Word, Sense and Contextualized Embeddings: Vector Representations of Meaning in NLP

Jose Camacho Collados

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

AI Wales, Cardiff, 6 February 2020

About me

- 5-year BSc in **Mathematics** and 2-year Masters in **Natural Language Processing (NLP)**: Spain-France-UK
- **Research engineer** (2013-2014): ATILF-CNRS (France)
- **PhD in Computer Science** (2014-2018): *Università La Sapienza (Italy)* + *Google (Google AI Doctoral Fellowship in NLP)*
- **Postdoc** (2018) and **Lecturer** (2019-*): Cardiff University, computer science

At Cardiff University

Research in Natural Language Processing (NLP):

- ➢ Semantics - Embeddings
- ➢ Multilinguality
- ➢ Applications

At Cardiff University

Research in Natural Language Processing (NLP):

- Semantics - Embeddings
- Multilinguality
- Applications

At Cardiff University

Research in Natural Language Processing (NLP):

- Semantics - Embeddings
- Multilinguality
- Applications

Teaching:

Artificial Intelligence and **Data Science** MSc.

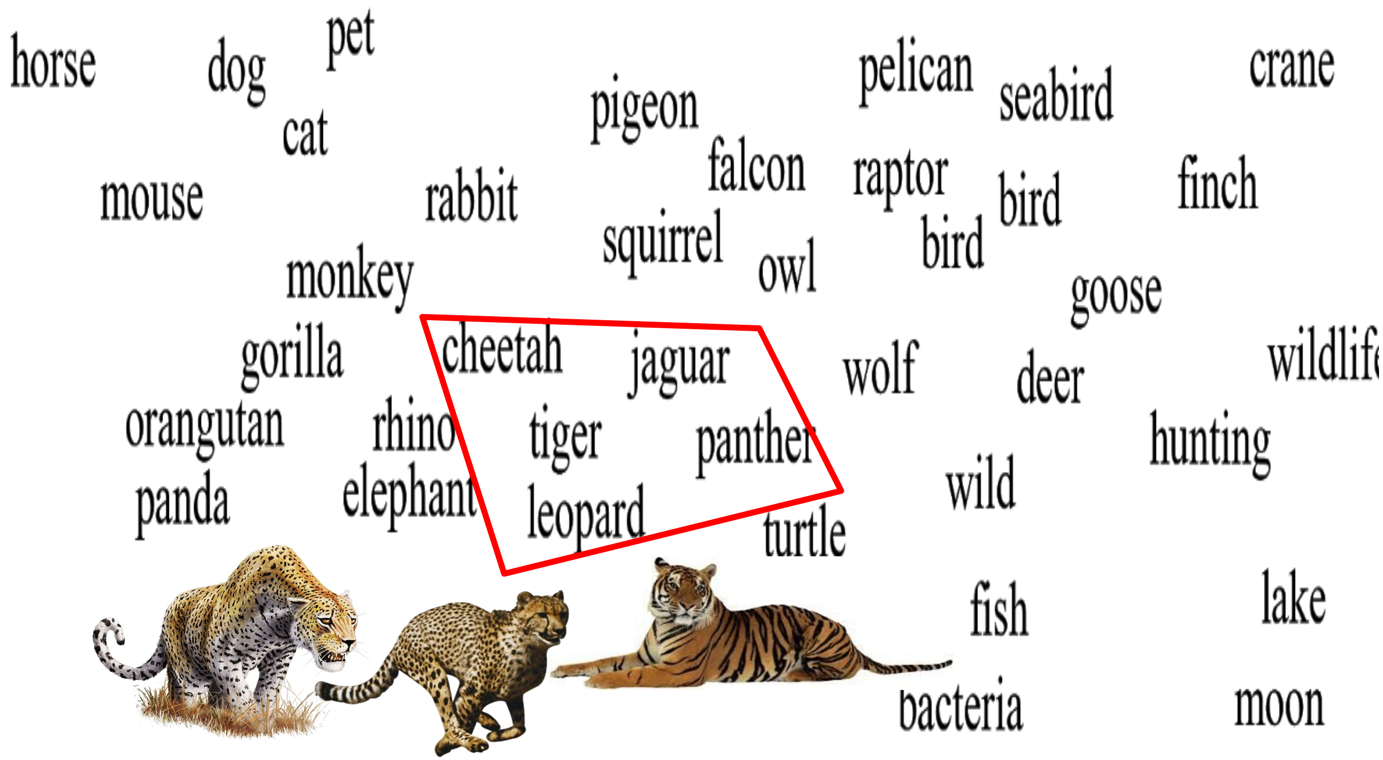
-> "**Applied Machine Learning**", with applications in NLP and Computer Vision.

Outline

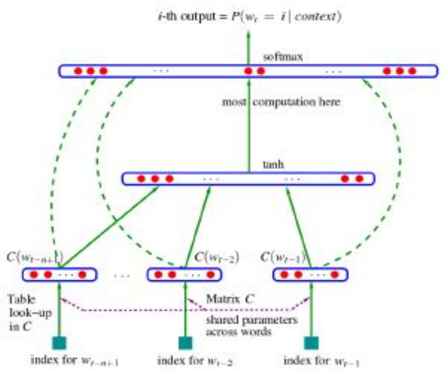
- Background
  - *Vector Space Models (word embeddings)*
  - *Lexical resources*
- Sense representations
  - *Knowledge-based: NASARI, SW2V*
  - *Contextualized: ELMo, BERT*
- Applications

Word vector space models

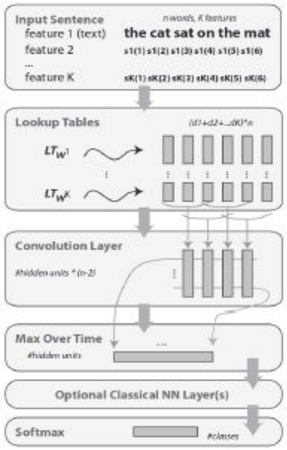
Words are represented as vectors: semantically similar words are close in the vector space



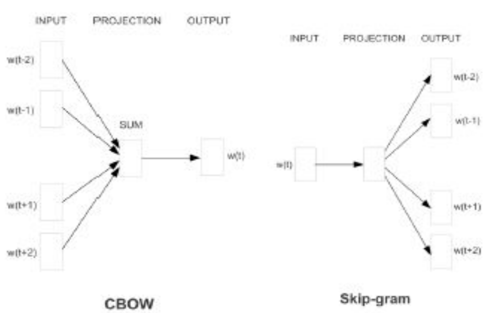
Neural networks for learning word vector representations from text corpora -> **word embeddings**



Bengio et al. (2003)



Collobert & Weston (2008)

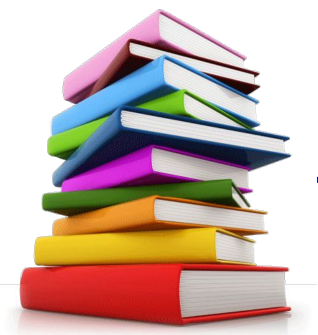


Mikolov et al. (2013)

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)

Word embeddings: How to learn them



... **London** is the **capital** of **UK** …

... Last night I **travelled** from **Cardiff** to **London**.

.

.

.

London

[0.25, 0.32, -0.1 …. 0.1]

Why word embeddings?

Embedded vector representations:

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

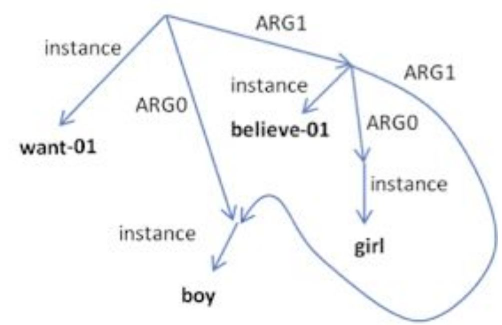
$$king - man + woman \approx queen$$

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)

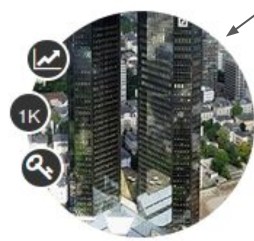
… and many more!

