Semantic Representations of Concepts and Entities and their Applications

Jose Camacho-Collados

Linguistic Computing Laboratory

SAPIENZA UNIVERSITÀ DI ROMA

University of Cambridge, 20 April 2017

Outline

- Background: Vector Space Models

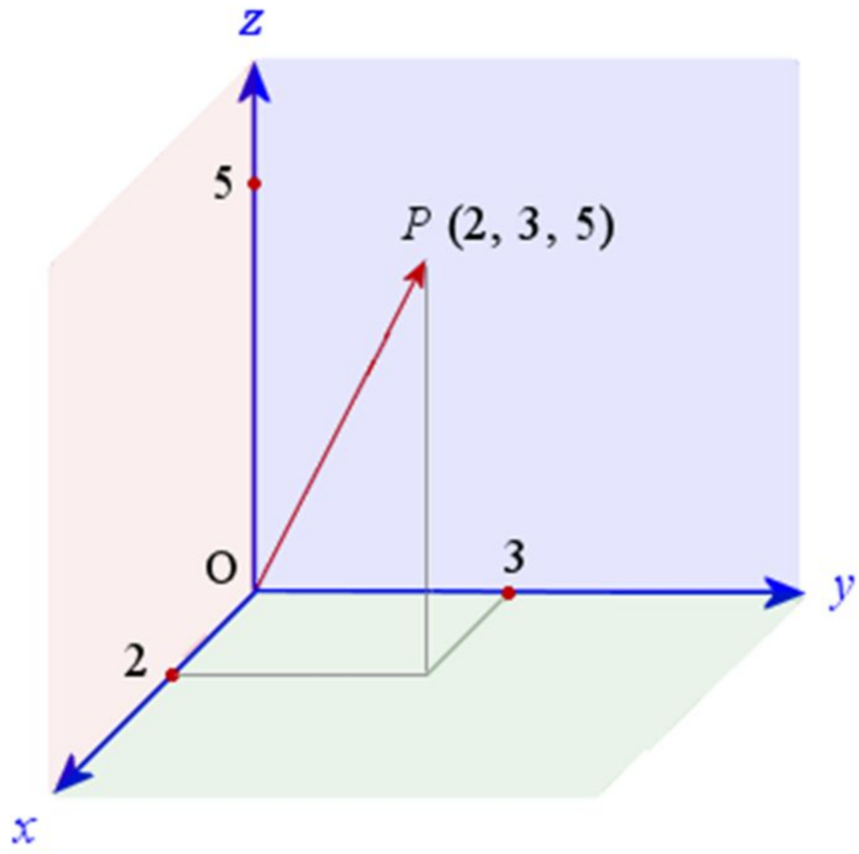
- Semantic representations for Senses, Concepts and Entities -> NASARI

- Applications

- Conclusions

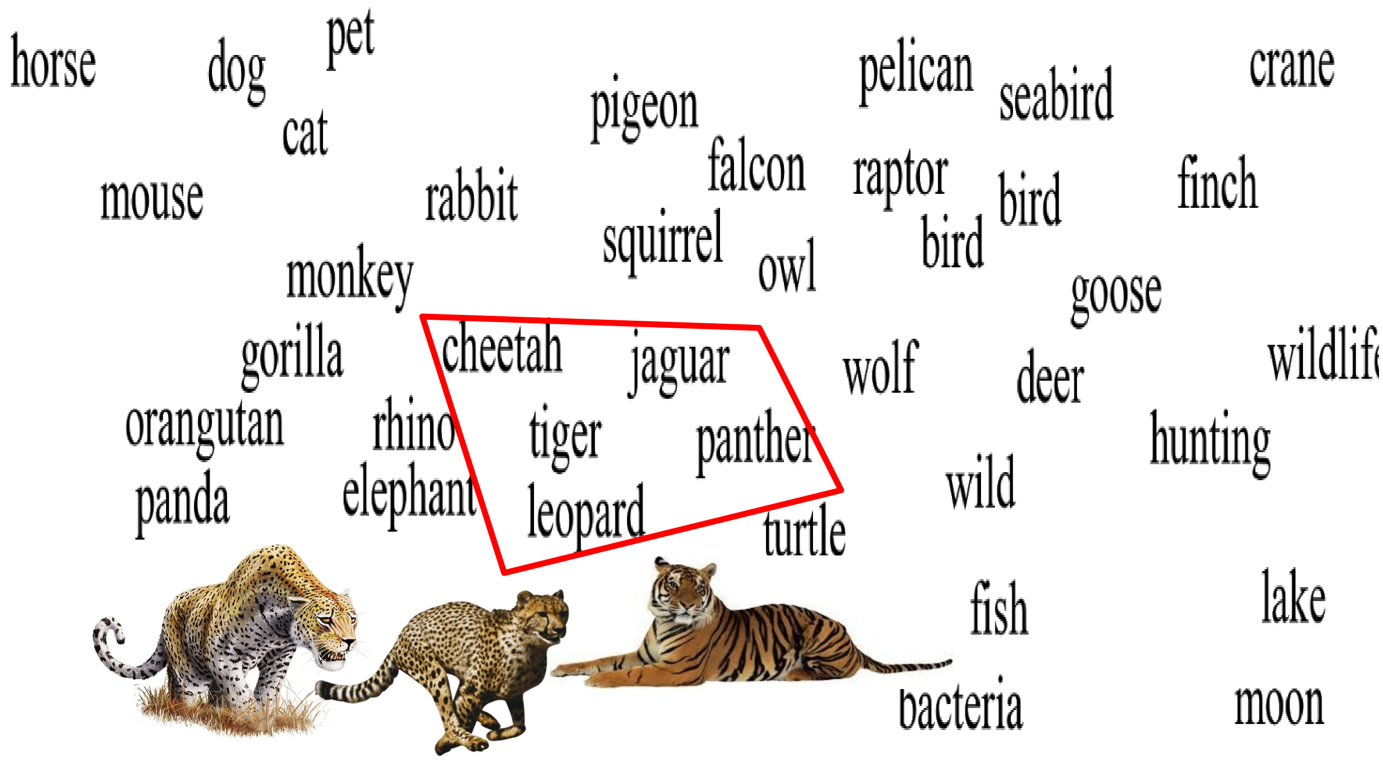
Vector Space Model

Turney and Pantel (2010): Survey on Vector Space Model of semantics

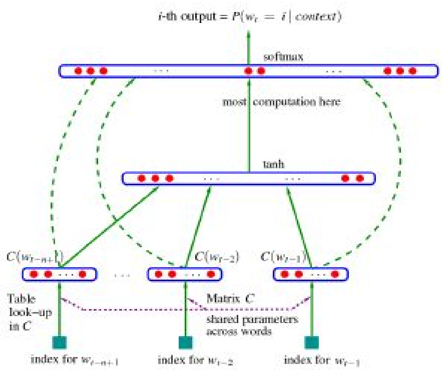


Word vector space models

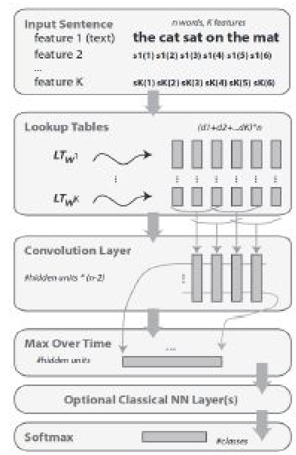
Words are represented as vectors: semantically similar words are close in the vector space



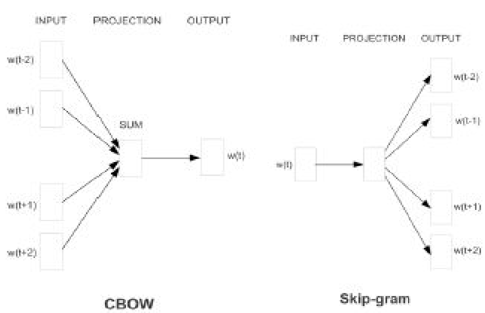
Neural networks for learning word vector representations from text corpora -> **word embeddings**



Bengio et al. (2003)



Collobert & Weston (2008)



Mikolov et al. (2013)

Probability and Ratio	$k = solid$	$k = gas$	$k = water$
$P(k\|ice)$	1.9×10^{-4}	6.6×10^{-5}	3.0×10^{-3}
$P(k\|steam)$	2.2×10^{-5}	7.8×10^{-4}	2.2×10^{-3}
$P(k\|ice)/P(k\|steam)$	8.9	8.5×10^{-2}	1.36

Pennington et al. (2014)

Why word embeddings?

Embedded vector representations:

- are compact and fast to compute
- preserve important relational information between words (actually, meanings):

$$king - man + woman \approx queen$$

- are geared towards general use

Applications for word representations

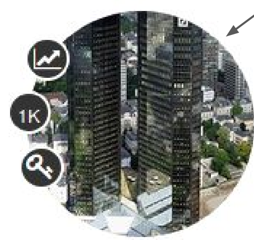
- Syntactic parsing (Weiss et al. 2015)
- Named Entity Recognition (Guo et al. 2014)
- Question Answering (Bordes et al. 2014)
- Machine Translation (Zou et al. 2013)
- Sentiment Analysis (Socher et al. 2013)

… and many more!

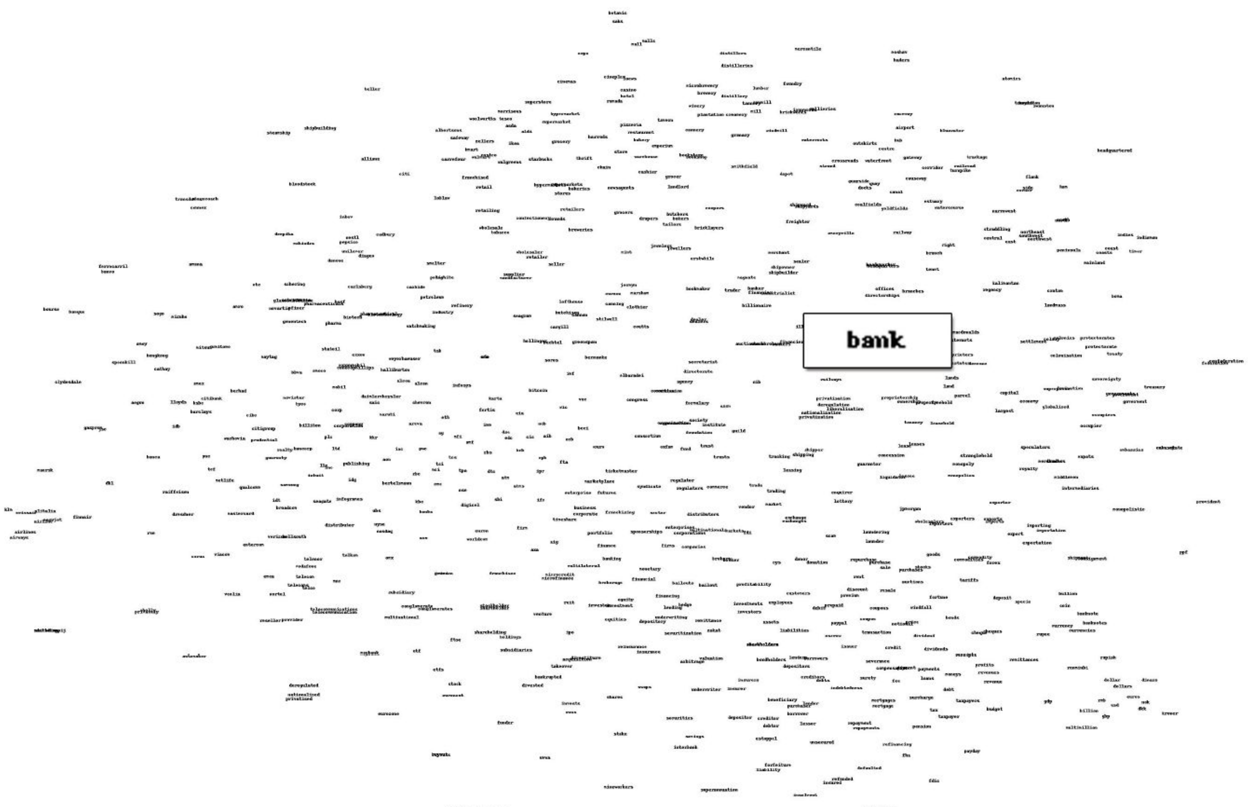
AI goal: language understanding

Limitations of word embeddings

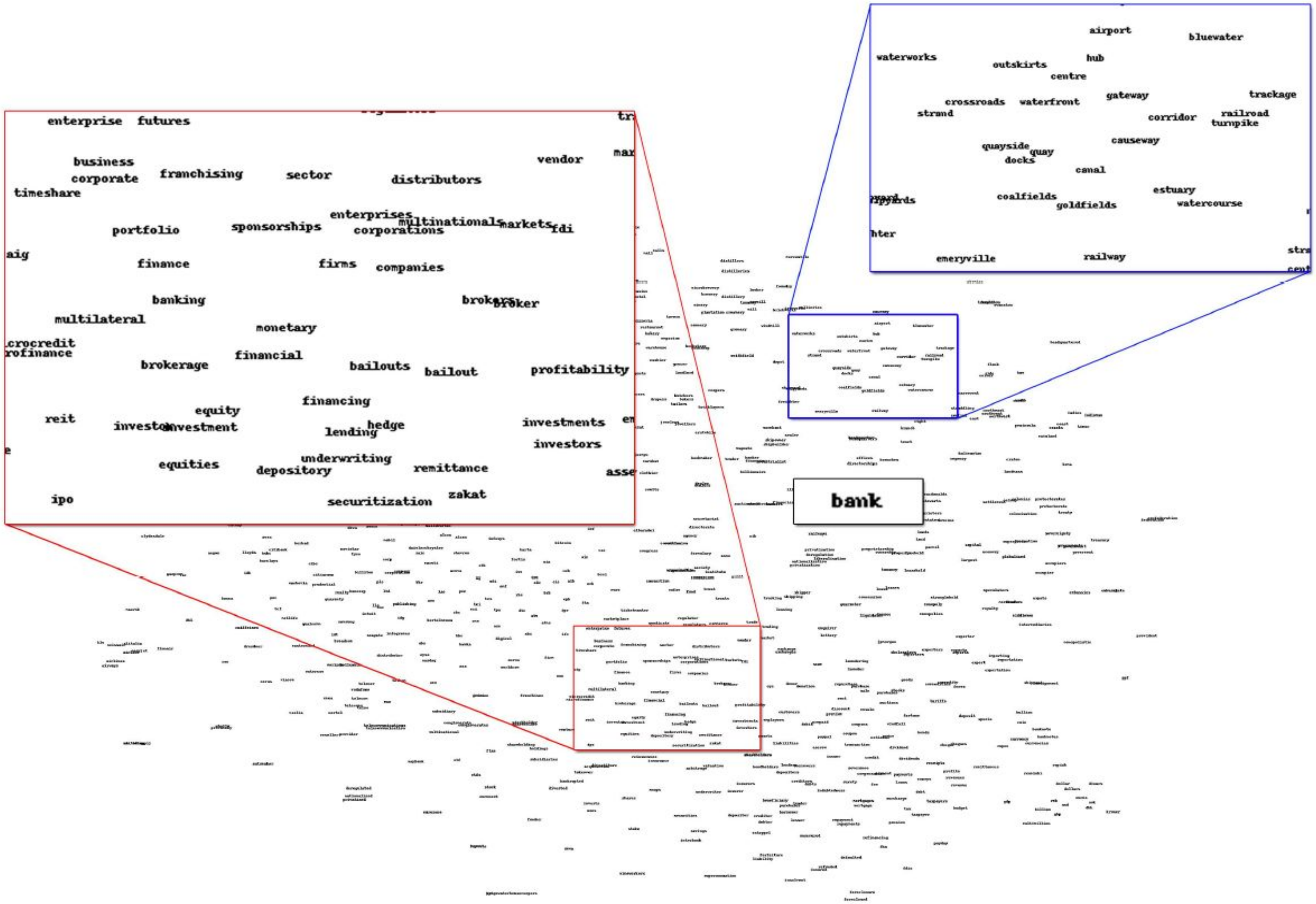
- Word representations cannot capture ambiguity. For instance, *bank*



