wiki) - all based on ISR-WN (resource that enriches WN semantic network from multiple lexical resources); WSD performed by an extension of the Personalized PageRank; RUN-1 uses all noun instances in the sentence as context, RUN-2 all noun instances in the document, and RUN-3 all words in the sentence;

NANYANG TECHNOLOGICAL UNIVERSITY

SemEval-2013 - Task#12

F1 score per language on BabelNet senses

Team	System	English	French	German	Italian	Spanish
DAEBAK!	PD	0.604	0.538	0.591	0.613	0.600
GETALP	BN-1	0.263	0.261	0.404	0.324	-
GETALP	BN-2	0.266	0.257	0.400	0.324	0.371
UMCC-DLSI	RUN-1	0.677	**0.605**	0.618	0.657	0.705
UMCC-DLSI	RUN-2	**0.685**	**0.605**	**0.621**	**0.658**	**0.710**
UMCC-DLSI	RUN-3	0.680	-	-	-	-
MFS		0.665	0.453	0.674	0.575	0.645

Systems' performance on Wordnet senses (English Only)

Team	System	Precision	Recall	F1
GETALP	WN-1	0.406	0.406	0.406
UMCC-DLSI	RUN-1	0.639	0.635	0.637
UMCC-DLSI	RUN-2	**0.649**	**0.645**	**0.647**
UMCC-DLSI	RUN-3	0.642	0.639	0.640
MFS		0.630	0.630	0.630

* tables obtained from [2]

NANYANG TECHNOLOGICAL UNIVERSITY

SemEval-2013 - Task#12

F1 score per instance type, averaged across all languages

Team	System	Single term	Multiword expression	Named Entity
DAEBAK!	PD	0.502	0.801	**0.910**
GETALP	BN-1	0.232	0.724	0.677
GETALP	BN-2	0.235	0.740	0.656
UMCC-DLSI	RUN-1	0.582	0.806	0.865
UMCC-DLSI	RUN-2	**0.584**	**0.809**	0.864
MFS		0.511	0.853	0.920

* table obtained from [2]

NANYANG TECHNOLOGICAL UNIVERSITY

SemEval-2013 - Task#12

- No system used cross lingual WSD - did not used bitext nor multilingual structure of BabelNet;
- The task organisers tweaked the submitted systems to check the utility of simple multilingual sense analysis;
  - Sense assignments were only kept iff at least two other aligned sentences have the same sense assigned to some word;

Post Hoc: scores for outputs where at least two other aligned sentences (cross lingual) were used

		English			French			German			Italian			Spanish		
Team	System	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1
DAEBAK	PD	0.769	0.364	0.494	0.747	0.387	0.510	0.762	0.307	0.438	0.778	0.425	0.550	0.778	0.450	0.570
GETALP	BN-2	0.793	0.111	0.195	0.623	0.130	0.215	0.679	0.124	0.210	0.647	0.141	0.231	0.688	0.177	0.282
UMCC-DLSI	RUN-1	0.787	0.421	0.549	0.754	0.441	0.557	0.741	0.330	0.457	0.796	0.461	0.584	0.830	0.525	0.643
UMCC-DLSI	RUN-2	0.791	0.419	0.548	0.760	0.436	0.554	0.746	0.332	0.460	0.799	0.453	0.578	0.837	0.530	0.649

* tables obtained from [2]