AI goal: language understanding

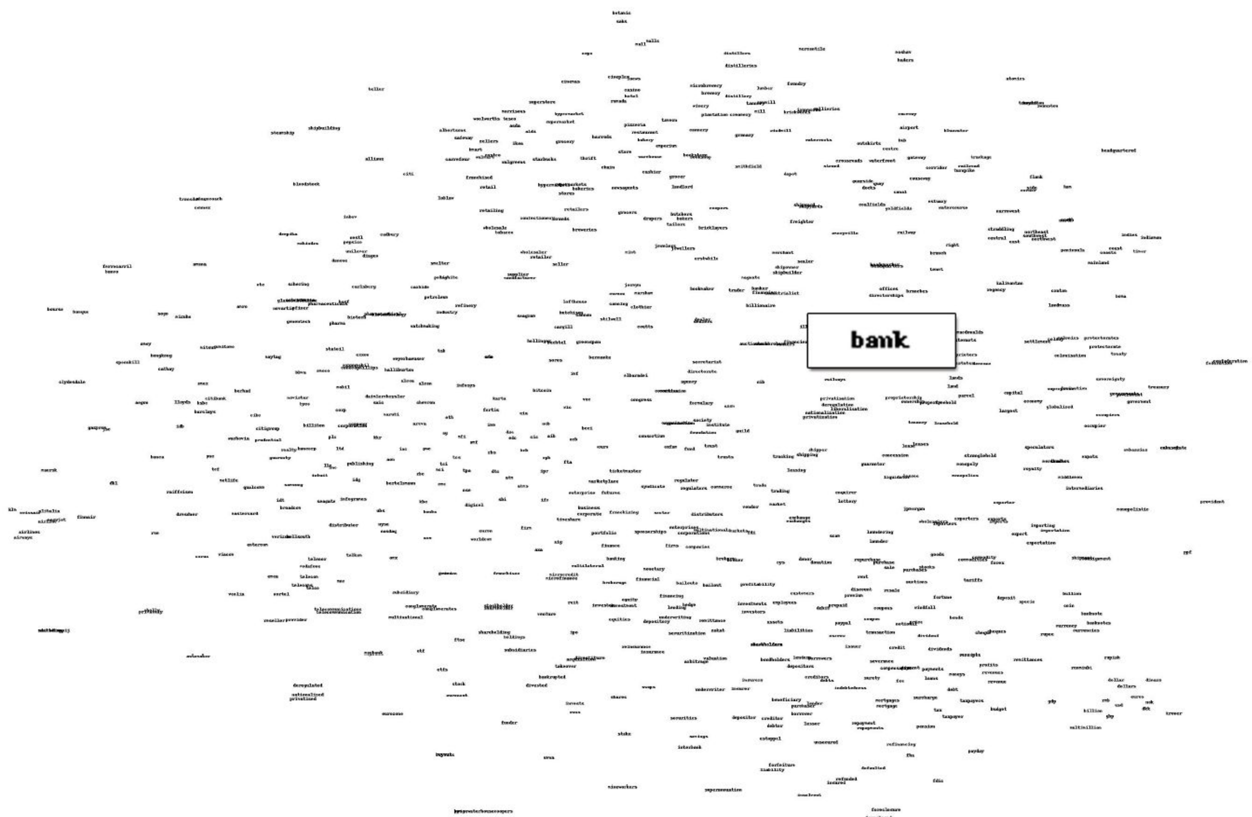


Limitations of word embeddings

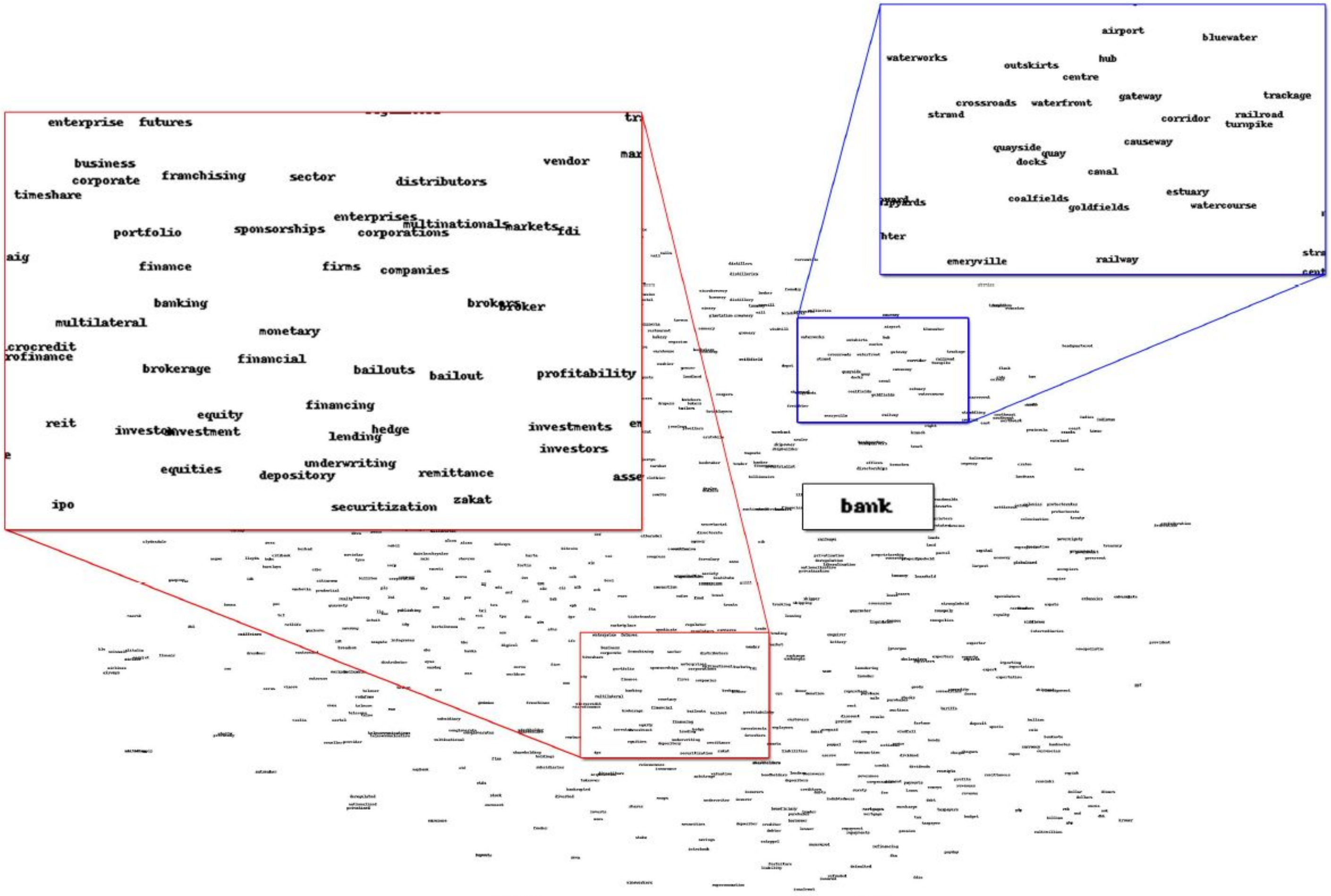
- Word representations cannot capture ambiguity. For instance, <u>*bank*</u>



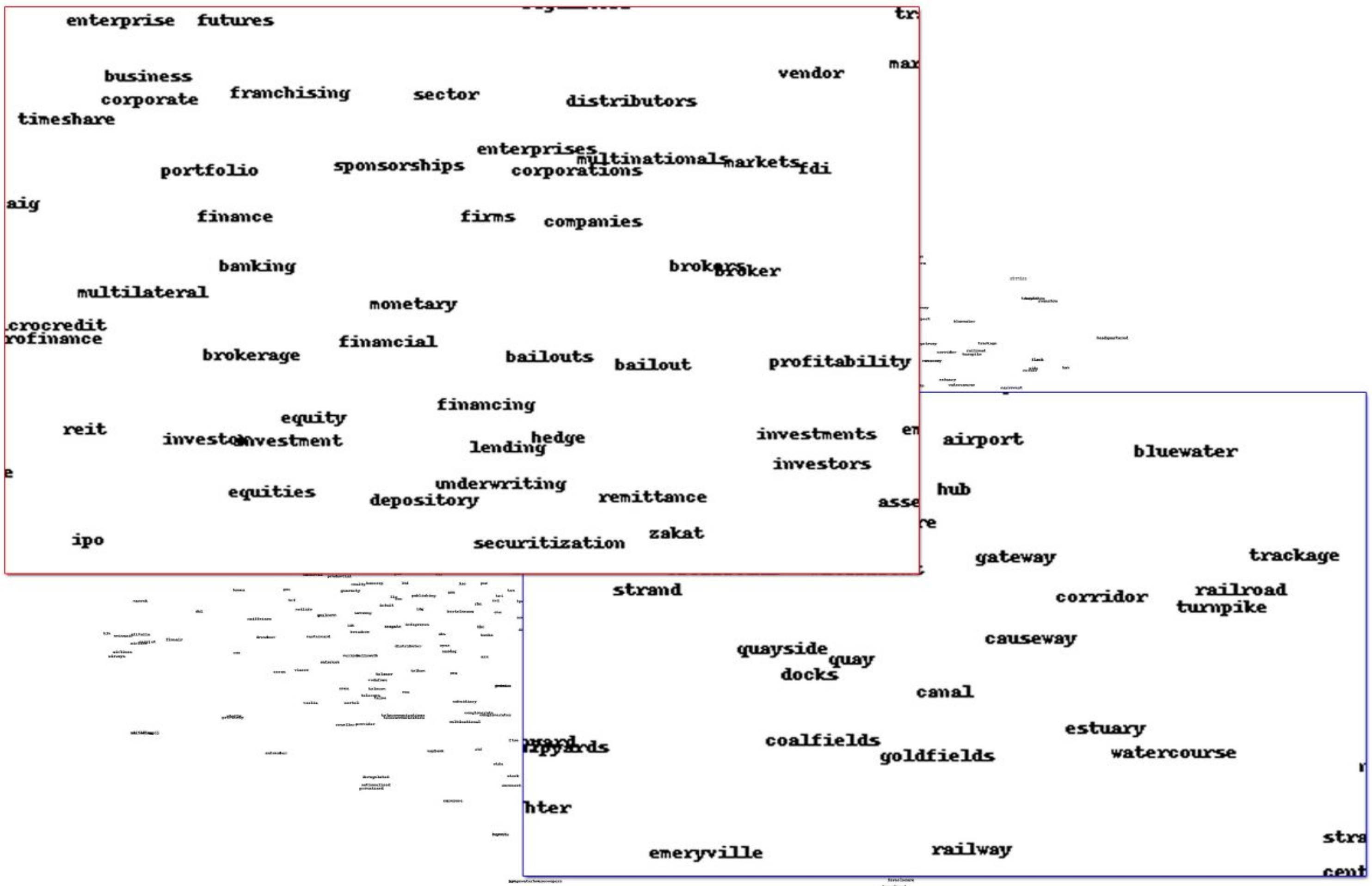
Problem 1:
word representations cannot capture ambiguity



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

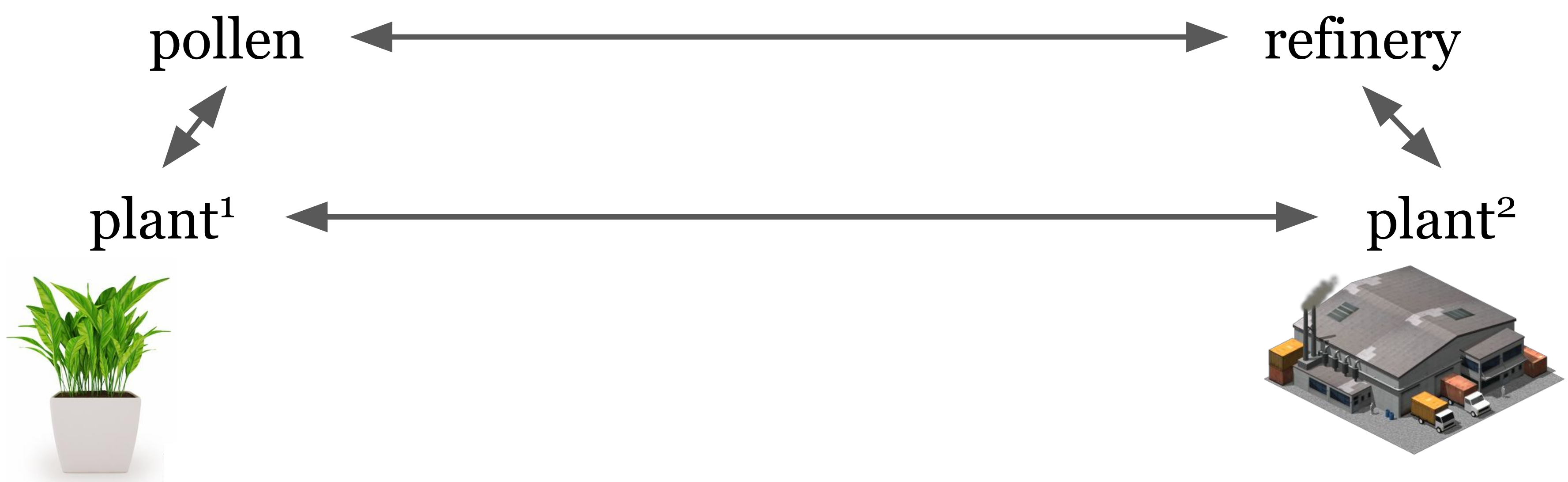
pollen refinery

plant


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

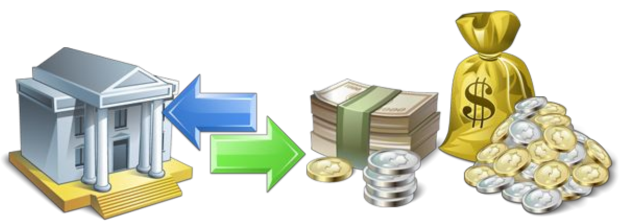
- Word representations do not exploit knowledge from existing lexical resources.