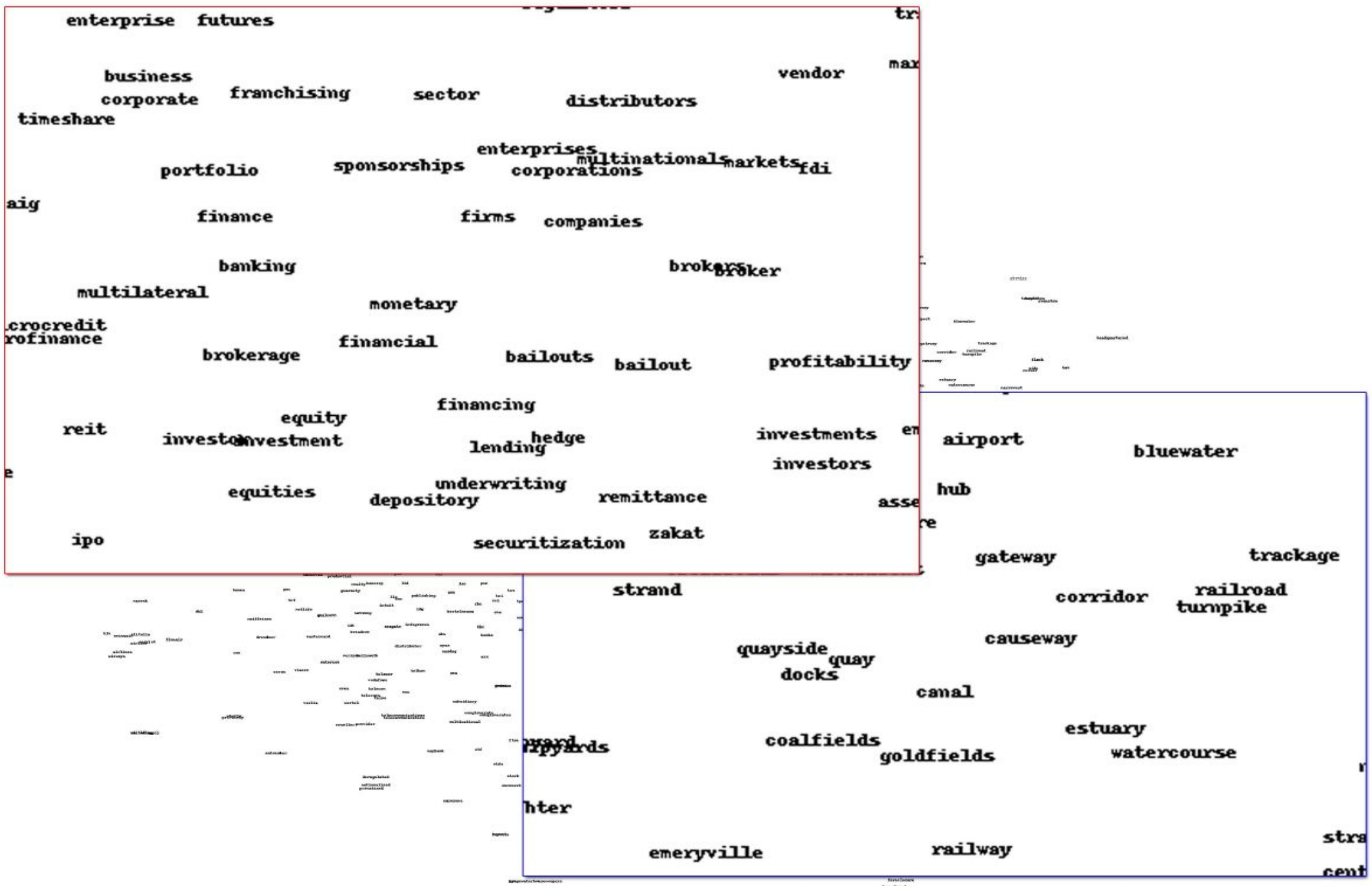
Problem 1:
word representations cannot capture ambiguity



Problem 1:
word representations cannot capture ambiguity



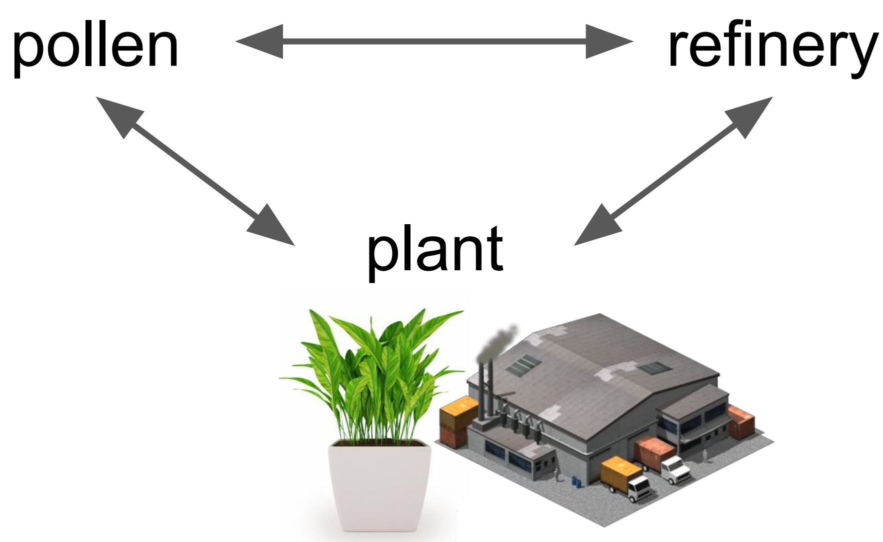
Problem 1: word representations cannot capture ambiguity



Word representations and the triangular inequality

Example from Neelakantan et al (2014)

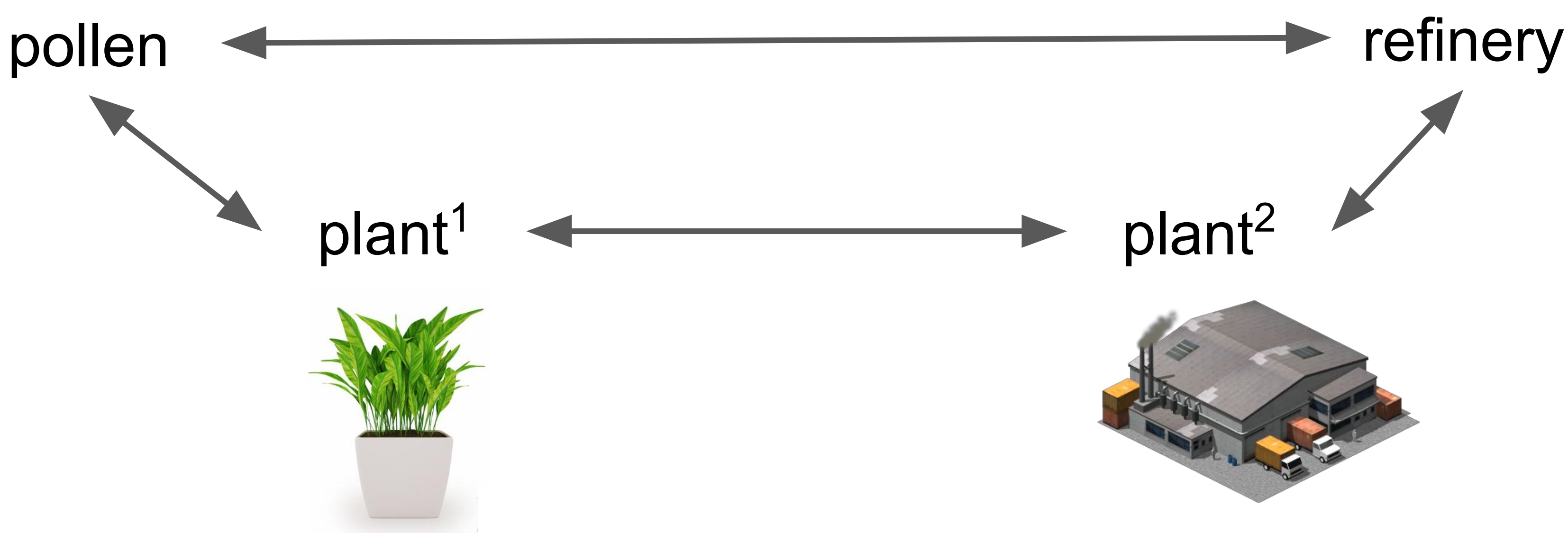
For distance d, $d(a,c) \leq d(a,b) + d(b,c)$.



Word representations and the triangular inequality

Example from Neelakantan et al (2014)

For distance d, $d(a, c) \leq d(a, b) + d(b, c)$.



Limitations of word representations

- They cannot capture ambiguity. For instance,

bank

-> They neglect rare senses and infrequent words

- Word representations do not exploit knowledge from existing lexical resources.

WIKIPEDIA
The Free Encyclopedia

BabelNet

NASARI

a Novel Approach to a Semantically-Aware Representations of Items

http://lcl.uniroma1.it/nasari/

NASARI semantic representations

- NASARI 1.0 (April 2015): *Lexical and unified vector representations for WordNet synsets and Wikipedia pages for English.*

José Camacho Collados, Mohammad Taher Pilehvar and Roberto Navigli. *NASARI: a Novel Approach to a Semantically-Aware Representation of Items.* **NAACL 2015**, Denver, USA, pp. 567-577.

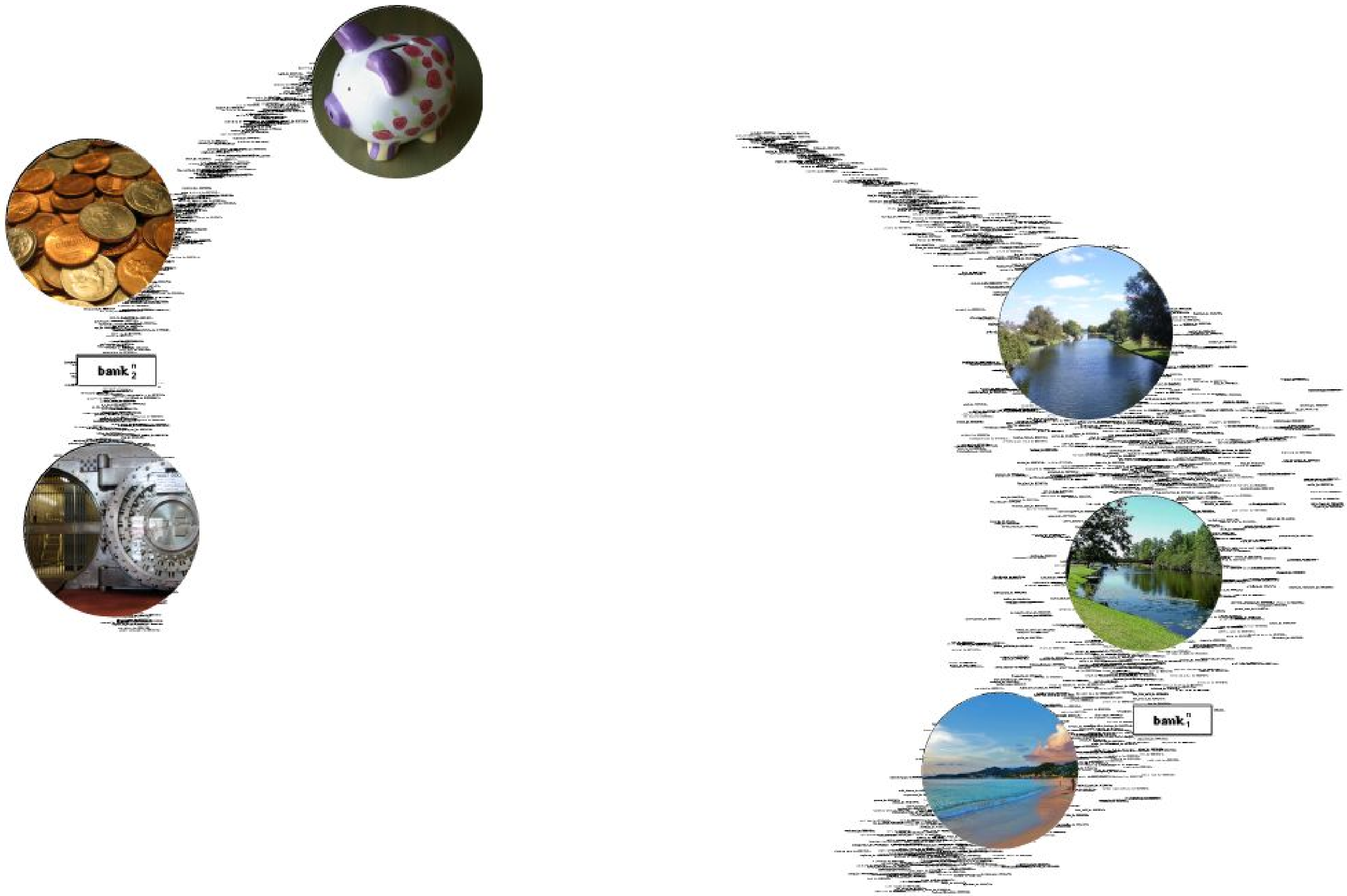
- NASARI 2.0 (August 2015): + Multilingual extension.

José Camacho Collados, Mohammad Taher Pilehvar and Roberto Navigli. *A Unified Multilingual Semantic Representation of Concepts.* **ACL 2015**, Beijing, China, pp. 741-751.

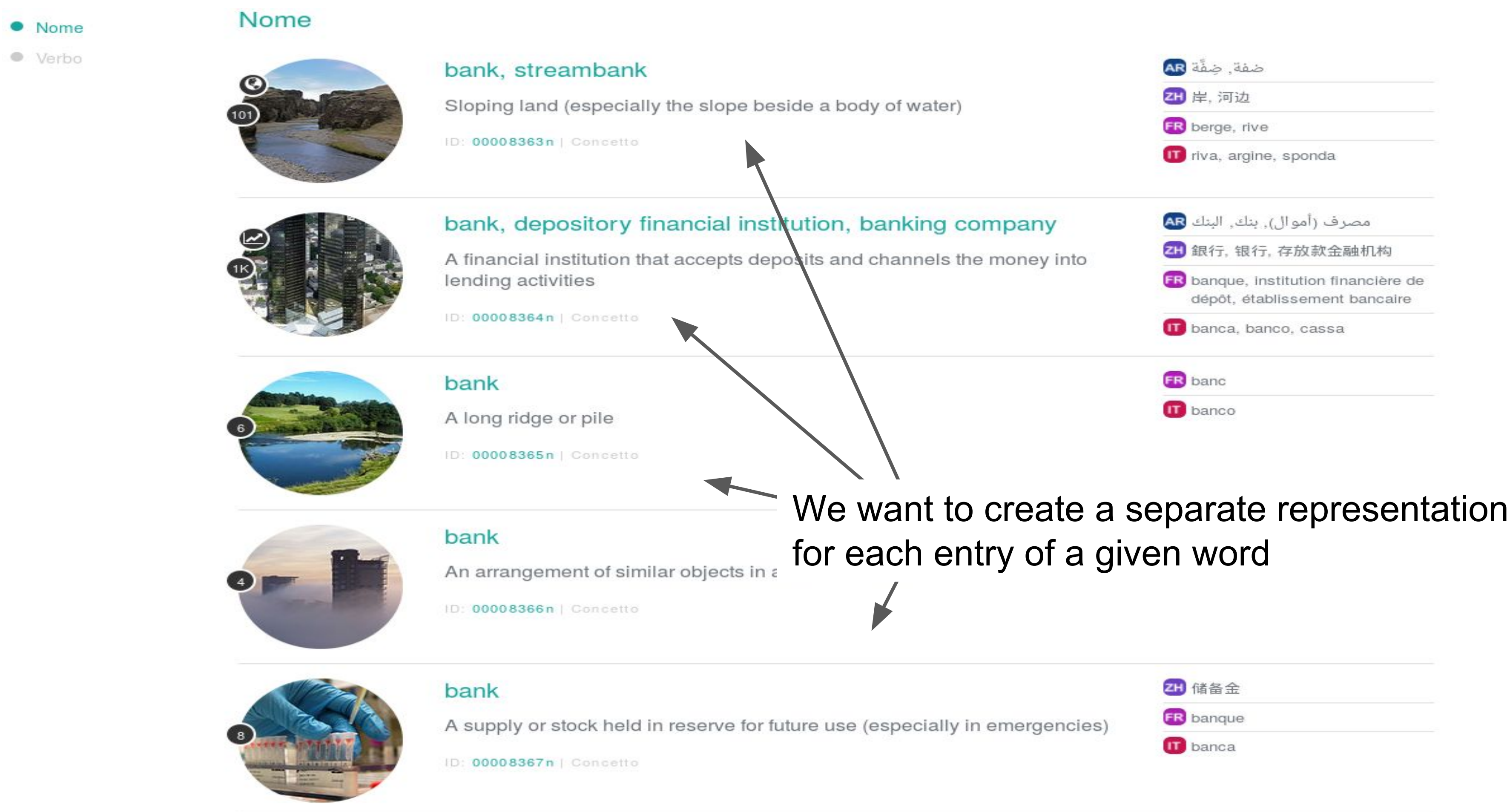
- NASARI 3.0 (March 2016): + Embedded representations, new applications.

José Camacho Collados, Mohammad Taher Pilehvar and Roberto Navigli. *Nasari: Integrating explicit knowledge and corpus statistics for a multilingual representation of concepts and entities.* **Artificial Intelligence Journal, 2016,** 240, 36-64.