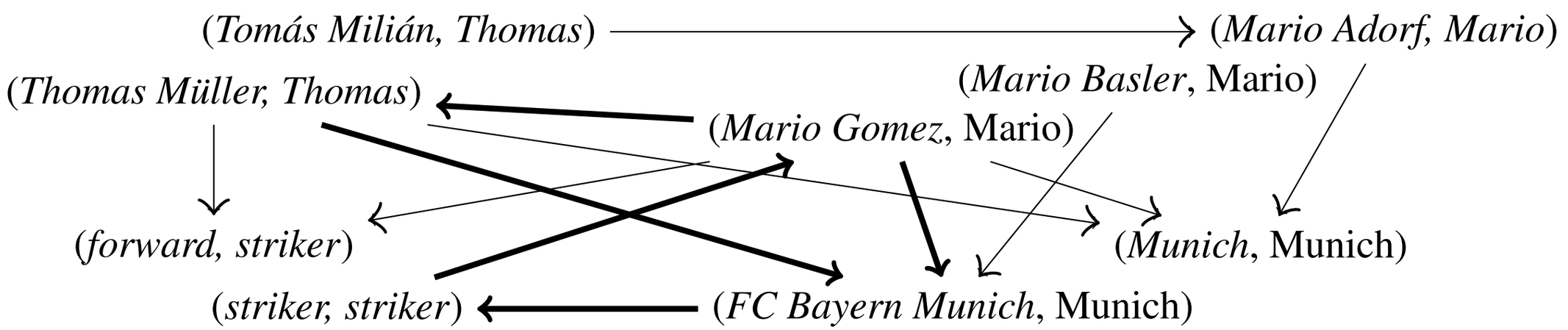
NANYANG TECHNOLOGICAL UNIVERSITY

Babelfy

"Word sense disambiguation and entity linking together!"

Babelfy

Relating Concepts and Named Entities in WSD



An excerpt of the semantic interpretation graph automatically built for the sentence *Thomas and Mario are strikers playing in Munich* (the edges connecting the correct meanings are in bold).

* graph obtained from [3]

NANYANG TECHNOLOGICAL UNIVERSITY

- [EN] The queen of England was eating a delicious hot dog while she was waiting for her driver by the bank.
- [PT] A rainha de Inglaterra estava a comer um delicioso cachorro quente enquanto esperava pelo seu motorista junto ao banco.
- [IT] La regina di Inghilterra stava mangiando un hot dog delizioso mentre aspettava il suo autista alla banca.
- [CH1] 英国女王在银行等待她司机的时候吃了一个美味的热狗。
- [CH2] 当英格兰女王在银行等她的司机的时候，她正在吃着一个美味的热狗。
- [JP] イギリスの女王様が銀行の手前で運転手を待っていながら美味しいホットドッグを食べていた。
- [KO] 영국 여왕은 은행 옆에서 운전수를 기다리고 있을 때 맛있는 핫도그를 먹고 있었다.

Some considerations

- BabelNet is ‘grand’ but a bit noisy;
- The WN disambiguation context for mapping Wiki to WN is too weak… (did they mentioned everything?)
- And from a cognitive perspective, weighted relations for WSD also seem the right choice for me; but also hierarchical (if a a threshold is reached by stronger relations, disregard weaker)
- Should we have the gloss relations explicit in the OMW?
  - and could we ask ILI to try to provide WSD’ted definitions?
- This would be even more interesting with a stronger cross lingual disambiguation when mapping WikiSenses to WNSenses; (it may have happened in V2.5)

NANYANG TECHNOLOGICAL UNIVERSITY

References

- [1] R. Navigli and S. Ponzetto. BabelNet: The Automatic Construction, Evaluation and Application of a Wide-Coverage Multilingual Semantic Network. Artificial Intelligence, 193, Elsevier, 2012, pp. 217-250.
- [2] R. Navigli, D. A. Jurgens, D. Vannella. SemEval-2013 Task 12: Multilingual Word Sense Disambiguation. Proc. of 7th International Workshop on Semantic Evaluation (SemEval), in the Second Joint Conference on Lexical and Computational Semantics (*SEM 2013), Atlanta, USA, June 14-15th, 2013, pp. 222-231.
- [3] Andrea Moro, Alessandro Raganato and Roberto Navigli. Entity Linking meets Word Sense Disambiguation: A Unified Approach. Transactions of the Association for Computational Linguistics (TACL), 2, 2014.

NANYANG TECHNOLOGICAL UNIVERSITY