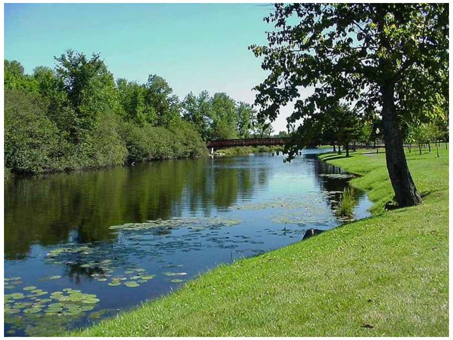
WIKIPEDIA
The Free Encyclopedia

BabelNet

Motivation: Model senses instead of only words

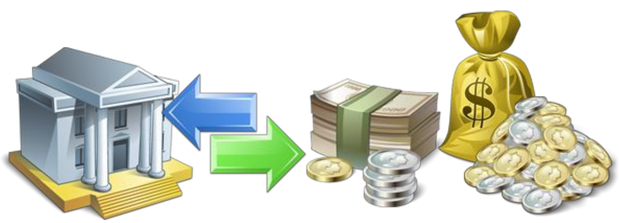
He withdrew money from the ***bank****.*



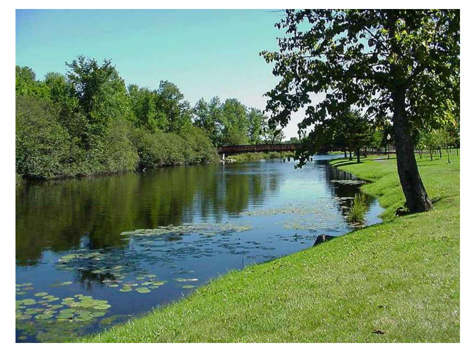


Motivation: Model senses instead of only words

He withdrew money from the ***bank***.



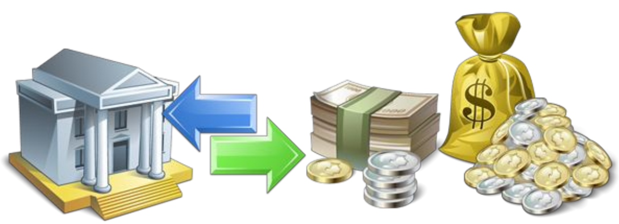
bank#1



bank#2

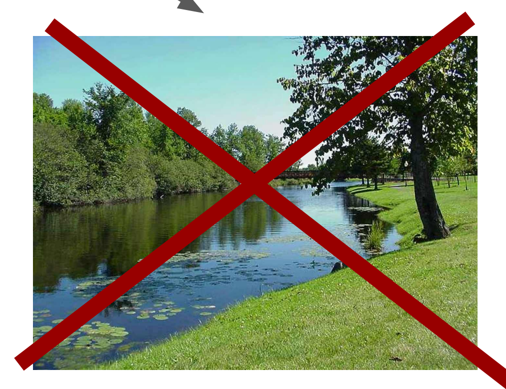
Motivation: Model senses instead of only words

*He withdrew money from the **bank**.*



bank#1

...



bank#2

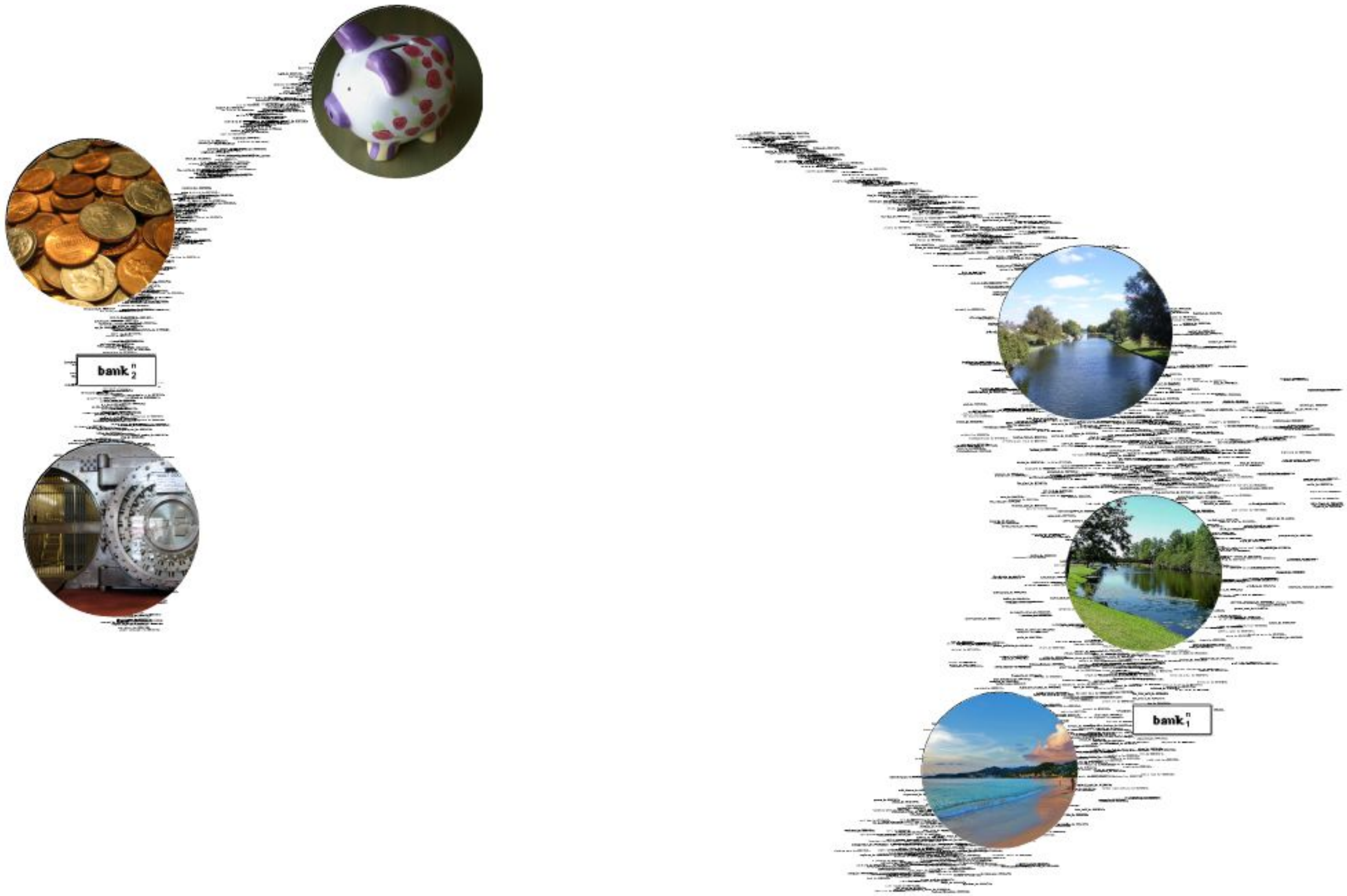
...

NASARI

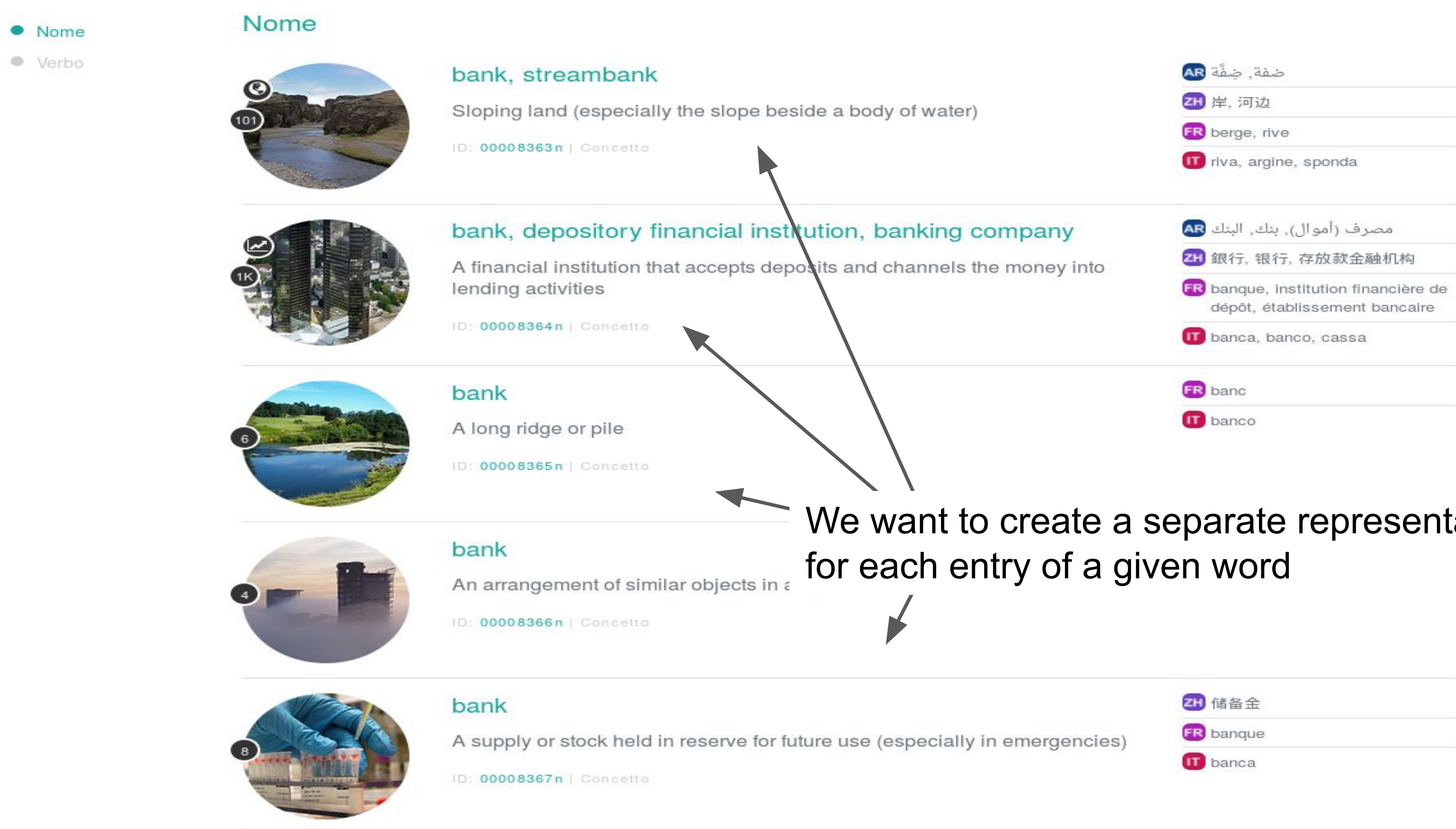
a Novel Approach to a Semantically-Aware Representations of Items