Key goal: obtain sense representations



Key goal: obtain sense representations

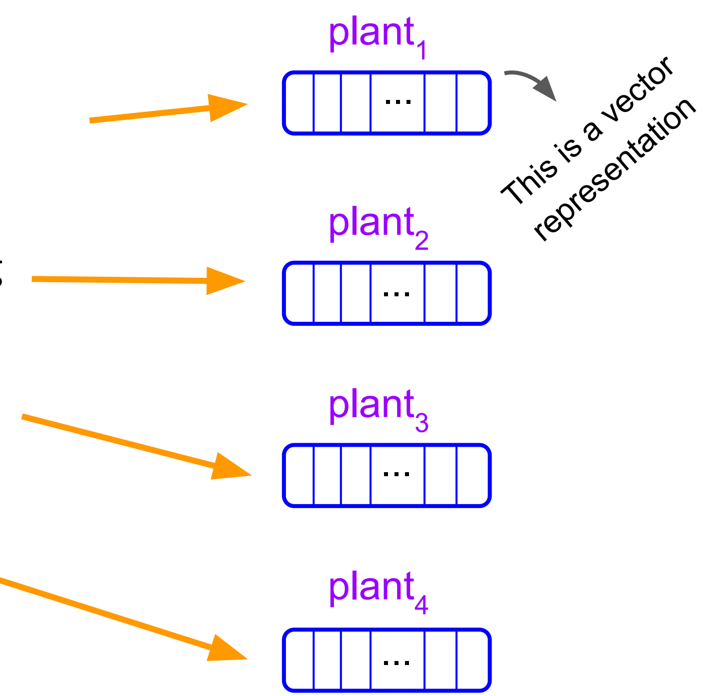


Knowledge-based Sense Representations

Represent word senses as defined by sense inventories

plant

- ***plant, works, industrial plant*** (buildings for carrying on industrial labor)
- ***plant, flora, plant life*** ((botany) a living organism lacking the power of locomotion)
- ***plant*** (an actor situated in the audience whose acting is rehearsed but seems spontaneous to the audience)
- ***plant*** (something planted secretly for discovery by another)



Idea

Encyclopedic knowledge

Lexicographic knowledge

WIKIPEDIA
The Free Encyclopedia

WordNet

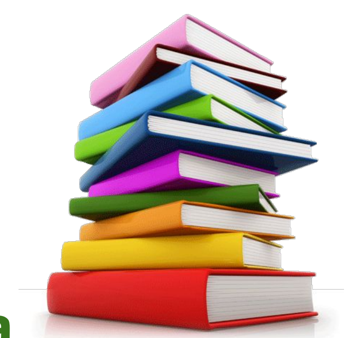
Idea

Encyclopedic knowledge

Wikipedia
The Free Encyclopedia

Lexicographic knowledge

WordNet

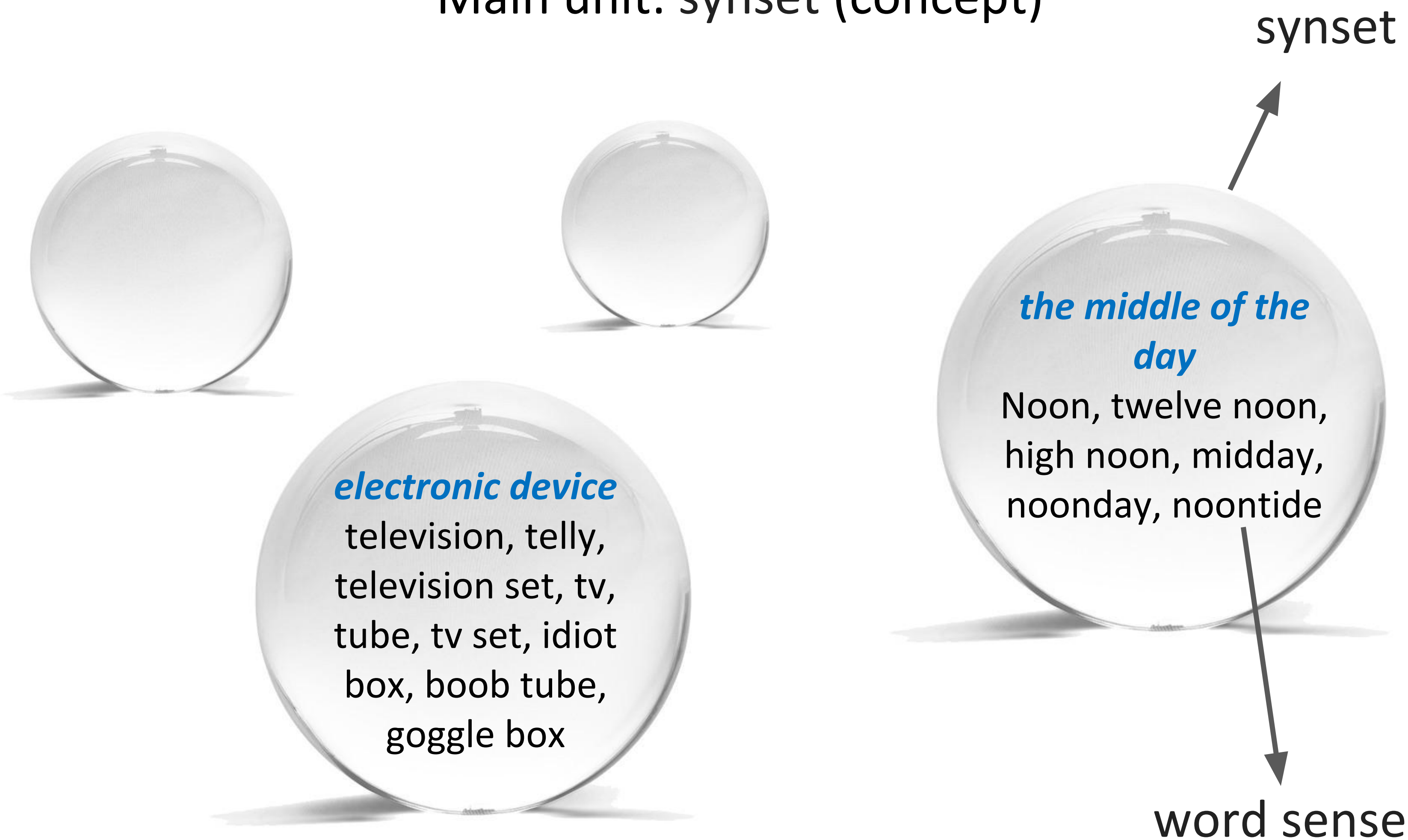


Information from text corpora

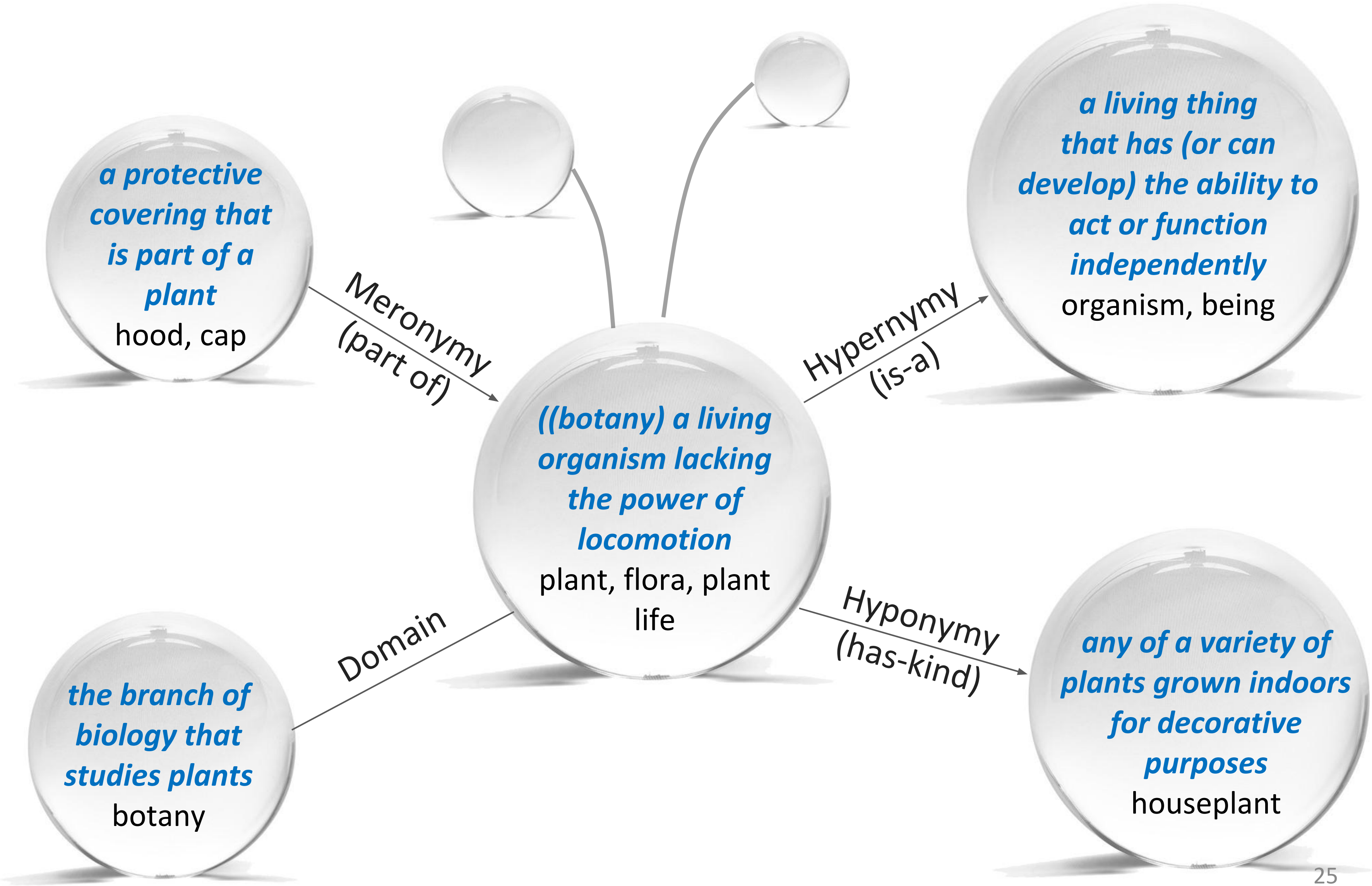
WordNet

WordNet

Main unit: synset (concept)



WordNet semantic relations



WordNet

WordNet Search - 3.1

- WordNet home page - Glossary - Help

Word to search for: plant [Search WordNet]

Display Options: (Select option to change) [Change]

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

Noun

- S: (n) **plant**, works, industrial plant (buildings for carrying on industrial labor) *"they built a large plant to manufacture automobiles"*
- S: (n) **plant**, flora, plant life ((botany) a living organism lacking the power of locomotion)
- S: (n) **plant** (an actor situated in the audience whose acting is rehearsed but seems spontaneous to the audience)
- S: (n) **plant** (something planted secretly for discovery by another) *"the police used a plant to trick the thieves"; "he claimed that the evidence against him was a plant"*

Verb

- S: (v) **plant**, set (put or set (seeds, seedlings, or plants) into the ground) *"Let's plant flowers in the garden"*
- S: (v) implant, engraft, embed, imbed, **plant** (fix or set securely or deeply) *"He planted a knee in the back of his opponent"; "The dentist implanted a tooth in the gum"*
- S: (v) establish, found, **plant**, constitute, institute (set up or lay the groundwork for) *"establish a new department"*
- S: (v) **plant** (place into a river) *"plant fish"*
- S: (v) **plant** (place something or someone in a certain position in order to secretly observe or deceive) *"Plant a spy in Moscow"; "plant bugs in the dissident's apartment"*
- S: (v) **plant**, implant (put firmly in the mind) *"Plant a thought in the students' minds"*

Link to online browser

Knowledge-based Sense Representations using WordNet

M. T. Pilehvar, D. Jurgens and R. Navigli: **Align, Disambiguate and Walk: A Unified Approach for Measuring Semantic Similarity** (ACL 2013)

X. Chen, Z. Liu, M. Sun: **A Unified Model for Word Sense Representation and Disambiguation** (EMNLP 2014)

S. Rothe and H. Schutze: **AutoExtend: Extending Word Embeddings to Embeddings for Synsets and Lexemes** (ACL 2015)

S. K. Jauhar, C. Dyer, E. Hovy: **Ontologically Grounded Multi-sense Representation Learning for Semantic Vector Space Models** (NAACL 2015)

M. T. Pilehvar and N. Collier: **De-Conflated Semantic Representations** (EMNLP 2016)

Wikipedia

WIKIPEDIA
The Free Encyclopedia

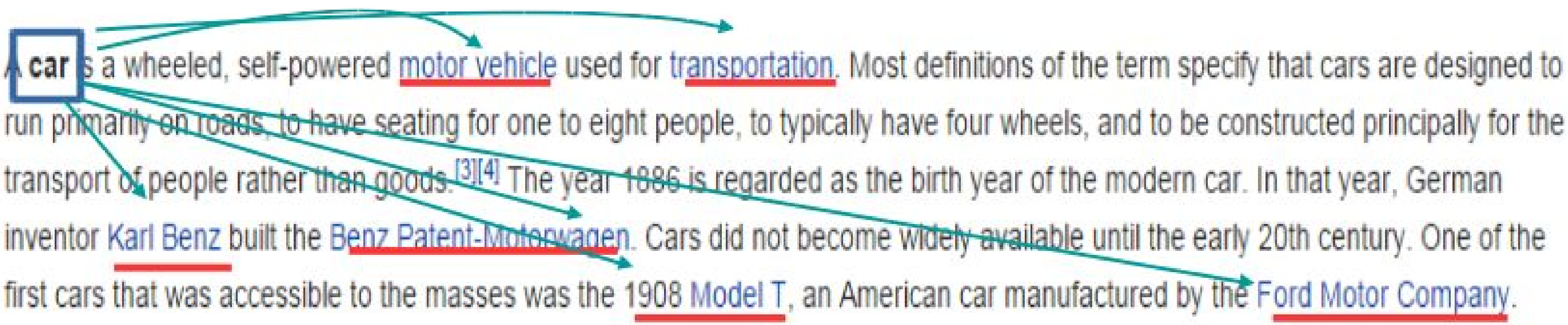
Wikipedia

High coverage of **named entities** and **specialized concepts** from different domains

Wikipedia hyperlinks

A **car** is a wheeled, self-powered motor vehicle used for transportation. Most definitions of the term specify that cars are designed to run primarily on roads, to have seating for one to eight people, to typically have four wheels, and to be constructed principally for the transport of people rather than goods.[3][4] The year 1886 is regarded as the birth year of the modern car. In that year, German inventor Karl Benz built the Benz Patent-Motorwagen. Cars did not become widely available until the early 20th century. One of the first cars that was accessible to the masses was the 1908 Model T, an American car manufactured by the Ford Motor Company.

Wikipedia hyperlinks



A **car** is a wheeled, self-powered motor vehicle used for transportation. Most definitions of the term specify that cars are designed to run primarily on roads, to have seating for one to eight people, to typically have four wheels, and to be constructed principally for the transport of people rather than goods.[3][4] The year 1886 is regarded as the birth year of the modern car. In that year, German inventor Karl Benz built the Benz Patent-Motorwagen. Cars did not become widely available until the early 20th century. One of the first cars that was accessible to the masses was the 1908 Model T, an American car manufactured by the Ford Motor Company.

BabelNet

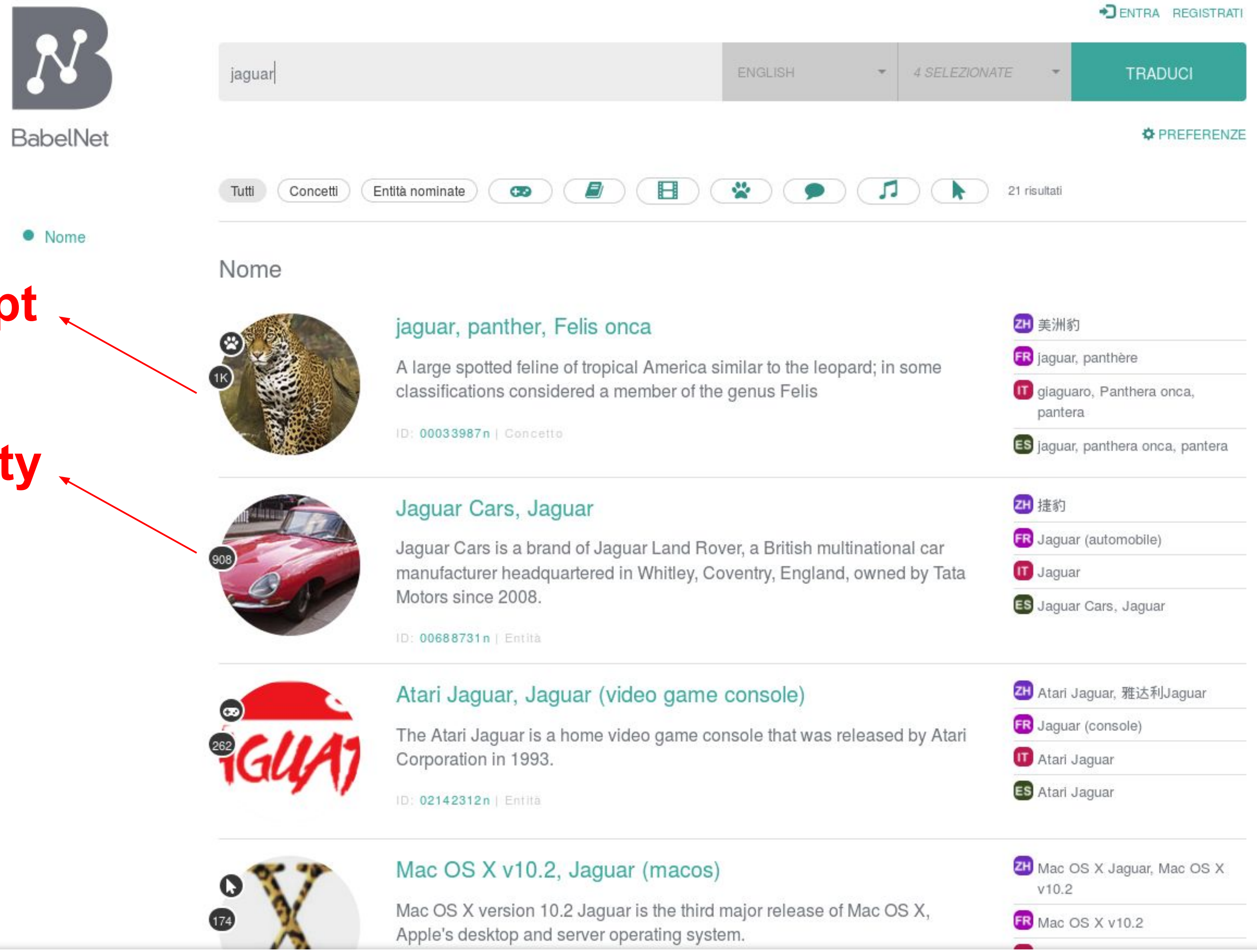
Thanks to an automatic mapping algorithm, **BabelNet integrates Wikipedia and WordNet**, among other resources (Wiktionary, OmegaWiki, WikiData…).

Key feature: **Multilinguality** (271 languages)

BabelNet

Concept

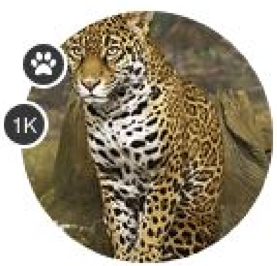
Entity



BabelNet

It follows the same structure of WordNet:
synsets are the main units

Nome



jaguar, panther, Felis onca

A large spotted feline of tropical America similar to the leopard; in some classifications considered a member of the genus Felis

ID: 00033987n | Concetto

ZH 美洲豹

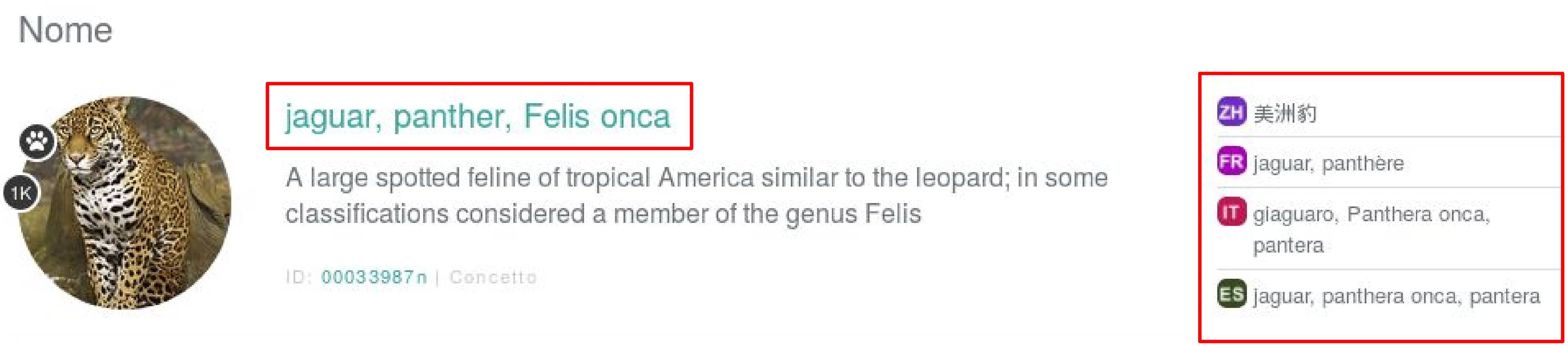
FR jaguar, panthère

IT giaguaro, Panthera onca, pantera

ES jaguar, panthera onca, pantera

BabelNet

In this case, **synsets are multilingual**



NASARI: Integrating Explicit Knowledge and Corpus Statistics for a Multilingual Representation of Concepts and Entities

(Camacho-Collados et al., AIJ 2016)

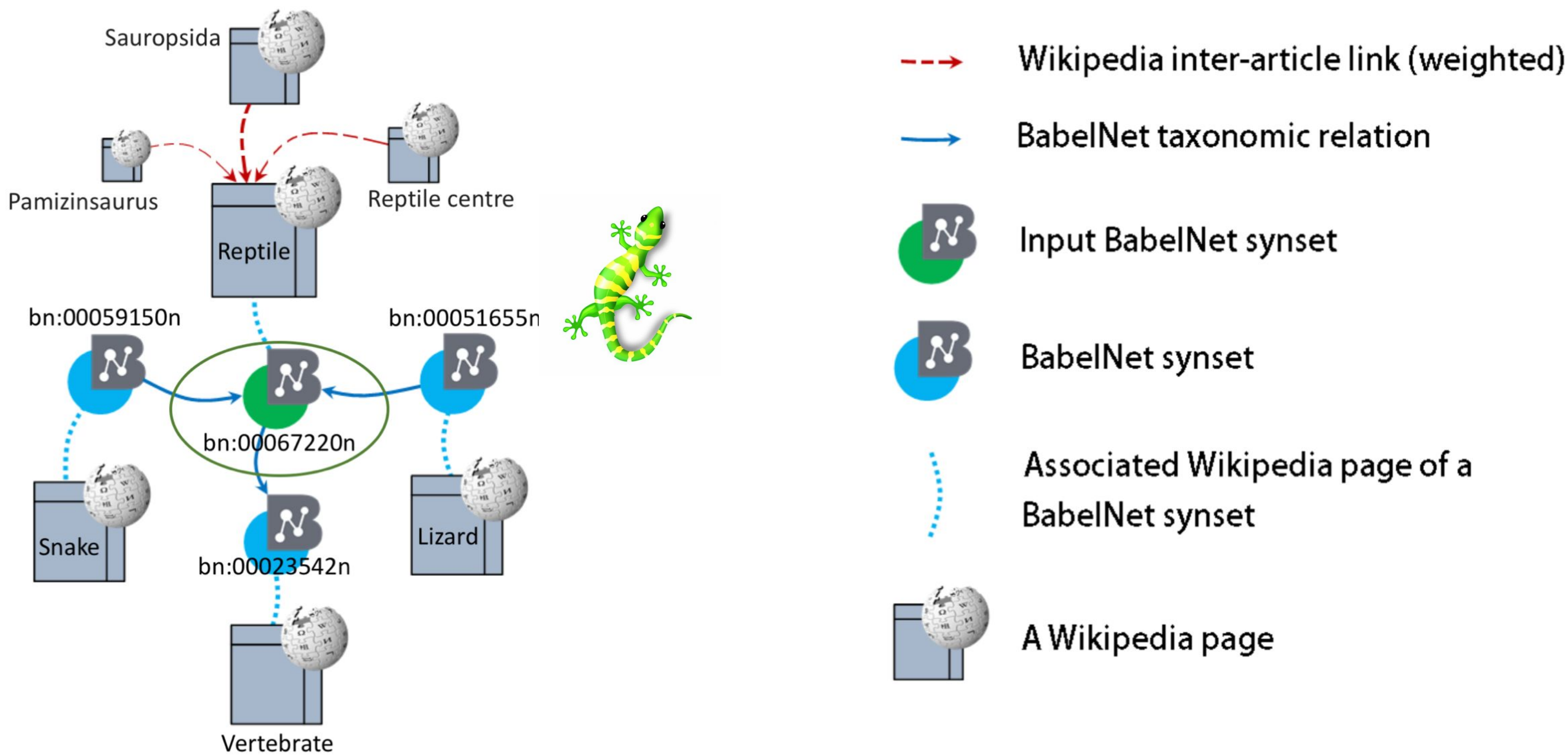
Goal

Build vector representations for multilingual BabelNet synsets.

How?

We exploit **Wikipedia semantic network** and **WordNet taxonomy** to construct a subcorpus (contextual information) for any given BabelNet synset.

Pipeline



Process of obtaining contextual information for a BabelNet synset exploiting BabelNet taxonomy and Wikipedia as a semantic network

Three types of vector representations

Three types of vector representations:

- **Lexical** (dimensions are words)

- **Unified** (dimensions are multilingual BabelNet synsets)

- **Embedded** (latent dimensions)

Three types of vector representations

Three types of vector representations:

- **Lexical** (dimensions are words): Dimensions are weighted via **lexical specificity,** a statistical measure based on the hypergeometric distribution.

- **Unified** (dimensions are multilingual BabelNet synsets)

- **Embedded** (latent dimensions)

Lexical specificity

It is a statistical measure based on the **hypergeometric distribution**, particularly suitable for term extraction tasks.

Thanks to its statistical nature, it is **less sensitive to corpus sizes** than the conventional *tf-idf* (in our setting, it **consistently outperforms *tf-idf* weighting**).

Three types of vector representations

Three types of vector representations:

- **Lexical** (dimensions are words):
- **Unified** (dimensions are multilingual BabelNet synsets): This representation uses a **hypernym-based clustering technique** and can be used in **cross-lingual** applications
- **Embedded** (latent dimensions)

Three types of vector representations

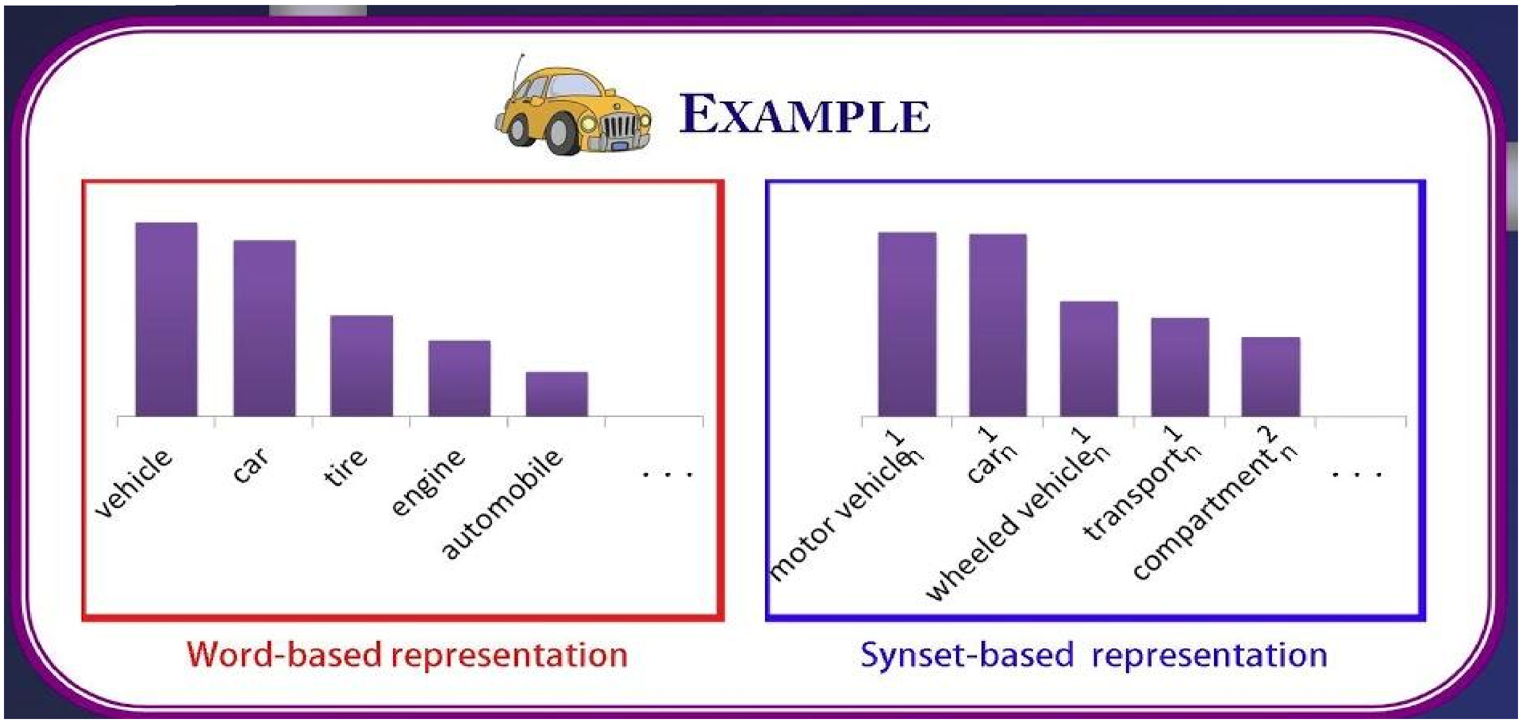
Three types of vector representations:

- **Lexical** (dimensions are words):

- **Unified** (dimensions are multilingual BabelNet synsets): This representation uses a **hypernym-based clustering technique** and can be used in **cross-lingual** applications

- **Embedded** (latent dimensions)

Lexical and unified vector representations



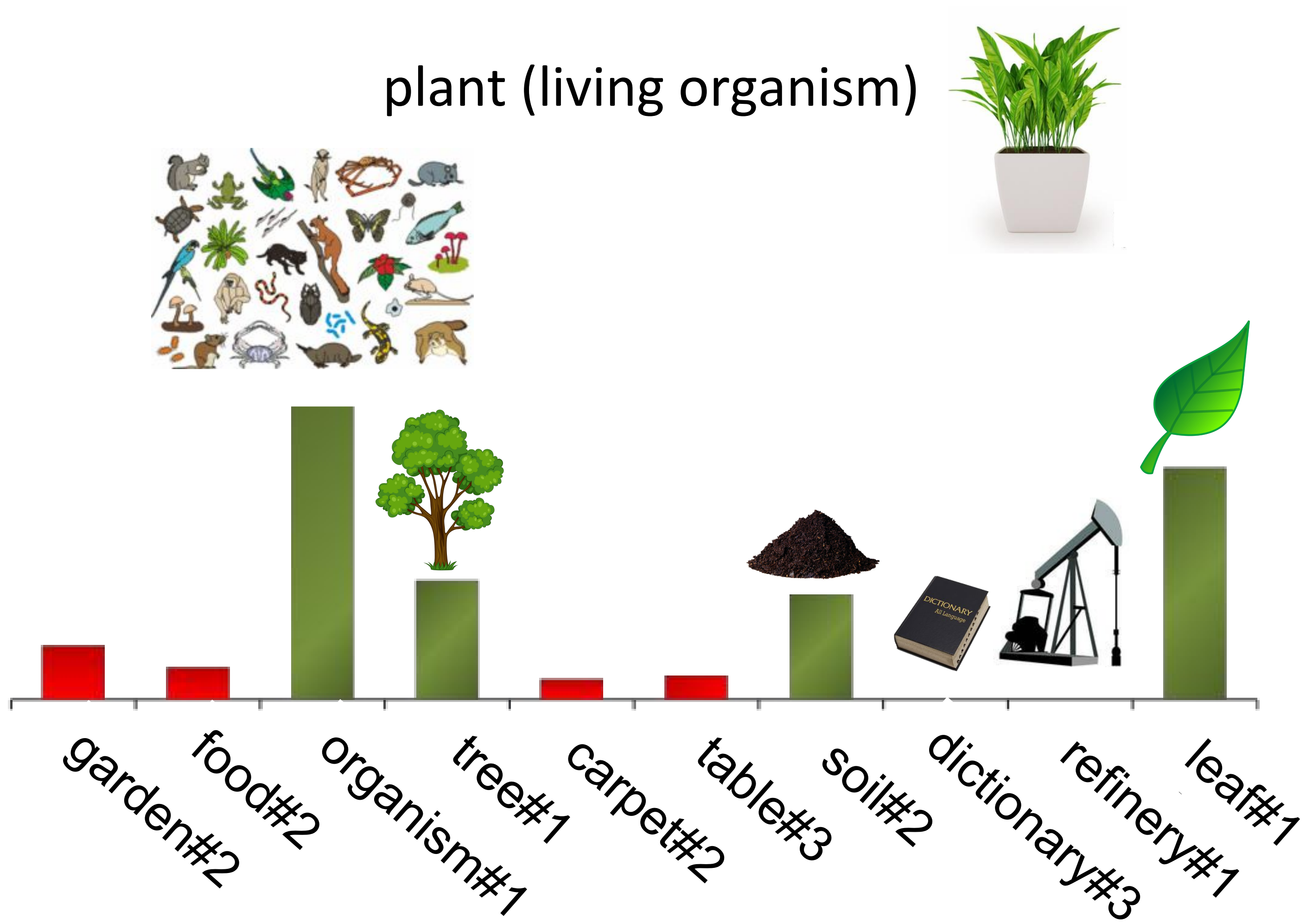
From a lexical vector to a unified vector

Lexical vector= (automobile, car, engine, vehicle, motorcycle, …)

motor_vehicle$^{1}_{n}$

Unified vector= (motor_vehicle$^{1}_{n}$, …)

Human-interpretable dimensions



Three types of vector representations

Three types of vector representations:

- **Lexical** (dimensions are words)
- **Unified** (dimensions are multilingual BabelNet synsets)

- **Embedded**: Low-dimensional vectors (latent) exploiting **word embeddings** obtained from **text corpora**. This representation is obtained by plugging word embeddings on the lexical vector representations.

Three types of vector representations

Three types of vector representations:

- **Lexical** (dimensions are words)
- **Unified** (dimensions are multilingual BabelNet synsets)

- **Embedded**: Low-dimensional vectors (latent) exploiting **word embeddings** obtained from **text corpora**. This representation is obtained by plugging word embeddings on the lexical vector representations.

Word and synset embeddings share the same vector space!

Sense-based Semantic Similarity

Based on the semantic similarity between senses.