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

Key goal: obtain sense representations



Key goal: obtain sense representations



We want to create a separate representation for each entry of a given word

Idea

Encyclopedic knowledge

WIKIPEDIA
The Free Encyclopedia

Lexicographic knowledge

WordNet

Idea

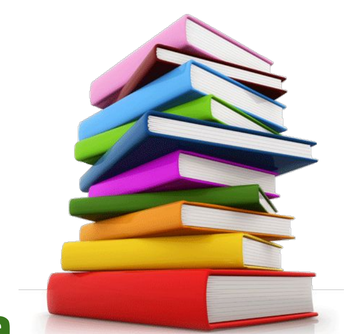
Encyclopedic knowledge

WIKIPEDIA
The Free Encyclopedia

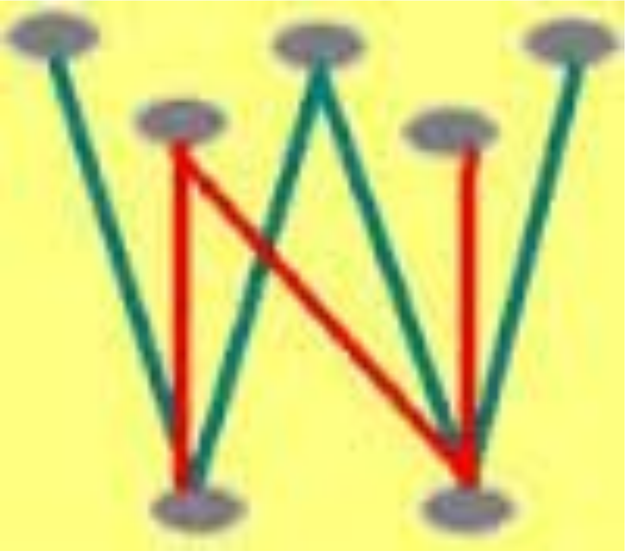
Lexicographic knowledge

WordNet

Information from text corpora

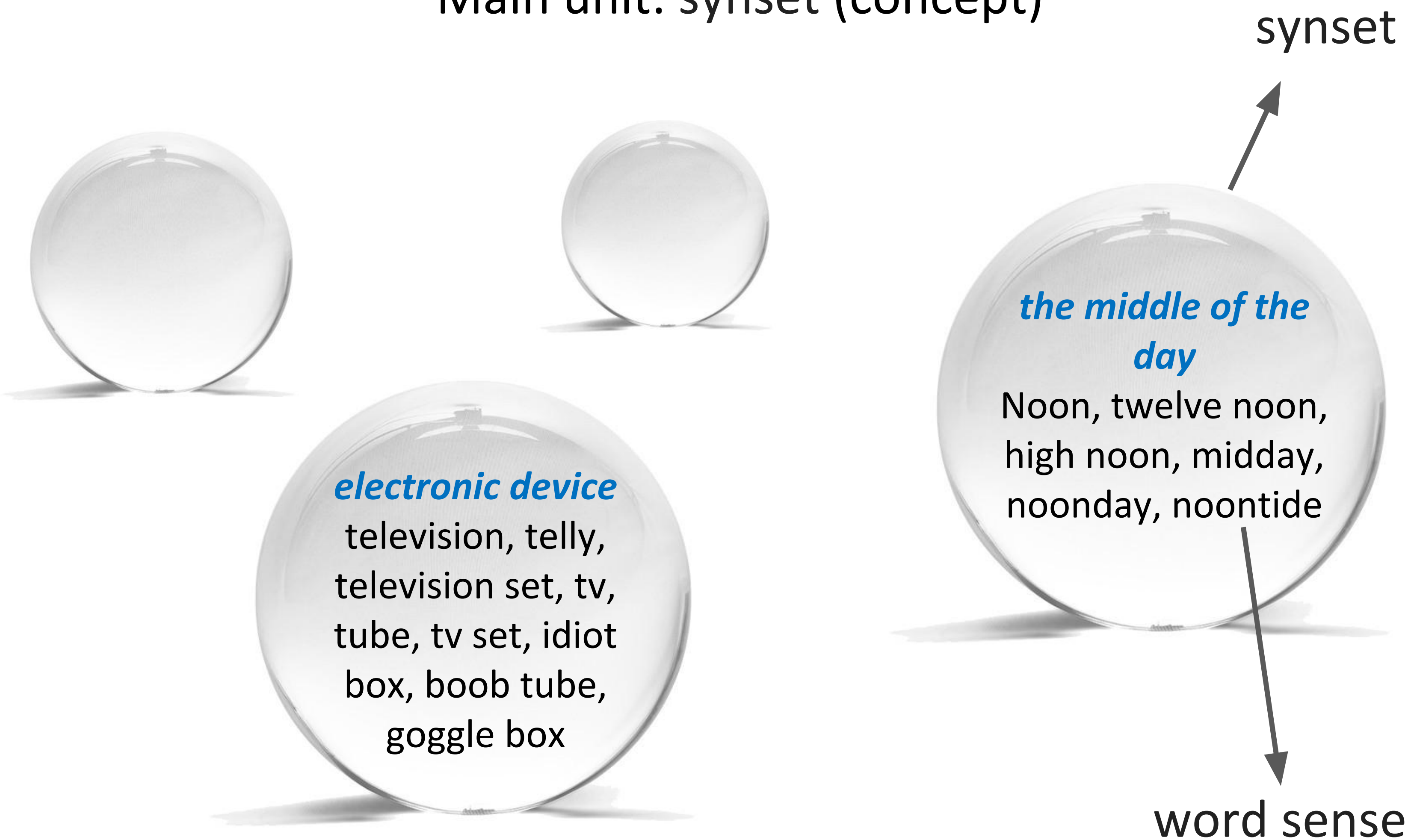


WordNet

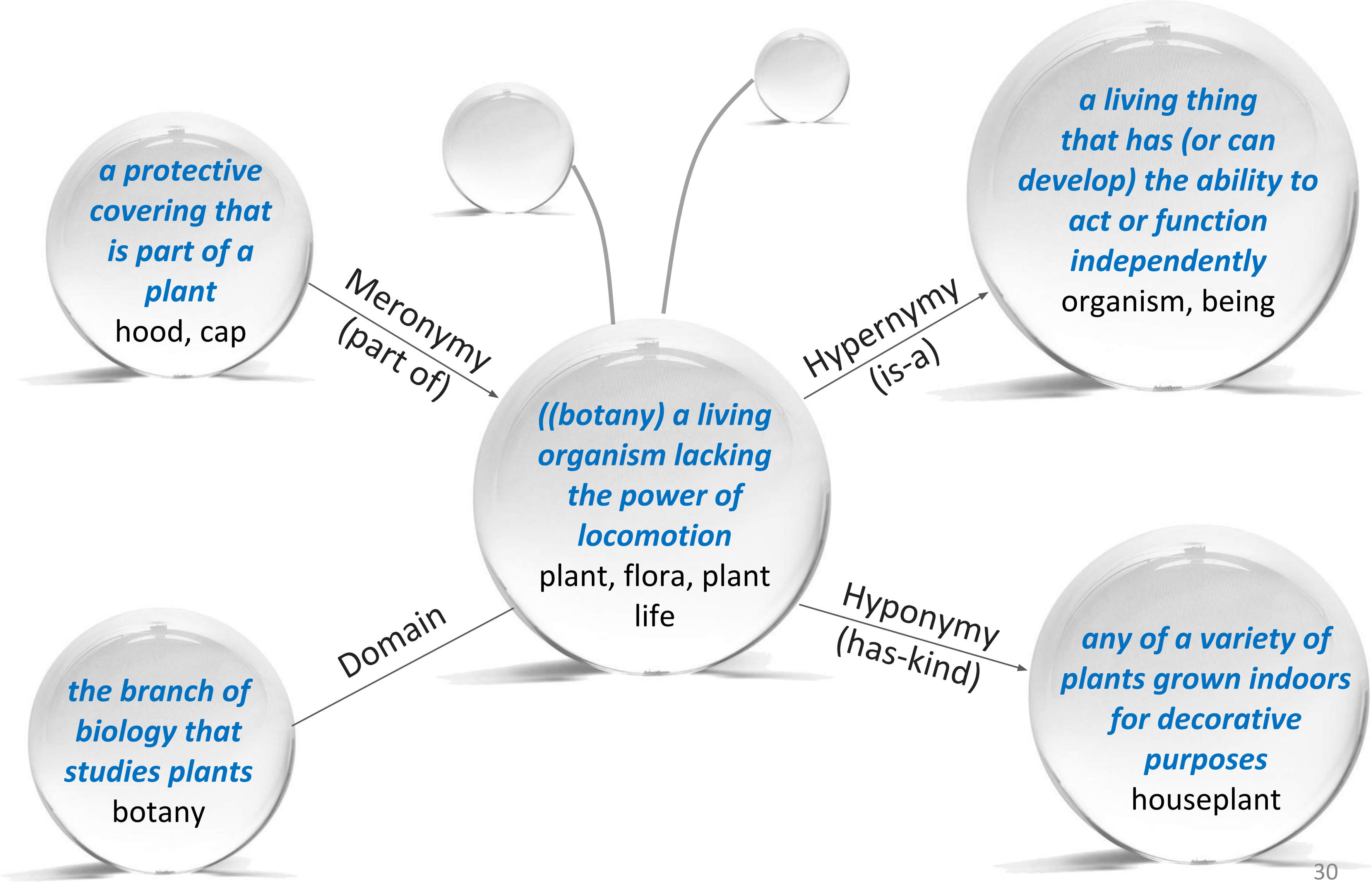


WordNet

Main unit: synset (concept)



WordNet semantic relations



Knowledge-based Representations (WordNet)

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)

☆ Faruqui, M., Dodge, J., Jauhar, S. K., Dyer, C., Hovy, E., & Smith, N. A. **Retrofitting Word Vectors to Semantic Lexicons** (NAACL 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
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page

Print/export
Create a book

Article Talk Read Edit View history Search

University of California, Los Angeles

From Wikipedia, the free encyclopedia

Coordinates: 34°04′20.00″N 118°26′38.75″W

"UCLA", "Ucla", and "U.C.L.A." redirect here. For other uses, see UCLA (disambiguation).