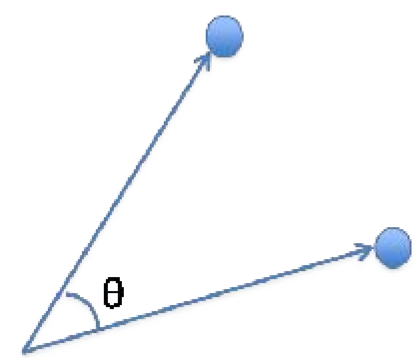
Two main measures:

- **Cosine similarity** for low-dimensional vectors
- **Weighted Overlap** for sparse high-dimensional vectors (interpretable)

Vector Comparison
Cosine Similarity

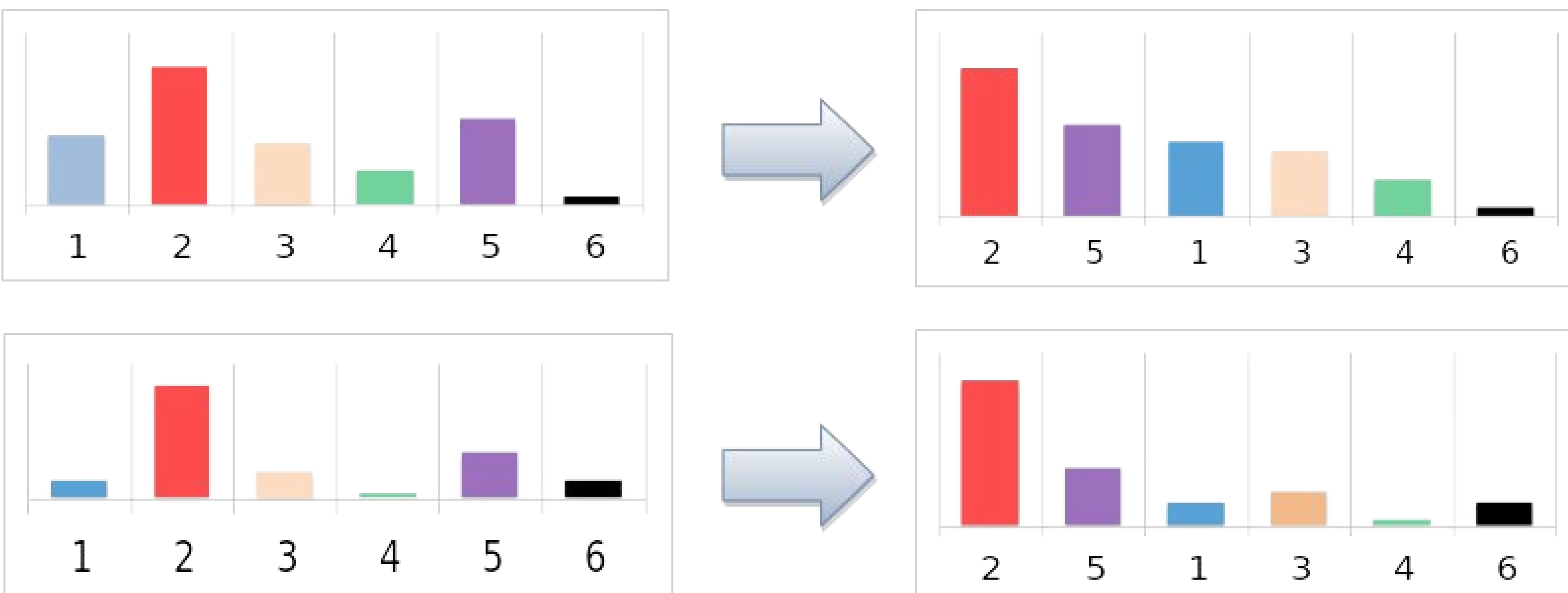
The most commonly used measure for the similarity of vector space model (sense) representations

$$sim(A,B) = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|}$$



Vector Comparison
Weighted Overlap

$$WO(v_1, v_2) = \frac{\sum_{q \in O} \left(rank(q, v_1) + rank(q, v_2) \right)^{-1}}{\sum_{i=1}^{|O|} (2i)^{-1}}$$



Embedded vector representation

Closest senses

Bank (financial institution)		Bank (geography)		*bank*	
Closest senses	Cosine	Closest senses	Cosine	Closest senses	Cosine
Deposit account	0.99	Stream bed	0.98	Bank (financial institution)	0.86
Universal bank	0.99	Current (stream)	0.97	Universal bank	0.86
British banking	0.98	River engineering	0.97	British banking	0.86
German banking	0.98	Braided river	0.97	German banking	0.85
Commercial bank	0.98	Fluvial terrace	0.97	Branch (banking)	0.85
Banking in Israel	0.98	Bar (river morphology)	0.97	McFadden Act	0.85
Financial institution	0.98	River	0.97	Four Northern Banks	0.84
Community bank	0.97	Perennial stream	0.96	State bank	0.84

NASARI semantic representations
Summary

- Three types of semantic representation: **lexical, unified and embedded.**

- **High coverage of concepts and named entities** in multiple languages (all Wikipedia pages covered).

NASARI semantic representations
Summary

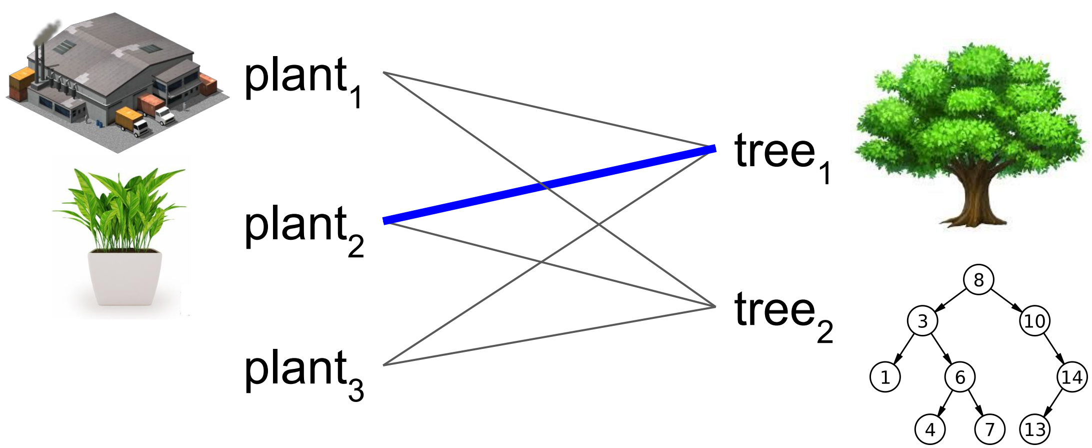
- Three types of semantic representation: **lexical, unified and embedded.**

- **High coverage of concepts and named entities** in multiple languages (all Wikipedia pages covered).

What's next? **Evaluation** and **use** of these semantic representations in **NLP applications**.

How are sense representations used for word similarity?

1- **MaxSim**: similarity between the most similar senses across two words



$$\text{MaxSim}(w, w') \stackrel{\text{def}}{=} \max_{1 \leq j \leq K, 1 \leq k \leq K'} d(\pi_k(w), \pi_j(w'))$$

Intrinsic evaluation

Monolingual semantic similarity (English)

	MC-30		WS-Sim		SimLex-999 (nouns)		Average	
	r	ρ	r	ρ	r	ρ	r	ρ
NASARI	0.89	0.78	0.74	0.72	0.50	0.49	**0.71**	0.67
$\text{NASARI}_{\text{lexical}}$	0.88	0.81	0.74	0.73	0.51	0.49	**0.71**	**0.68**
$\text{NASARI}_{\text{unified}}$	0.88	0.78	0.72	0.70	0.49	0.48	0.70	0.65
$\text{NASARI}_{\text{embed}}$	**0.91**	0.83	0.68	0.68	0.48	0.46	0.69	0.66
ESA	0.59	0.65	0.45	0.53	0.16	0.23	0.40	0.47
Lin	0.76	0.72	0.66	0.62	**0.58**	**0.58**	0.67	0,64
ADW	0.79	0.83	0.63	0.67	0.44	0.45	0.62	0.65
Chen	0.82	0.82	0.63	0.64	0.48	0.44	0.64	0.63
Word2Vec	0.80	0.80	**0.76**	**0.77**	0.46	0.45	0.67	0.67
Best-Word2Vec	$0.83^{\ddagger}$	$0.83^{\ddagger}$	$0.76^{\ddagger}$	$0.78^{\ddagger}$	0.48	0.49	0.69	0.70
Best-PMI-SVD	$0.76^{\ddagger}$	$0.71^{\ddagger}$	$0.68^{\ddagger}$	$0.66^{\ddagger}$	0.40	0.40	0.61	0.59
SensEmbed	0.89	**0.88**	0.65	0.75	$0.46^{\dagger}$	$0.47^{\dagger}$	0.67	0.70

Intrinsic evaluation

(Camacho-Collados et al., ACL 2015)

Most current approaches are developed for English only and there are no many datasets to evaluate multilinguality.

To this end, we developed a semi-automatic framework to extend English datasets to other languages (and across languages):

Data available at

http://lcl.uniroma1.it/similarity-datasets/

Intrinsic evaluation

Multilingual semantic similarity

English	r	ρ	French	r	ρ	German	r	ρ	Spanish	r	ρ
NASARI	0.81	0.78	NASARI	**0.82**	0.73	NASARI	0.69	0.65	NASARI	**0.85**	0.79
NASARI$_{\text{lexical}}$	0.80	0.78	NASARI$_{\text{lexical}}$	0.80	0.70	NASARI$_{\text{lexical}}$	0.69	0.67	NASARI$_{\text{lexical}}$	**0.85**	0.79
NASARI$_{\text{unified}}$	0.80	0.76	NASARI$_{\text{unified}}$	**0.82**	**0.76**	NASARI$_{\text{unified}}$	**0.71**	**0.68**	NASARI$_{\text{unified}}$	0.82	0.77
NASARI$_{\text{embed}}$	**0.82**	**0.80**	–	–	–	–	–	–	NASARI$_{\text{embed}}$	0.79	0.77
SOC-PMI	0.61	–	SOC-PMI	0.19	–	SOC-PMI	0.27	–	–	–	–
PMI	0.41	–	PMI	0.34	–	PMI	0.40	–	–	–	–
LSA-Wiki	0.65	0.69	LSA-Wiki	0.57	0.52	–	–	–	–	–	–
Wiki-wup	0.59	–	–	–	–	Wiki-wup	0.65	–	–	–	–
Word2Vec	–	0.73	Word2Vec	–	0.47	Word2Vec	–	0.53	Best-Word2Vec	0.80	**0.80**
Retrofitting	–	0.77	Retrofitting	–	0.61	Retrofitting	–	0.60	–	–	–
NASARI$_{\text{poly-embed}}$	0.74	0.77	NASARI$_{\text{poly-embed}}$	0.60	0.69	NASARI$_{\text{poly-embed}}$	0.46	0.52	NASARI$_{\text{poly-embed}}$	0.68	0.74
Polyglot-embed	0.51	0.55	Polyglot-embed	0.38	0.35	Polyglot-embed	0.18	0.15	Polyglot-embed	0.51	0.56
IAA	0.85°	-	IAA	-	-	IAA	0.81	-	IAA	0.83	-

Intrinsic evaluation

Cross-lingual semantic similarity

Measure	**EN-FR**		**EN-DE**		**EN-ES**		**FR-DE**		**FR-ES**		**DE-ES**		**Average**	
	r	ρ	r	ρ	r	ρ	r	ρ	r	ρ	r	ρ	r	ρ
NASARI$_{\text{unified}}$	**0.84**	0.79	**0.79**	0.79	**0.84**	0.82	0.75	0.70	**0.86**	0.78	**0.81**	**0.80**	**0.82**	0.78
CL-MSR-2.0	0.30	–	–	–	–	–	–	–	–	–	–	–	–	–
NASARI$_{pivot}$	0.79	0.69	0.78	0.76	0.80	0.74	**0.79**	0.70	0.80	0.67	0.72	0.68	0.78	0.71
ADW$_{pivot}$	0.80	0.82	0.73	**0.82**	0.78	**0.84**	0.72	**0.77**	0.81	**0.81**	0.68	0.72	0.75	**0.80**
Word2Vec$_{pivot}$	0.77	0.82	0.70	0.73	0.76	0.80	0.65	0.70	0.75	0.76	0.64	0.63	0.71	0.74
Best-Word2Vec$_{pivot}$	0.75	**0.84**	0.69	0.76	0.75	0.82	0.77	0.73	0.74	0.79	0.64	0.64	0.72	0.76
Best-PMI-SVD$_{pivot}$	0.76	0.76	0.72	0.74	0.77	0.77	0.65	0.69	0.76	0.74	0.62	0.61	0.71	0.72

NEW: **SemEval 2017** task on multilingual and cross-lingual semantic word similarity

Large datasets to evaluate semantic similarity in **five languages** (within and across languages): English, Farsi, German, Italian and Spanish.

Additional challenges:

- **Multiwords:** *black hole*
- **Entities:** *Microsoft*
- **Domain-specific terms:** *chemotherapy*

Data available at

http://alt.qcri.org/semeval2017/task2/

Applications

- Domain labeling/adaptation
- Word Sense Disambiguation
- Sense Clustering
- Topic categorization and sentiment analysis

Domain labeling

(Camacho-Collados et al., AIJ 2016)

Annotate each **concept/entity** with its corresponding **domain of knowledge**.

To this end, we use the Wikipedia featured articles page, which includes 34 domains and a number of Wikipedia pages associated with each domain (*Biology*, *Geography*, *Mathematics*, *Music*, etc.).

Domain labeling

Wikipedia featured articles

Chemistry and mineralogy

Acetic acid • Antioxidant • Astatine • Caesium • Californium • Cyclol • Diamond • DNA nanotechnology • Enzyme • Enzyme inhibitor • Enzyme kinetics • Fluorine • Francium • Germanium • Helium • Hydrochloric acid • Hydrogen • Iridium • Lead(II) nitrate • Metalloid • Nicotinamide adenine dinucleotide • Niobium • Noble gas • Oxidative phosphorylation • Oxygen • Periodic table • Plutonium • Psilocybin • Rhodocene • Synthetic diamond • Technetium • Titanium • Ununoctium • Ununseptium • Uranium • Xenon • Yogo sapphire • Yttrium • Zinc

Chemistry and mineralogy biographies

James Bryant Conant • Joseph Priestley

Computing

4chan • Acid2 • Delrina • Folding@home • Macintosh Classic • Manchester Mark 1 • Manchester Small-Scale Experimental Machine • Microsoft Security Essentials • The Million Dollar Homepage • NeXT • Parallel computing • PowerBook 100 • Rosetta@home • ROT13 • Scene7

Culture and society

Aggie Bonfire • Hadji Ali • The Livestock Conservancy • Anna Anderson • Marshall Applewhite • Baden-Powell House • Isabella Beeton • Biddenden Maids • William D. Boyce • Guy Bradley • Burke and Hare murders • William Henry Bury • "The Bus Uncle" • Josephine Butler • The Chaser APEC pranks • Cleveland Street scandal • Cock Lane ghost • D. B. Cooper • Daylight saving time • Disco Demolition Night • Charles Domery • Dorset Ooser • Marjory Stoneman Douglas • Montague Druitt • W. E. B. Du Bois • Monroe Edwards • Female genital mutilation • Terry Fox • Ursula Franklin • Free Association of German Trade Unions • Margaret Fuller • E. Urner Goodman • Debora Green • Stanley Green • Green children of Woolpit • Gropecunt Lane • Guy Fawkes Night • Hanged, drawn and quartered • William Hillcourt • Fanny Imlay • Indigenous people of the Everglades region • *An Introduction to Animals and Political Theory* • Jack the Ripper • *Jack the Ripper: The Final Solution* • *Ketuanan Melayu* • Akhtar Hameed Khan • Kylfings • Daniel Lambert • Liberty Bell • Lynching of Jesse Washington • Macedonia (terminology) • Mantra-Rock Dance • Bob Marshall (wilderness activist) • Murder of Dwayne Jones • Florence Nagle • The Negro Motorist Green Book • Emmeline Pankhurst • Pig-faced women • Polish culture during World War II • Postage stamps of Ireland • Ramblin' Wreck • Rosewood massacre • Royal baccarat scandal • Same-sex marriage in Spain • Mark Satin • Scouting • John Martin Scripps • Sexuality after spinal cord injury • Grace Sherwood • Society of the Song dynasty • Stonewall riots • Taiwanese aborigines • Mary Toft • Toraja • Truthiness • Voluntary Human Extinction Movement • Whitechapel murders • Wife selling (English custom) • Wonderbra • Wood Badge • Robert Sterling Yard • Zong massacre

Education

Alpha Kappa Alpha • Amador Valley High School • ANAK Society • Avery Coonley School • Baltimore City College • Boden Professor of Sanskrit election, 1860 • James E. Boyd (scientist) • C. R. M. F. Cruttwell • Dartmouth College • Duke University • Florida Atlantic University • Georgetown University • The Green (Dartmouth College) • *The Guardian of Education* • History of Baltimore City College • History of Texas A&M University • The Judd School • Kappa Kappa Psi • *Lessons for Children* • Michigan State University • Ohio Wesleyan University • Oriel College, Oxford • Romney Literary Society • Royal National College for the Blind • School for Creative and Performing Arts • Shimer College • *Some Thoughts Concerning Education* • Stuyvesant High School • Texas A&M University • Texas Tech University • *Thoughts on the Education of Daughters* • Tuck School of Business • United States Academic Decathlon • United States Military Academy • University of California, Riverside • University of Michigan • Vkhutemas

Engineering and technology

2013 Rosario gas explosion • Apollo 8 • Atomic line filter • Caesar cipher • Calutron • CFM International CFM56 • Construction of the World Trade Center • Distributed element filter • Draining and development of the Everglades • Gas metal arc welding • Gas tungsten arc welding • Grand Coulee Dam • Halkett boat • Hanford Site • History of timekeeping devices • Hoover Dam • Mechanical filter • Oil shale • Panavision • Pigeon photography • Rampart Dam • Renewable energy in Scotland • Restoration of the Everglades • Rolls-Royce Merlin • Rolls-Royce R • Scout Moor Wind Farm • Shale oil extraction • Shielded metal arc welding • Shoe polish • Sholes and Glidden typewriter • Shuttle–*Mir* Program • Science and technology of the Song dynasty • Waveguide filter • Webley Revolver • Welding • World Science Festival, 2008

Domain labeling

How to associate a synset with a domain?