The **University of California, Los Angeles** (**UCLA**) is a public research university located in the Westwood neighborhood of Los Angeles, California, United States. It became the University of California Southern Branch in 1919, making it the second-oldest undergraduate campus of the ten-campus system after the original University of California campus in Berkeley (1873).[11] It offers 337 undergraduate and graduate degree programs in a wide range of disciplines.[12] With an approximate enrollment of 30,000 undergraduate and 12,000 graduate students, UCLA has the highest enrollment of any university in California[6] and is the most applied to university in the United States with over 112,000 applications for fall 2015.[13]

The university is organized into five undergraduate colleges, seven professional schools, and four professional health science schools. The undergraduate colleges are the College of Letters and Science; Henry Samueli School of Engineering and Applied Science (HSSEAS); School of the Arts and Architecture; School of Theater, Film, and Television; and School of Nursing. Fifteen[14][15] Nobel laureates, one Fields Medalist,[16] and three Turing Award winners[17] have been faculty, researchers, or alumni. Among the current faculty members, 55 have been elected to the National Academy of Sciences, 28 to the National Academy of Engineering, 39 to the Institute of Medicine, and 124 to the American Academy of Arts and Sciences.[18] The university was elected to the Association of American Universities in 1974.[19]

UCLA student-athletes compete as the Bruins in the Pacific-12 Conference. The Bruins have won 125 national championships, including 112 NCAA team championships.[20][21] UCLA student-athletes have won 250 Olympic medals: 125 gold, 65 silver and 60 bronze.[22] The Bruins have competed in every Olympics since 1920 with one exception (1924), and have won a gold medal in every Olympics that the United States has participated in since 1932.[23]

Contents [hide]
1 History

University of California, Los Angeles

UCLA official seal

Former names	State Normal School at Los Angeles (1882-1919) University of California Southern Branch (1919–1927) University of California at Los Angeles (1927–1958)
Motto	*Fiat lux* (Latin)
Motto in English	Let there be light

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

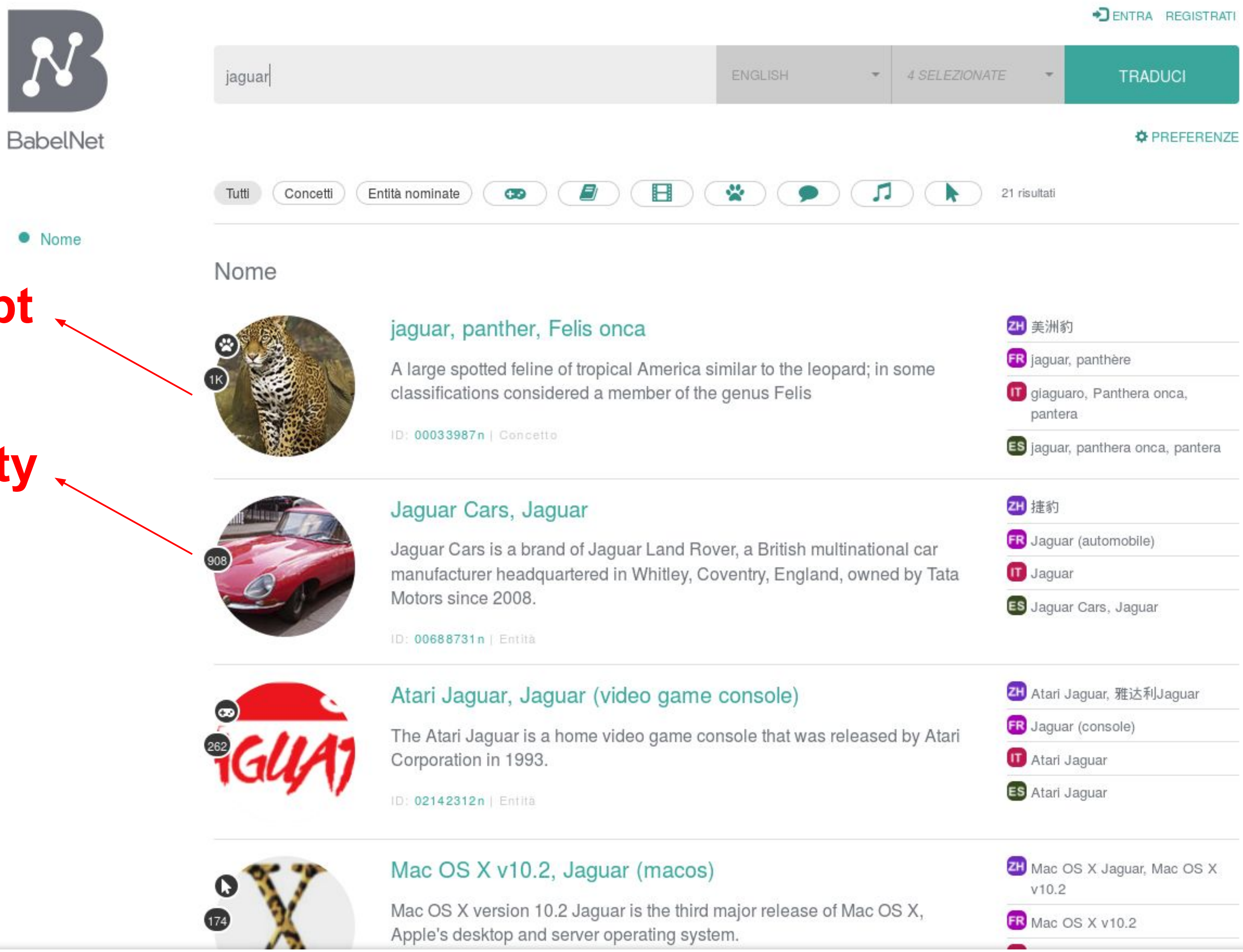
Thanks to an automatic mapping algorithm, **BabelNet integrates Wikipedia and WordNet**, among other resources (Wiktionary, OmegaWiki, WikiData…).

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

BabelNet (https://babelnet.org/)

Concept

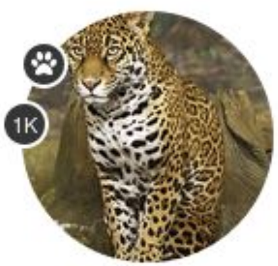
Entity



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

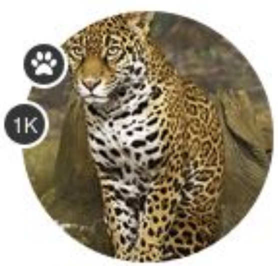
IT giaguaro, Panthera onca, pantera

ES jaguar, panthera onca, pantera

BabelNet

In this case, **synsets are multilingual**

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

NASARI

(Camacho-Collados et al., AIJ 2016)

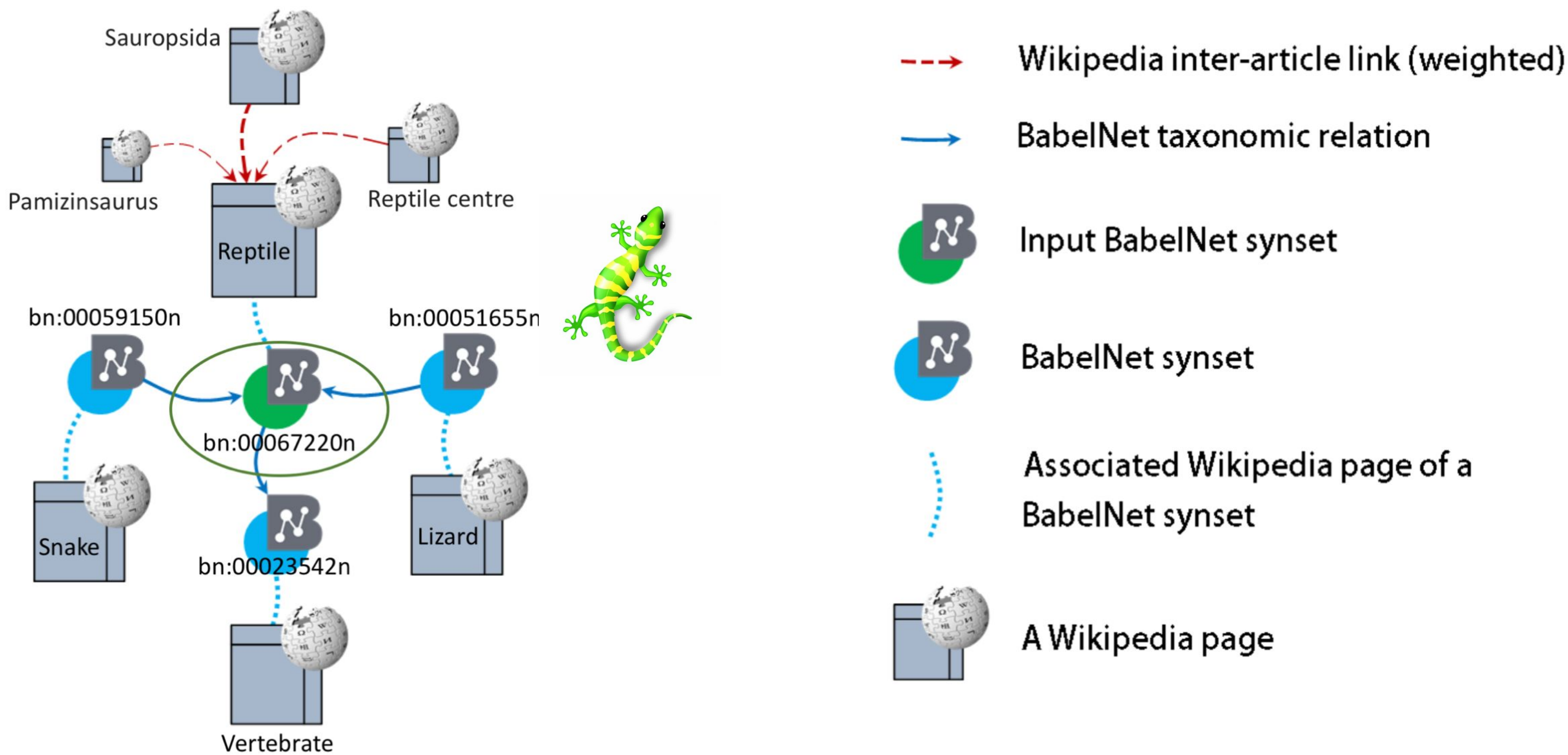
Goal

Build vector representations for multilingual BabelNet synsets.

How?