- We first construct a **NASARI lexical vector** for the **concatenation of all Wikipedia pages associated with a given domain** in the featured article page.

- Then, we calculate the **semantic similarity** between the corresponding NASARI vectors of the synset and all domains:

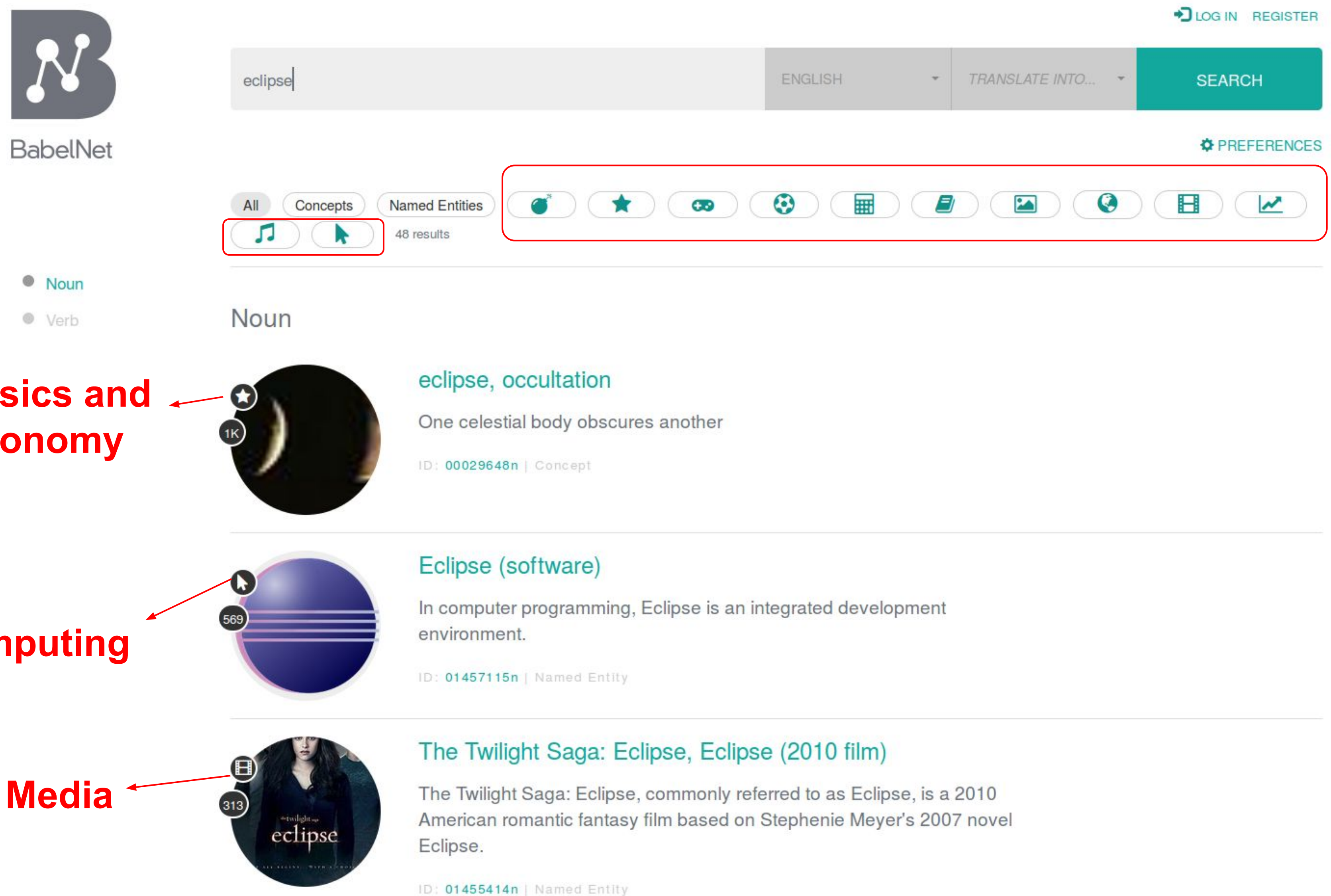
$$\hat{d}(s) = \underset{d \in D}{\arg\max}\, WO\,(\overrightarrow{\text{NASARI}}_{lex}(s), \vec{v}_{lex}(d))$$

Domain labeling

This results in **over 1.5M synsets** associated with a domain of knowledge.

This domain information has already been integrated in the last version of BabelNet.

Domain labeling



Physics and astronomy

Computing

Media

Domain labeling

	WordNet dataset			BabelNet dataset		
	Precision	Recall	F-Measure	Precision	Recall	F-Measure
NASARI$_{\text{lexical}}$	**77.9**	70.1	**73.8**	**62.3**	**40.5**	**49.1**
Wikipedia-TF	25.4	16.4	19.9	3.4	2.5	2.9
Wikipedia-TFidf	45.9	29.7	36.1	8.8	6.5	7.5
Taxo-Prop (WN)	71.3	70.7	71.0	–	–	–
Taxo-Prop (BN)	73.5	**73.5**	73.5	48.3	37.2	42.0
WN-Domains-3.2	93.6	64.4	76.3	–	–	–

Domain labeling results on WordNet and BabelNet

BabelDomains

(Camacho-Collados and Navigli, EACL 2017)

As a result:

Unified resource with information about domains of knowledge

*BabelDomains available for **BabelNet, Wikipedia and WordNet** available at*

http://lcl.uniroma1.it/babeldomains

Already integrated into BabelNet (online interface and API)

Domain filtering for supervised distributional hypernym discovery

(Espinosa-Anke et al., EMNLP 2016;
Camacho-Collados and Navigli, EACL 2017)

Apple

is a

Fruit

Task: Given a term, predict its hypernym(s)

Model: Distributional supervised system based on the transformation matrix of Mikolov et al. (2013).

Idea: Training data filtered by domain of knowledge

Domain filtering for supervised distributional hypernym discovery

Domain-filtered training data

	art			biology			education			geography			health		
Train	MRR	MAP	R-P	MRR	MAP	R-P	MRR	MAP	R-P	MRR	MAP	R-P	MRR	MAP	R-P
5k	0.12	0.12	0.12	0.63	0.63	0.59	0.00	0.00	0.00	0.08	0.07	0.07	0.08	0.08	0.07
15k	0.21	0.20	0.18	**0.84**	0.72	0.79	0.22	0.22	0.21	0.15	0.14	0.14	0.08	0.07	0.07
25k	**0.29**	0.27	**0.26**	**0.84**	**0.83**	**0.81**	0.33	0.32	0.30	**0.23**	**0.22**	**0.21**	0.09	0.09	0.08
25k+K^{d}_{1k}	**0.29**	**0.28**	**0.26**	**0.84**	0.80	0.79	0.32	0.29	0.27	0.22	**0.22**	**0.21**	0.09	0.09	0.08
25k+K^{d}_{25k}	0.26	0.24	0.22	0.70	0.63	0.56	**0.38**	**0.36**	**0.33**	0.15	0.13	0.12	0.11	0.11	0.10
25k+K^{r}_{50k}	0.28	0.26	0.24	0.82	0.77	0.72	0.36	0.33	0.30	0.17	0.16	0.16	**0.12**	0.11	0.10
100k$^{r}_{wd}$	0.00	0.00	0.00	**0.84**	0.81	0.77	0.00	0.00	0.00	0.01	0.01	0.01	0.07	0.06	0.06
100k$^{r}_{kbu}$	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	**0.12**	0.12	0.11
Baseline	0.13	0.12	0.10	0.58	0.57	**0.57**	0.10	0.10	0.09	0.12	0.09	0.05	0.07	**0.13**	**0.14**

Non-filtered training data

Results on the hypernym discovery task for five domains

Conclusion: Filtering training data by domains prove to be clearly beneficial

Word Sense Disambiguation

Kobe, which is one of Japan's largest cities, [...]

?

Word Sense Disambiguation

Kobe, which is one of Japan's largest cities, [...]

X

Word Sense Disambiguation

Kobe, which is one of Japan's largest cities, [...]

Word Sense Disambiguation

(Camacho-Collados et al., AIJ 2016)

Basic idea

Select the sense which is semantically closer to the semantic representation of the whole document (**global context**).

$$\hat{d}(s) = \operatorname*{argmax}_{d \in D} WO(\vec{\text{NASARI}}_{lex}(s), \vec{v}_{lex}(d))$$

Word Sense Disambiguation

System	English	French	Italian	German	Spanish	Average
NASARI	86.3	**76.2**	83.7	**83.2**	82.9	**82.5**
MUFFIN	84.5	71.4	81.9	83.1	**85.1**	81.2
Babelfy	**87.4**	71.6	**84.3**	81.6	83.8	81.7
UMCC-DLSI	54.8	60.5	58.3	61.0	58.1	58.5
MFS	80.2	74.9	82.2	83.0	82.1	79.3

Multilingual Word Sense Disambiguation using Wikipedia as sense inventory (F-Measure)

Word Sense Disambiguation

System	SemEval-2013	SemEval-2007
NASARI	66.7	66.7
NASARI+IMS	67.0	**68.5**
MUFFIN	66.0	66.0
Babelfy	65.9	62.7
UKB	61.3	56.0
UMCC-DLSI	64.7	–
Multi-Objective	**72.8**	66.0
IMS	65.3	67.3
MFS	63.2	65.8

All-words Word Sense Disambiguation using WordNet as sense inventory (F-Measure)

Word Sense Disambiguation

System	SemEval-2013	SemEval-2007
NASARI	66.7	66.7
NASARI+IMS	67.0	**68.5**
MUFFIN	66.0	66.0
Babelfy	65.9	62.7
UKB	61.3	56.0
UMCC-DLSI	64.7	–
Multi-Objective	**72.8**	66.0
IMS	65.3	67.3
MFS	63.2	65.8

All-words Word Sense Disambiguation using WordNet as sense inventory (F-Measure)

Word Sense Disambiguation: Empirical Comparison

(Raganato et al., EACL 2017)

- Supervised systems clearly outperform knowledge-based systems, but they **only exploit local context** (future direction -> integration of both)

- Supervised systems perform well when trained on **large amounts of sense-annotated data** (even if not manually annotated).

Data and results available at

http://lcl.uniroma1.it/wsdeval/

Word Sense Disambiguation on textual definitions

(Camacho-Collados et al., LREC 2016)

Combination of a graph-based disambiguation system (Babelfy) with NASARI to **disambiguate** the concepts and named entities of **over 35M definitions** in **256 languages**.

Sense-annotated corpus freely available at

http://lcl.uniroma1.it/disambiguated-glosses/

Context-rich WSD

castling *(chess)*

*Interchanging the positions of the **king** and a **rook**.*

Context-rich WSD

castling *(chess)*

Interchanging the positions of the ***king*** *and a* ***rook****.*

Castling *is a move in the game of* ***chess*** *involving a player's* ***king*** *and either of the player's original* ***rooks****.*

A move in which the ***king*** *moves two* ***squares*** *towards a* ***rook****, and the* ***rook*** *moves to the other side of the* ***king.***

Context-rich WSD

castling *(chess)*

*Interchanging the positions of the **king** and a **rook**.*

***Castling** is a move in the game of **chess** involving a player's **king** and either of the player's original **rooks**.*

*A move in which the **king** moves two **squares** towards a **rook**, and the **rook** moves to the other side of the **king.***

*Manœuvre du jeu d'**échecs***

***Rošáda** je zvláštní tah v šachu, při kterém táhne zároveň **král** a **věž**.*

*Spielzug im **Schach**, bei dem **König** und **Turm** einer Farbe bewegt werden*

*El **enroque** es un movimiento especial en el juego de **ajedrez** que involucra al **rey** y a una de las **torres** del jugador.*

***Rokade** er et spesialtrekk i **sjakk**.*

***Rok** İngilizce'de kaleye **rook** denmektedir.*

Το ροκέ είναι μια ειδική **κίνηση** στο **σκάκι** που συμμετέχουν ο βασιλιάς και ένας από τους δυο **πύργους**.

Context-rich WSD

castling *(chess)*

*Interchanging the positions of the **king** and a **rook**.*

W ***Castling** is a move in the game of **chess** involving a ...'s **king** and either of the player... s.*

*...hich the **king** moves two ...ards a **rook**, and the **rook** ... other side of the **king.***

*W Manc... d'**éche**...*

*W **Rošáda** je zvláštní tah v šachu, při kterém táhne zároveň **král** a **věž**.*

*Spielzug im **Schach**, bei dem **König** und **Turm** einer Farbe bewegt werden*

*W El **enroque** es un movimiento especial en el juego de **ajedrez** que involucra al **rey** y a una de las **torres** del jugador.*

*W **Rokade** er et spesialtrekk i **sjakk**.*

*W **Rok** İngilizce'de kaleye **rook** denmektedir.*

W Το ροκέ είναι μια ειδική **κίνηση** στο **σκάκι** που συμμετέχουν ο βασιλιάς και ένας από τους δυο **πύργους**.

Context-rich WSD exploiting parallel corpora

(Delli Bovi et al., ACL 2017)

Applying the same method to provide **high-quality sense annotation from parallel corpora** (Europarl): 120M+ sense annotations for 21 languages.

Extrinsic evaluation: Improved performance of a standard supervised WSD system using this automatically sense-annotated corpora.

Sense Clustering

- Current sense inventories suffer from the **high granularity** of their sense inventories.

- A meaningful clustering of senses would help **boost the performance on downstream applications** (Hovy et al., 2013)

Example:

- Parameter (computer programming) - Parameter

Sense Clustering

Idea

Using a clustering algorithm based on the **semantic similarity between sense vectors**

Sense Clustering

(Camacho-Collados et al., AIJ 2016)

Measure	System type	500-pair		SemEval	
		Acc.	**F1**	**Acc.**	**F1**
NASARI	unsupervised	83.8	**70.5**	**87.4**	**63.1**
$\text{NASARI}_{\text{lexical}}$	unsupervised	81.6	65.4	85.7	57.4
$\text{NASARI}_{\text{unified}}$	unsupervised	82.6	69.5	87.2	**63.1**
$\text{NASARI}_{\text{embed}}$	unsupervised	81.2	65.9	86.3	45.5
SVM-monolingual	supervised	77.4	-	83.5	-
SVM-multilingual	supervised	**84.4**	-	85.5	-
$\text{Baseline}_{\text{no-cluster}}$	-	71.4	0.0	82.5	0.0
$\text{Baseline}_{\text{cluster}}$	-	28.6	44.5	17.5	29.8

Clustering of Wikipedia pages

Towards a seamless integration of senses in downstream NLP applications

(Pilehvar et al., ACL 2017)

Question: What if we apply WSD and inject sense embeddings to a standard neural classifier?

Problems:

Towards a seamless integration of senses in downstream NLP applications

(Pilehvar et al., ACL 2017)

Question: What if we apply WSD and inject sense embeddings to a standard neural classifier?

Problems:

- WSD is not perfect

Towards a seamless integration of senses in downstream NLP applications

(Pilehvar et al., ACL 2017)

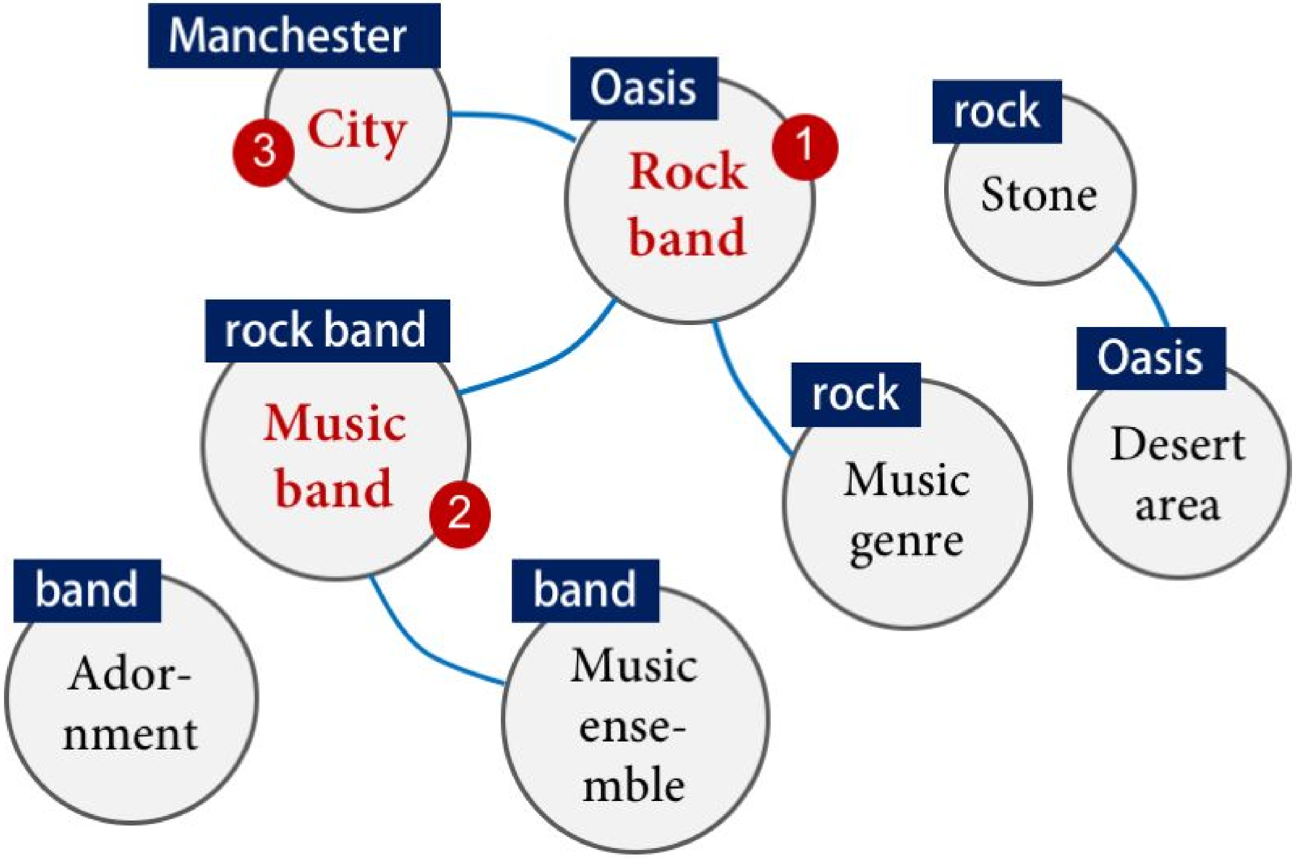
Question: What if we apply WSD and inject sense embeddings to a standard neural classifier?