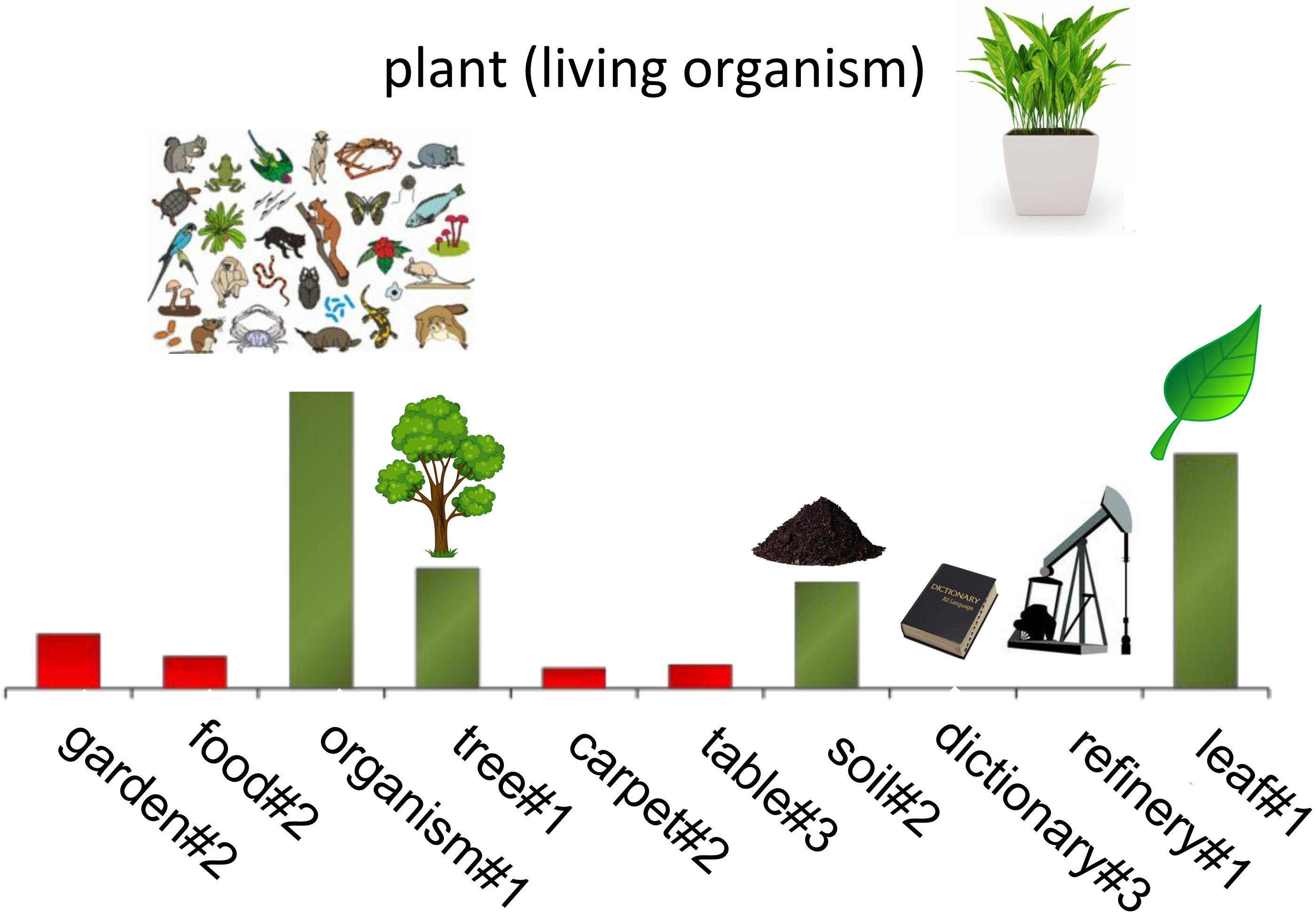
We exploit **Wikipedia semantic network** and **WordNet taxonomy** to construct a subcorpus (contextual information) for any given BabelNet synset.

Pipeline



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

Human-interpretable dimensions



Embedded vector representation (low-dimensional)

Closest senses

Bank (financial institution)	
Closest senses	Cosine
Deposit account	0.99
Universal bank	0.99
British banking	0.98
German banking	0.98
Commercial bank	0.98
Banking in Israel	0.98
Financial institution	0.98
Community bank	0.97

Bank (geography)	
Closest senses	Cosine
Stream bed	0.98
Current (stream)	0.97
River engineering	0.97
Braided river	0.97
Fluvial terrace	0.97
Bar (river morphology)	0.97
River	0.97
Perennial stream	0.96

bank	
Closest senses	Cosine
Bank (financial institution)	0.86
Universal bank	0.86
British banking	0.86
German banking	0.85
Branch (banking)	0.85
McFadden Act	0.85
Four Northern Banks	0.84
State bank	0.84

SW2V: *Senses and Words to Vectors*

(Mancini and Camacho-Collados et al., CoNLL 2017)

A word is the surface form of a sense: we can exploit this intrinsic relationship for **jointly training word and sense embeddings**.

SW2V: *Senses and Words to Vectors*

(Mancini and Camacho-Collados et al., CoNLL 2017)

A word is the surface form of a sense: we can exploit this intrinsic relationship for **jointly training word and sense embeddings**.

How?

Updating the representation of the word and its associated senses interchangeably.

SW2V: Idea

Given as input a **corpus** and a **semantic network**:

1. Use a semantic network to link to each word its *associated senses in context*.

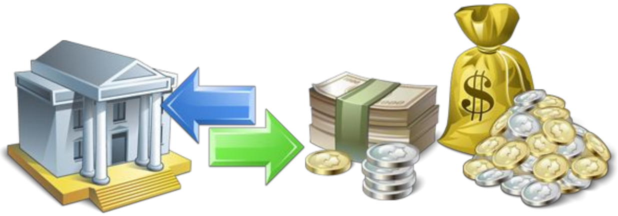
He withdrew money from the ***bank****.*

SW2V: Idea

Given as input a **corpus** and a **semantic network**:

1. Use a semantic network to link to each word its *associated senses in context*.

*He withdrew money from the **bank**.*



SW2V: Idea

Given as input a corpus and a semantic network:

1. Use a semantic network to link to each word its *associated senses in context*.

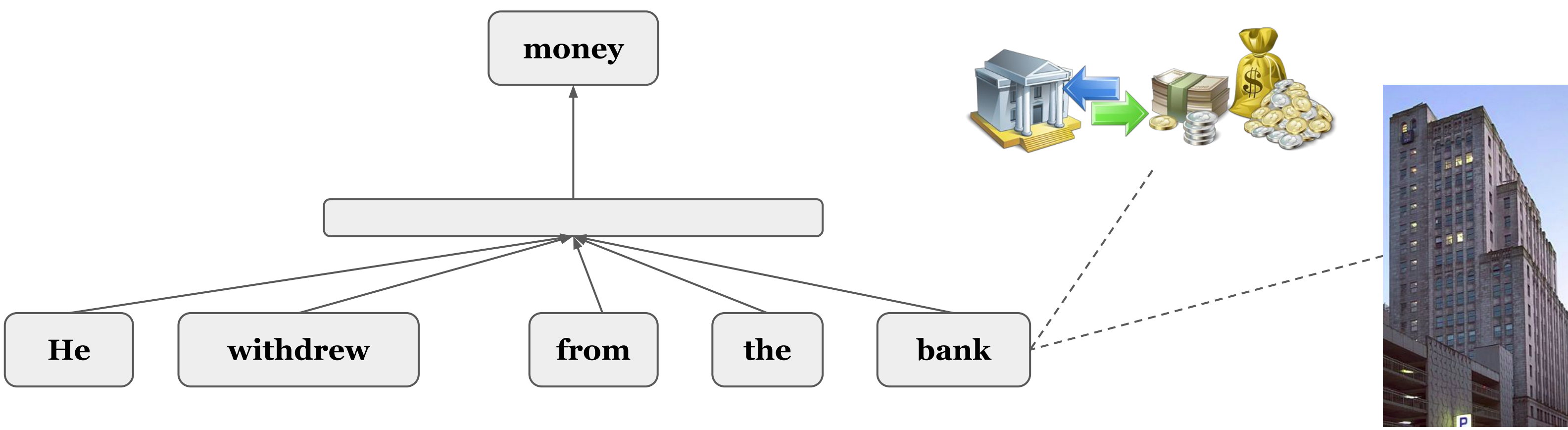
2. Use a **neural network** where the **update of word and sense embeddings is linked**, exploiting *virtual* connections.

SW2V: Idea

Given as input a corpus and a semantic network:

1. Use a semantic network to link to each word its *associated senses in context*.

2. Use a neural network where the update of word and sense embeddings is linked, exploiting *virtual* connections.

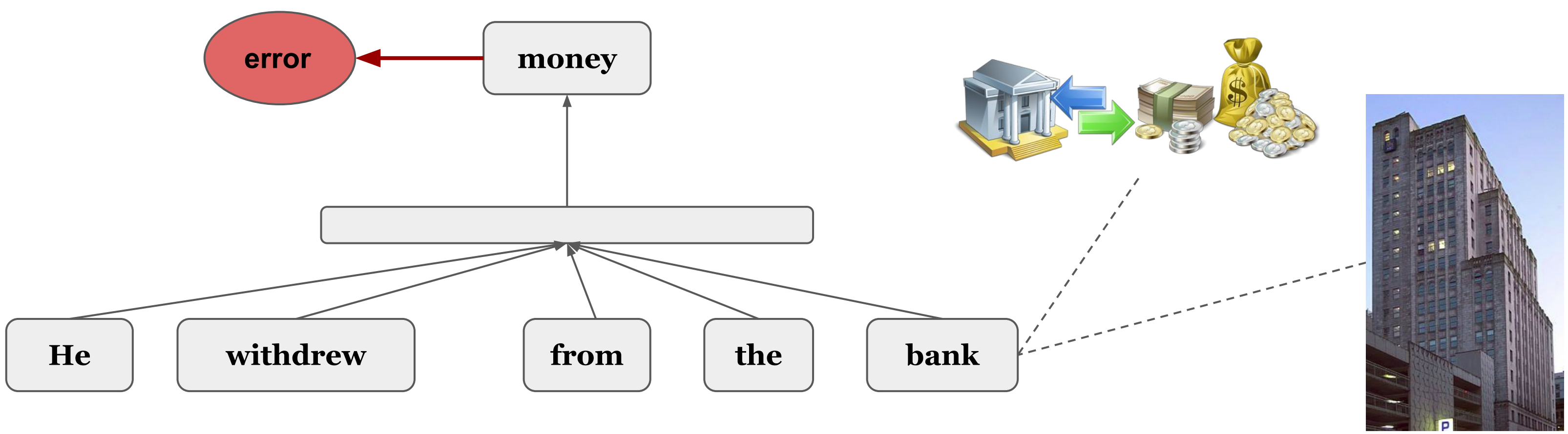


SW2V: Idea

Given as input a corpus and a semantic network:

1. Use a semantic network to link to each word its *associated senses in context*.

2. Use a neural network where the update of word and sense embeddings is linked, exploiting *virtual* connections.

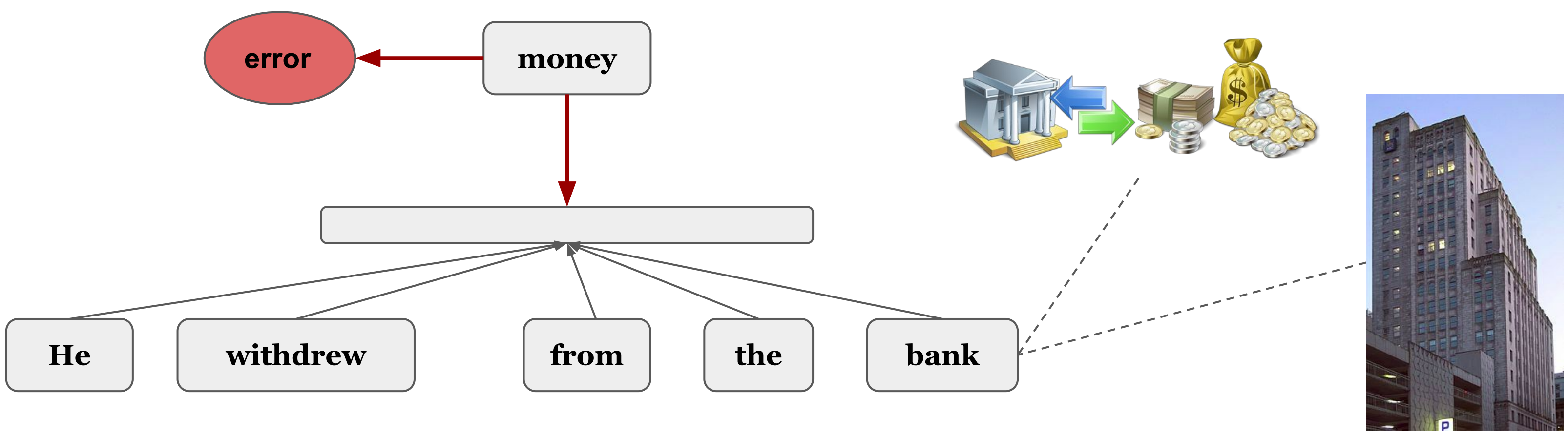


SW2V: Idea

Given as input a corpus and a semantic network:

1. Use a semantic network to link to each word its *associated senses in context*.

2. Use a neural network where the update of word and sense embeddings is linked, exploiting *virtual* connections.

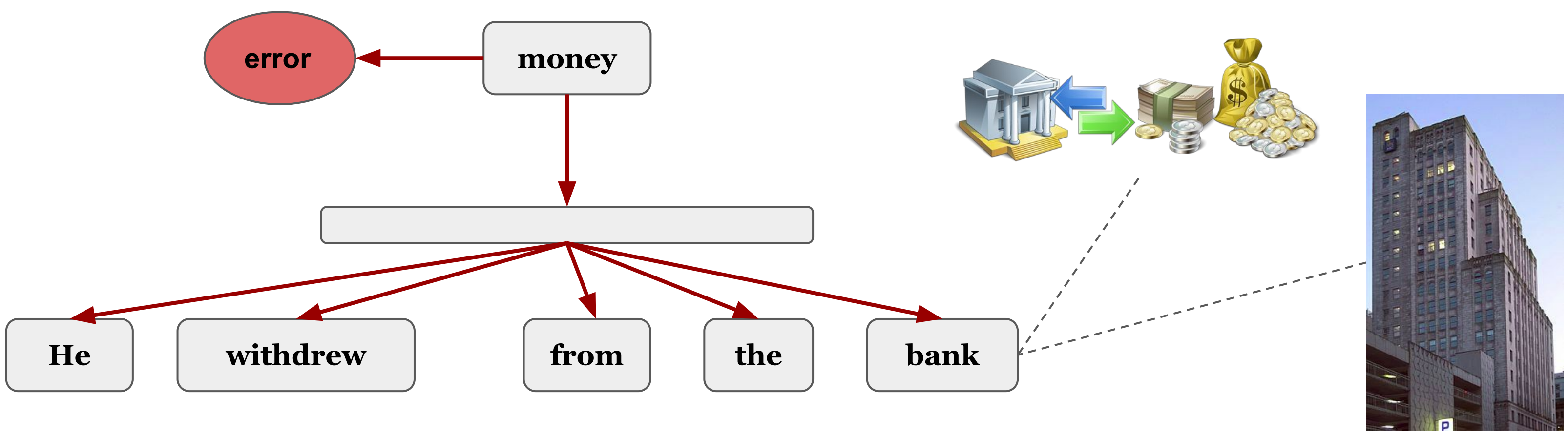


SW2V: Idea

Given as input a corpus and a semantic network:

1. Use a semantic network to link to each word its *associated senses in context*.

2. Use a neural network where the update of word and sense embeddings is linked, exploiting *virtual* connections.

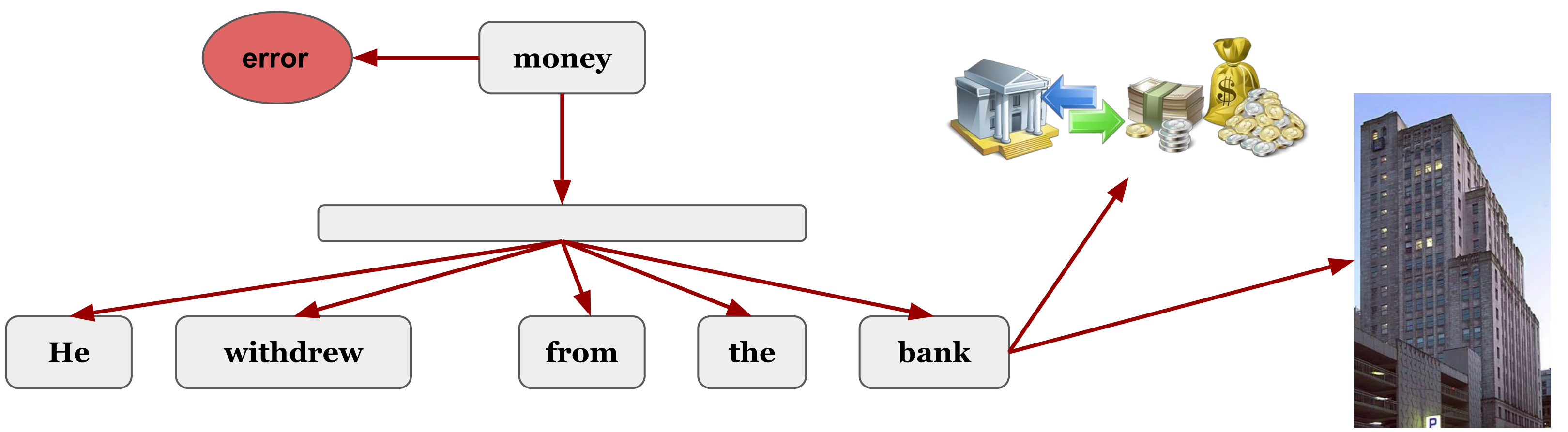


SW2V: Idea

Given as input a corpus and a semantic network:

1. Use a semantic network to link to each word its *associated senses in context*.

2. Use a neural network where the update of word and sense embeddings is linked, exploiting *virtual* connections.



SW2V: Idea

Given as input a corpus and a semantic network:

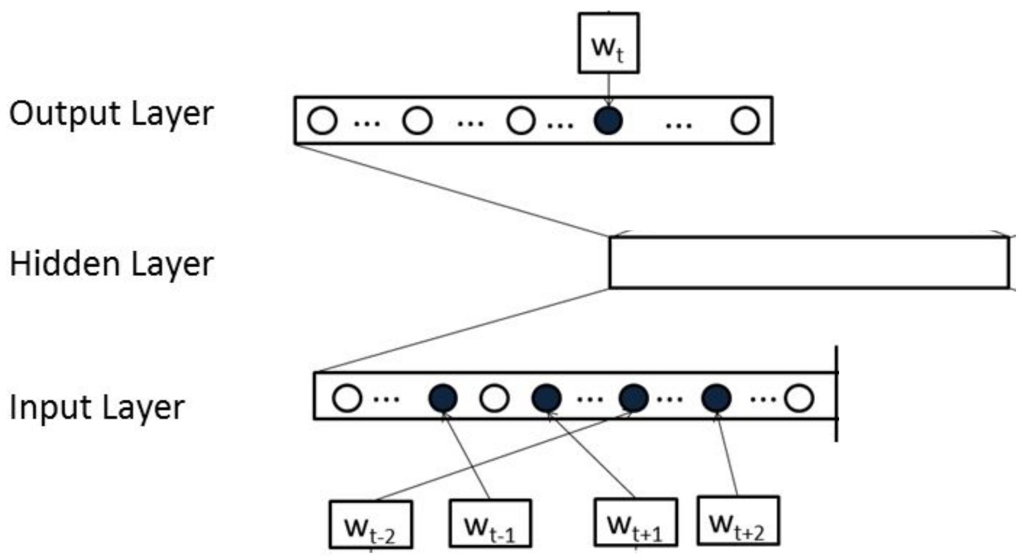
1. Use a semantic network to link to each word its *associated senses in context*.

2. Use a neural network where the update of word and sense embeddings is linked, exploiting *virtual* connections.

In this way it is possible to learn word and sense/synset embeddings jointly on a ***single training****.*

Full architecture of W2V (Mikolov et al., 2013)

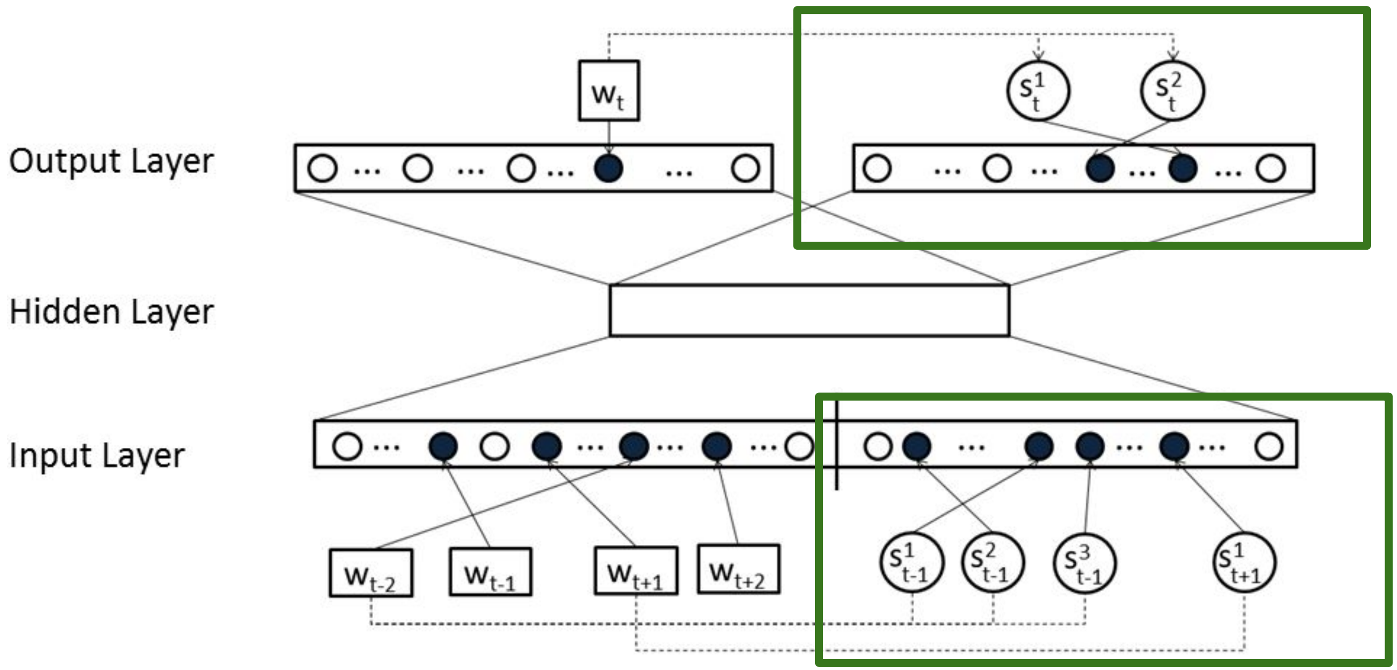
$$E=-\log(p(w_t|W^t))$$



Words and associated senses used both as input and output.