Problems:

- WSD is not perfect **-> Solution:** High-confidence disambiguation

High confidence graph-based disambiguation

Oasis was a rock band formed in Manchester.



Towards a seamless integration of senses in downstream NLP applications

(Pilehvar et al., ACL 2017)

Question: What if we apply WSD and inject sense embeddings to a standard neural classifier?

Problems:

- WSD is not perfect **-> Solution:** High-confidence disambiguation

- Senses in WordNet are too fine-grained

Towards a seamless integration of senses in downstream NLP applications

(Pilehvar et al., ACL 2017)

Question: What if we apply WSD and inject sense embeddings to a standard neural classifier?

Problems:

- WSD is not perfect **-> Solution:** High-confidence disambiguation
- Senses in WordNet are too fine-grained **-> Solution:** Supersenses

Towards a seamless integration of senses in downstream NLP applications

(Pilehvar et al., ACL 2017)

Question: What if we apply WSD and inject sense embeddings to a standard neural classifier?

Problems:

- WSD is not perfect **-> Solution:** High-confidence disambiguation
- Senses in WordNet are too fine-grained **-> Solution:** Supersenses
- WordNet lacks coverage

Towards a seamless integration of senses in downstream NLP applications

(Pilehvar et al., ACL 2017)

Question: What if we apply WSD and inject sense embeddings to a standard neural classifier?

Problems:

- WSD is not perfect **-> Solution:** High-confidence disambiguation
- Senses in WordNet are too fine-grained **-> Solution:** Supersenses
- WordNet lacks coverage **-> Solution:** Use of Wikipedia

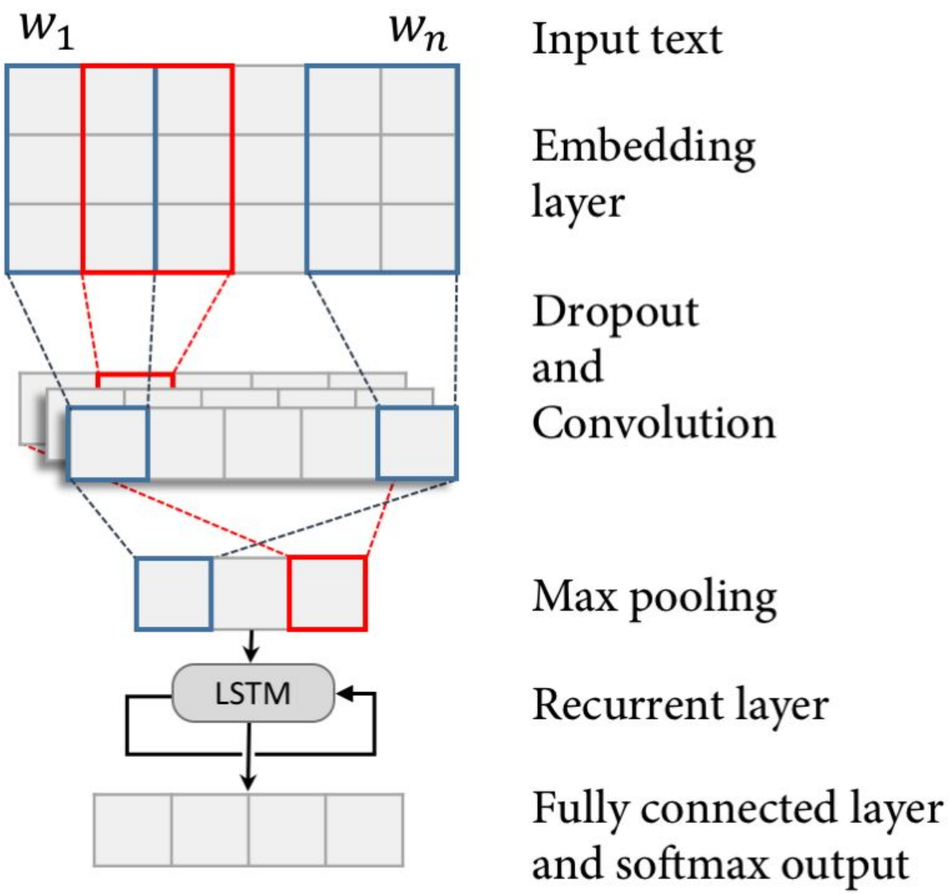
Tasks: Topic categorization and sentiment analysis (polarity detection)

Topic categorization: Given a text, assign it a label (i.e. topic).

Polarity detection: Predict the sentiment of the sentence/review as either positive or negative.

Classification model

Standard CNN classifier inspired by Kim (2014)



Sense-based vs. word-based: Conclusions

- Coarse-grained senses (*supersenses*) better than fine-grained senses.

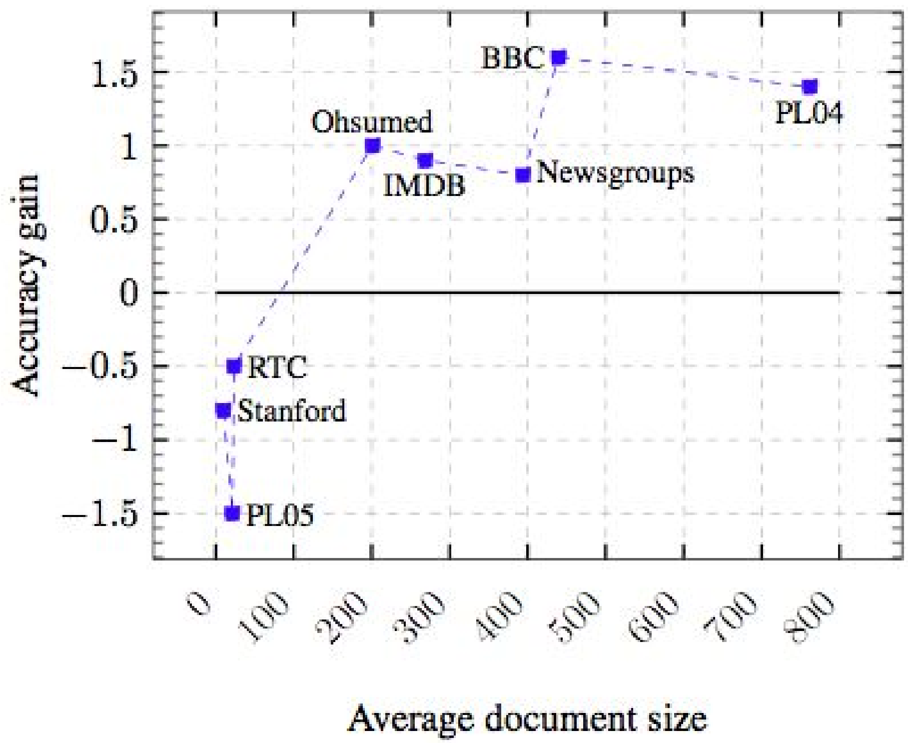
Sense-based vs. word-based: Conclusions

- Coarse-grained senses (*supersenses*) better than fine-grained senses.

- Sense-based **better** than word-based... when the **input text is large enough**

Sense-based vs. word-based:

Sense-based **better** than word-based... when the **input text is large enough**:



Why
does the input text size matter?

- Graph-based WSD works better in larger texts (Moro et al. 2014; Raganato et al. 2017)

- Disambiguation increases sparsity

Conclusions of the talk

- Novel approach to **represent concepts and entities in a multilingual vector space** (**NASARI**).

- These knowledge-based sense representations can be easily integrated in several applications, acting as a glue for **combining corpus-based information and knowledge from lexical resources**, while enabling:

  - **Multilinguality**

  - **Work at the deeper sense level**

For more information on other sense-based representations and their applications:

- ACL 2016 Tutorial on “**Semantic representations of word senses and concepts”:** http://acl2016.org/index.php?article_id=58

- EACL 2017 workshop on “Sense, Concept and Entity Representations and their Applications”: https://sites.google.com/site/senseworkshop2017/

Thank you!

Questions please!

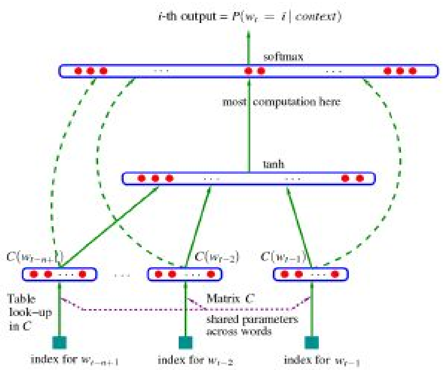
collados@di.uniroma1.it

Secret Slides

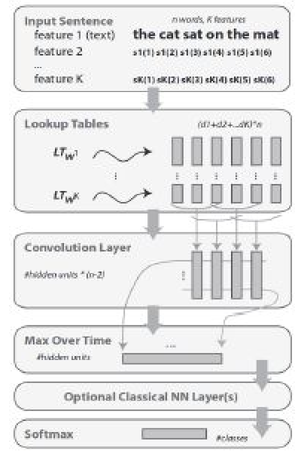
Word vector space models

Words are represented as vectors: semantically similar words are close in the space

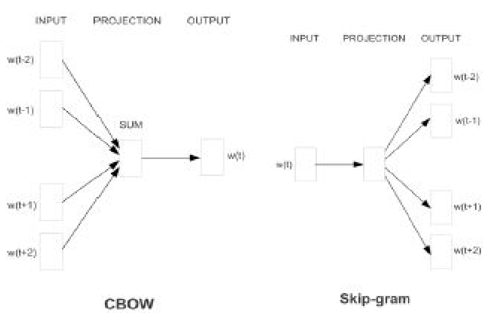
Neural networks for learning word vector representations from text corpora -> **word embeddings**



Bengio et al. (2003)



Collobert & Weston (2008)

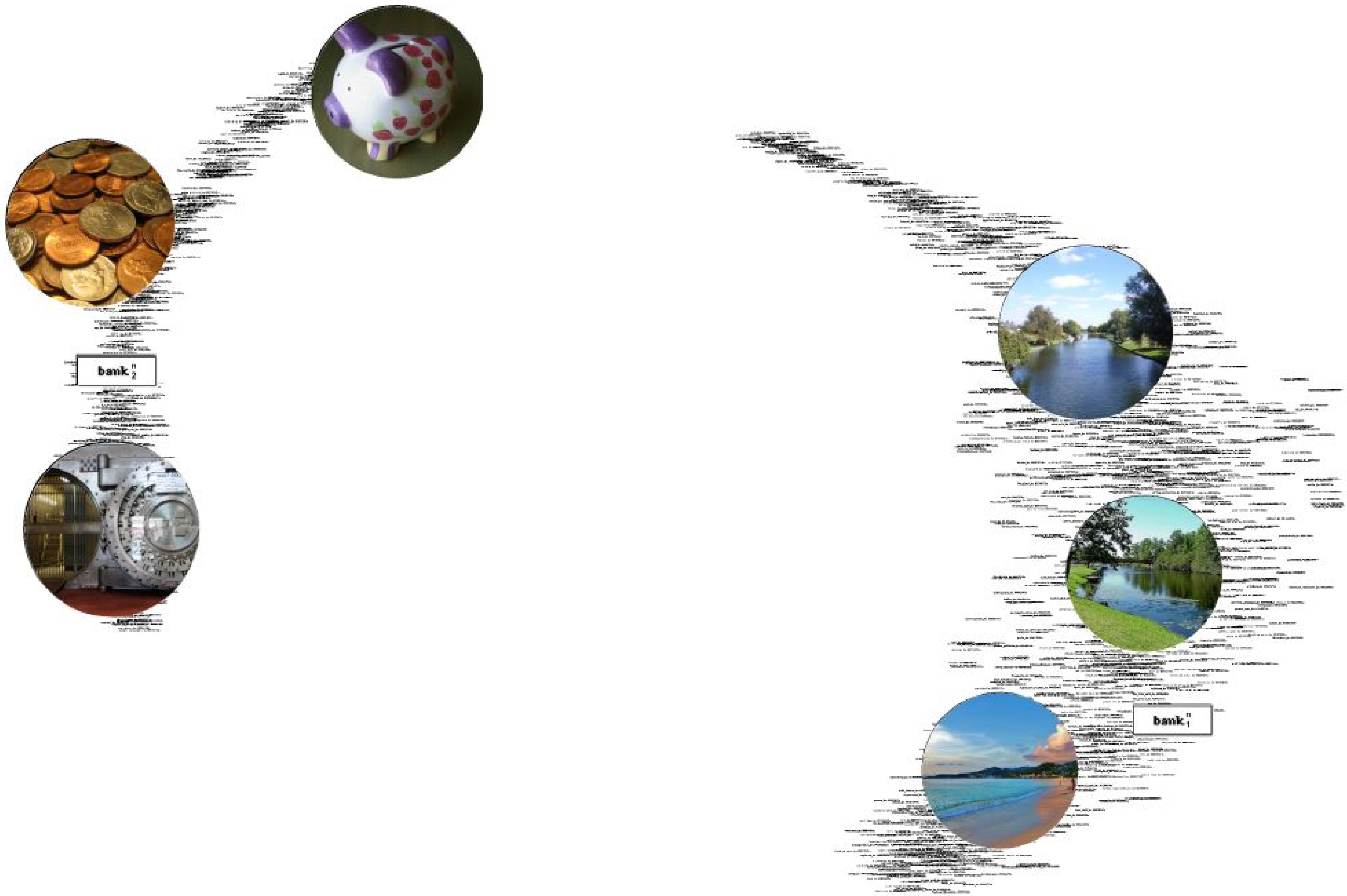


Mikolov et al. (2013)

Probability and Ratio	$k = solid$	$k = gas$	$k = water$
$P(k\|ice)$	1.9×10^{-4}	6.6×10^{-5}	3.0×10^{-3}
$P(k\|steam)$	2.2×10^{-5}	7.8×10^{-4}	2.2×10^{-3}
$P(k\|ice)/P(k\|steam)$	8.9	8.5×10^{-2}	1.36

Pennington et al. (2014)

Key goal: obtain sense representations



NASARI semantic representations

- NASARI 1.0 (April 2015): *Lexical and unified vector representations for WordNet synsets and Wikipedia pages for English.*

José Camacho Collados, Mohammad Taher Pilehvar and Roberto Navigli. *NASARI: a Novel Approach to a Semantically-Aware Representation of Items.* **NAACL 2015**, Denver, USA, pp. 567-577.

NASARI semantic representations

- NASARI 1.0 (April 2015): *Lexical and unified vector representations for WordNet synsets and Wikipedia pages for English.*

José Camacho Collados, Mohammad Taher Pilehvar and Roberto Navigli. *NASARI: a Novel Approach to a Semantically-Aware Representation of Items.* **NAACL 2015**, Denver, USA, pp. 567-577.

- NASARI 2.0 (August 2015): + Multilingual extension.

José Camacho Collados, Mohammad Taher Pilehvar and Roberto Navigli. *A Unified Multilingual Semantic Representation of Concepts.* **ACL 2015**, Beijing, China, pp. 741-751.

NASARI semantic representations

- NASARI 1.0 (April 2015): *Lexical and unified vector representations for WordNet synsets and Wikipedia pages for English.*

José Camacho Collados, Mohammad Taher Pilehvar and Roberto Navigli. *NASARI: a Novel Approach to a Semantically-Aware Representation of Items.* **NAACL 2015**, Denver, USA, pp. 567-577.

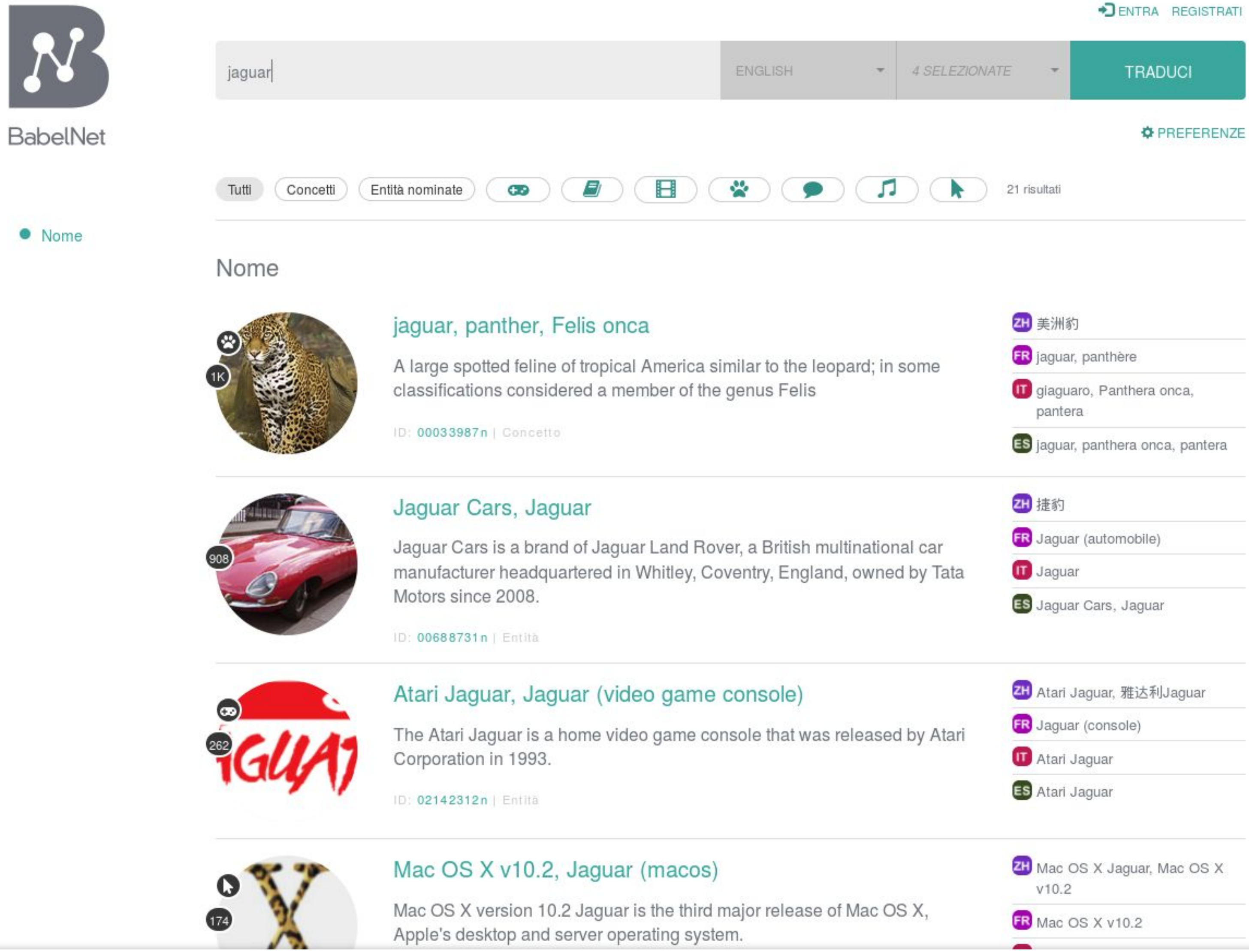
- NASARI 2.0 (August 2015): + Multilingual extension.

José Camacho Collados, Mohammad Taher Pilehvar and Roberto Navigli. *A Unified Multilingual Semantic Representation of Concepts.* **ACL 2015**, Beijing, China, pp. 741-751.

- NASARI 3.0 (March 2016): + Embedded representations, new applications.

José Camacho Collados, Mohammad Taher Pilehvar and Roberto Navigli. *Nasari: Integrating explicit knowledge and corpus statistics for a multilingual representation of concepts and entities.* **Artificial Intelligence Journal, 2016,** 240, 36-64.

BabelNet



BabelNet
ENTRA REGISTRATI
jaguar
ENGLISH
4 SELEZIONATE
TRADUCI
PREFERENZE
Tutti
Concetti
Entità nominate
21 risultati
Nome
Nome
jaguar, panther, Felis onca
A large spotted feline of tropical America similar to the leopard; in some classifications considered a member of the genus Felis
ID: 00033987n | Concetto
1K
ZH 美洲豹
FR jaguar, panthère
IT giaguaro, Panthera onca, pantera
ES jaguar, panthera onca, pantera
Jaguar Cars, Jaguar
Jaguar Cars is a brand of Jaguar Land Rover, a British multinational car manufacturer headquartered in Whitley, Coventry, England, owned by Tata Motors since 2008.
ID: 00688731n | Entità
908
ZH 捷豹
FR Jaguar (automobile)
IT Jaguar
ES Jaguar Cars, Jaguar
Atari Jaguar, Jaguar (video game console)
The Atari Jaguar is a home video game console that was released by Atari Corporation in 1993.
ID: 02142312n | Entità
262
ZH Atari Jaguar, 雅达利Jaguar
FR Jaguar (console)
IT Atari Jaguar
ES Atari Jaguar
Mac OS X v10.2, Jaguar (macos)
Mac OS X version 10.2 Jaguar is the third major release of Mac OS X, Apple's desktop and server operating system.
174
ZH Mac OS X Jaguar, Mac OS X v10.2
FR Mac OS X v10.2

Three types of vector representations

Three types of vector representations:

- **Lexical** (dimensions are words): Dimensions are weighted via **lexical specificity** (statistical measure based on the hypergeometric distribution)

- **Unified** (dimensions are multilingual BabelNet synsets): This representation uses a **hypernym-based clustering technique** and can be used in **cross-lingual** applications

- **Embedded** (latent dimensions)

Key points

- **What** do we want to **represent**?
- What does "**semantic representation**" **mean**?
- **Why** semantic representations?
- What **problems** affect mainstream representations?
- How to **address** these problems?
- What comes **next**?

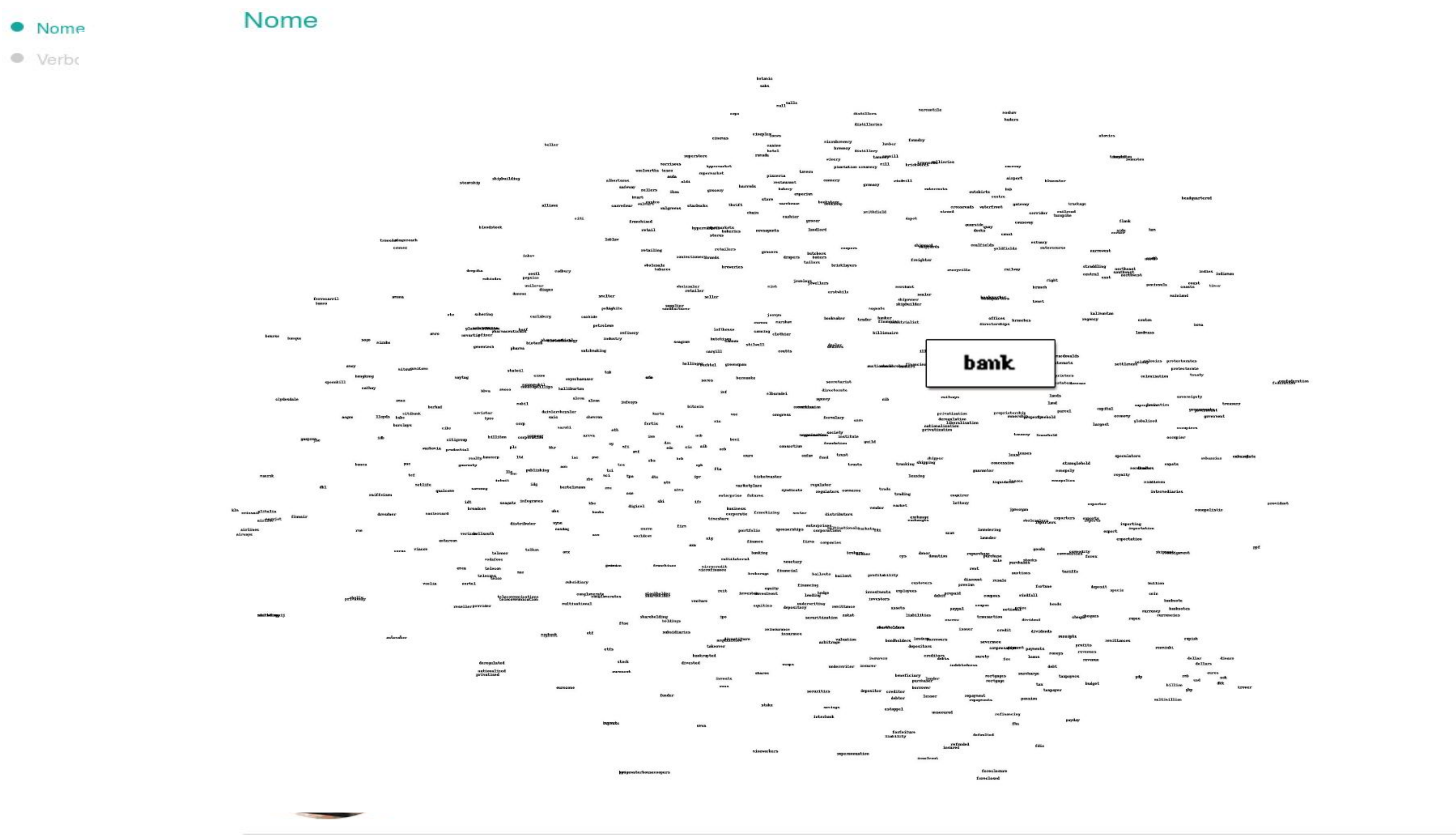
Problem 2: word representations do not take advantage of existing semantic resources

WIKIDATA

BabelNet

WIKIPEDIA

Key goal: obtain sense representations



on

Named Entity Disambiguation

System	Type	F-Measure
$\text{NASARI}_{\text{lexical}}$	unsupervised	87.1
DFKI	supervised	**88.9**
SUDOKU	unsupervised	87.0
el92	systems mix	86.1
MFS	–	85.7

Named Entity Disambiguation using BabelNet as sense inventory on the SemEval-2015 dataset

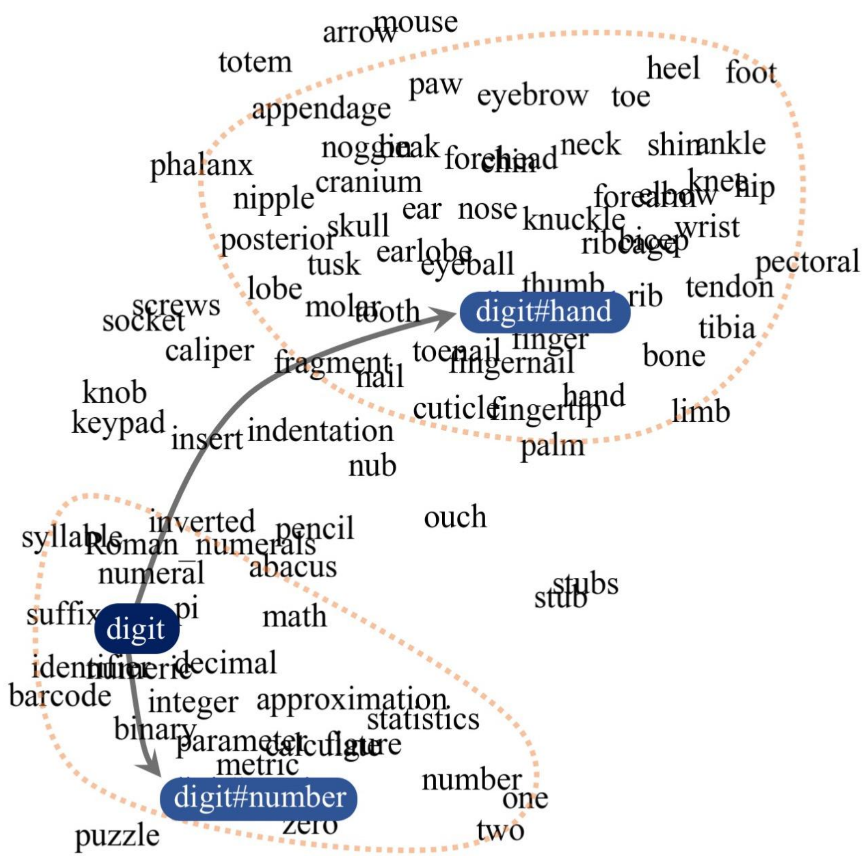
Word Sense Disambiguation

Open problem

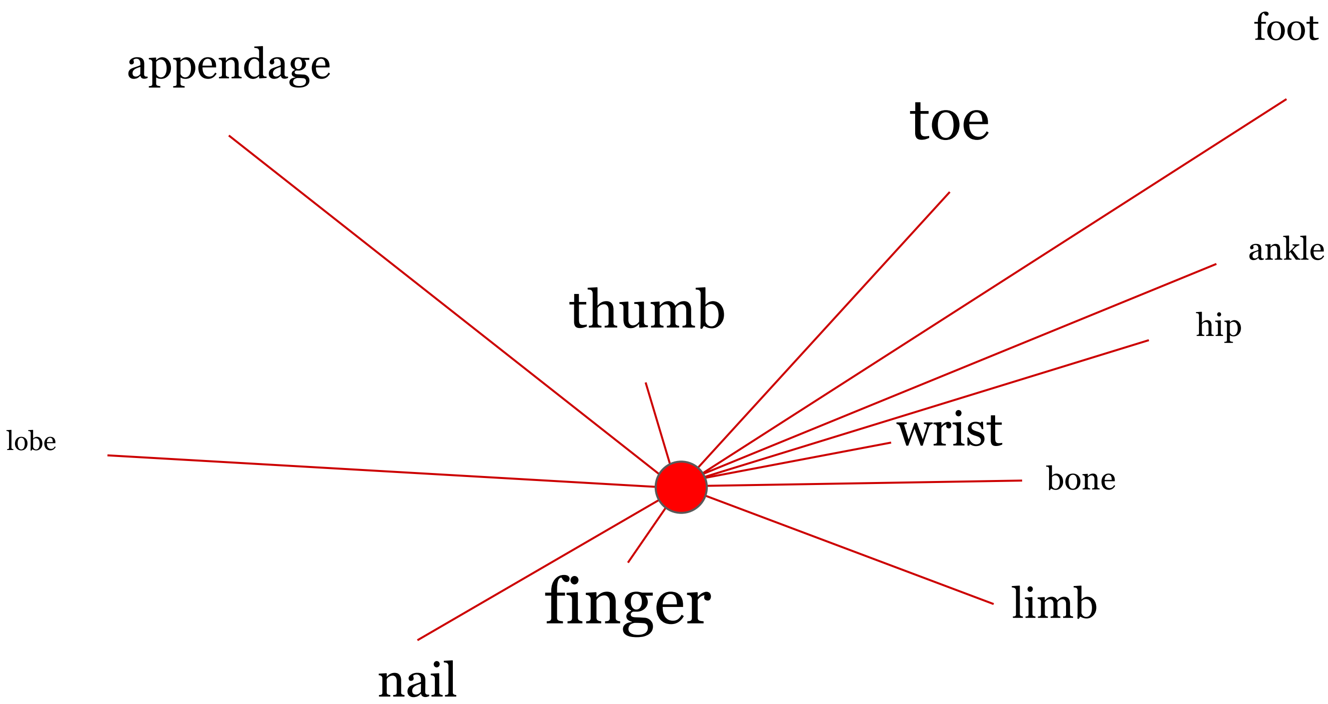
Integration of **knowledge-based** (exploiting global contexts) and **supervised** (exploiting local contexts) systems to overcome the *knowledge-acquisition bottleneck*.

De-Conflated Semantic Representations

M. T. Pilehvar and N. Collier (EMNLP 2016)



De-Conflated Semantic Representations



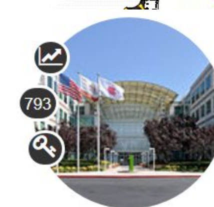
Open Problems and Future Work

1. Improve evaluation
   - Move from word similarity gold standards to end-to-end applications
     - Integration in Natural Language Understanding tasks (Li and Jurafsky, EMNLP 2015)
     - SemEval task? see e.g. WSD & Induction within an end user application @ SemEval 2013

Open Problems and Future Work

2. Make semantic representations more meaningful

- unsupervised representations are hard to inspect (clustering is hard to evaluate)
- but also knowledge-based approaches have issues:
  - e.g. top-10 closest vectors to the military sense of “company” in AutoExtend



AutoExtend
company_n^9
company
company_n^8
company_n^6
company_n^7
company_v^1
firm
business_n^1
firm_n^2
company_n^1

Open Problems and Future Work

3. Interpretability
   – The reason why things work or do not work is not obvious
     - E.g. avgSimC and maxSimC are based on implicit disambiguation that improves word similarity, but is not proven to disambiguate well
     - Many approaches are tuned to the task
   – Embeddings are difficult to interpret and debug

Open Problems and Future Work

4. Link the representations to rich semantic resources like WikiData and BabelNet
   – Enabling applications that can readily take advantage of huge amounts of multilinguality and information about concepts and entities
   – Improving the representation of low-frequency/isolated meanings

Open Problems and Future Work

5. Scaling semantic representations to sentences and documents
   – Sensitivity to word order
   – Combine vectors into syntactic-semantic structures
   – Requires disambiguation, semantic parsing, etc.
   – Compositionality

Open Problems and Future Work

6. Addressing multilinguality
   – a key trend in today's NLP research

- We are already able to perform POS tagging and dependency parsing in dozens of languages
  – Also mixing up languages

Open Problems and Future Work

- We can perform Word Sense Disambiguation and Entity Linking in hundreds of languages
  - Babelfy (Moro et al. 2014)
  - but with only a few sense vector representations
- Now: it is crucial that sense and concept representations are language-independent
- Enabling comparisons across languages
- Also useful in semantic parsing

Open Problems and Future Work

- Representations are most of the time evaluated in English
  - single words only
- It is important to evaluate sense representations in other languages and across languages
  - Check out the SemEval 2017 Task 2: multilingual and cross-lingual semantic word similarity (multilwords, entities, domain-specific, slang, etc.)

Open Problems and Future Work

7. Integrate sense representations into Neural Machine Translation
   - Previous results in the 2000s working on semantically-enhanced SMT are not very encouraging
   - However, many options have not been considered