Full architecture of SW2V

$$E=-\log(p(w_t|W^t,\mathbf{S^t})) \ \mathbf{-\sum_{s\in S_t} \log(p(s|W^t,S^t))}$$



Words and associated senses used both as input and output.

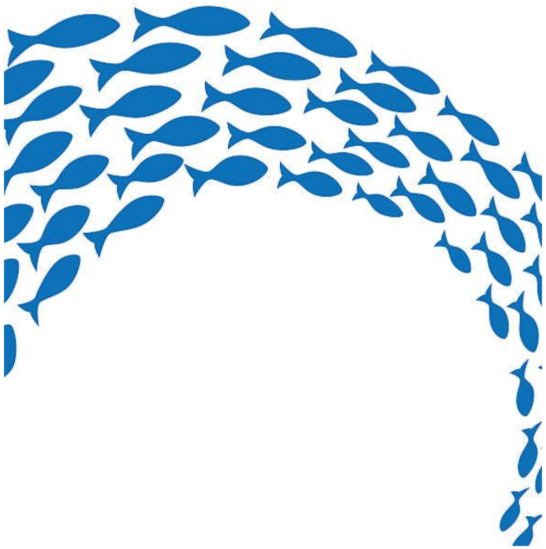
Word and senses connectivity: example 1

company$_n^2$ (military unit)

AutoExtend	**SW2V**
company$_n^9$	battalion$_n^1$
company	battalion
company$_n^8$	regiment$_n^1$
company$_n^6$	detachment$_n^4$
company$_n^7$	platoon$_n^1$
company$_v^1$	brigade$_n^1$
firm	regiment
business$_n^1$	corps$_n^1$
firm$_n^2$	brigade
company$_n^1$	platoon

Ten closest word and sense embeddings to the sense *company (military unit)*

Word and senses connectivity: example 2



$school^7_n$ (group of fish)

AutoExtend	SW2V
school	$schools^7_n$
$school^4_n$	$sharks^1_n$
$school^6_n$	sharks
$school^1_v$	$shoals^3_n$
$school^3_n$	$fish^1_n$
elementary	$dolphins^1_n$
schools	$pods^3_n$
$elementary^3_a$	eels
$school^5_n$	dolphins
$elementary^1_a$	$whales^2_n$

Ten closest word and sense embeddings to the sense *school (group of fish)*

Applications of knowledge-based sense representations

- **Taxonomy Learning** (Espinosa-Anke et al., EMNLP 2016)
- **Open Information Extraction** (Delli Bovi et al. EMNLP 2015).
- **Lexical entailment** (Nickel & Kiela, NIPS 2017)
- **Word Sense Disambiguation** (Tripodi & Navigli, EMNLP 2019)
- **Sentiment analysis** (Flekova & Gurevych, ACL 2016)
- **Lexical substitution** (Cocos et al., SENSE 2017)
- **Computer vision** (Young et al., ICRA 2017)
- **Text classification** (Sinoara et al., KB-Systems 2019)

…

Applications (in this talk)

- Domain labeling/adaptation
- Word Sense Disambiguation
- Downstream NLP applications (e.g. text classification)

Domain labeling

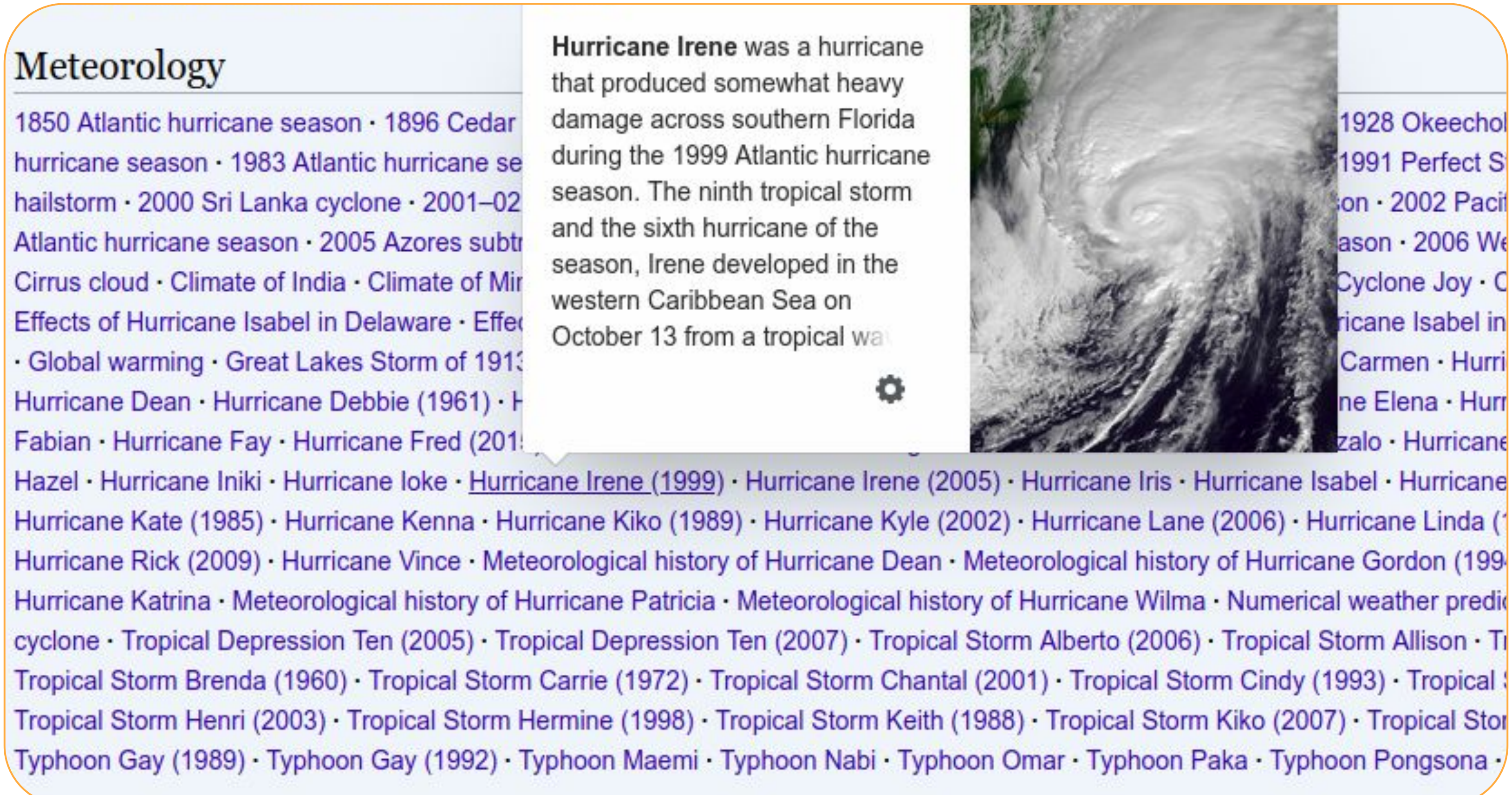
(Camacho-Collados and Navigli, EACL 2017)

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



Domain labeling

How to associate a concept with a domain?

1. Learn a **NASARI vector** for the concatenation of all Wikipedia pages associated with a given domain.

2. Exploit the **semantic similarity** between knowledge-based vectors and **graph properties** of the lexical resources.

BabelDomains

(Camacho-Collados and Navigli, EACL 2017)

As a result:

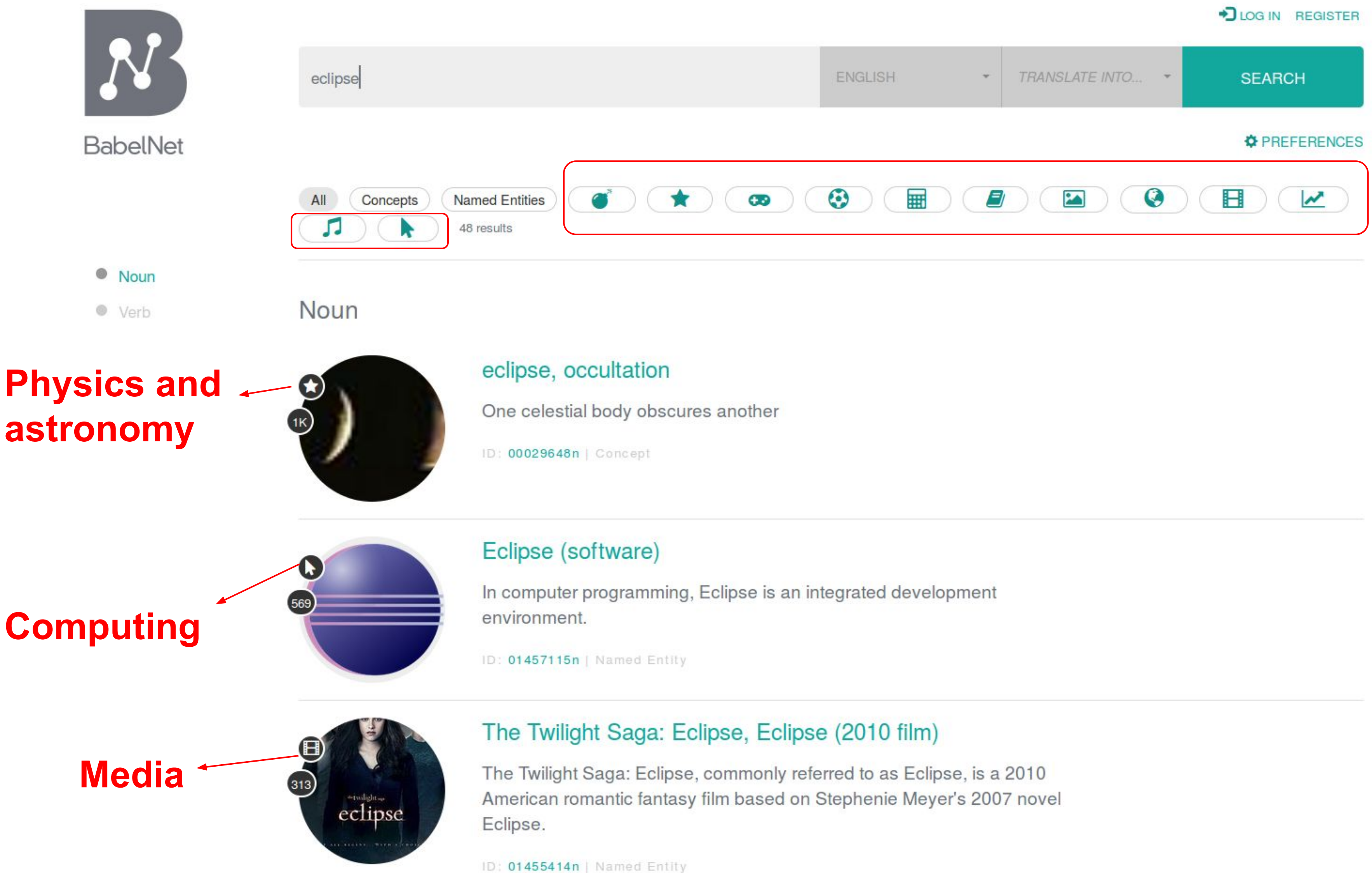
Unified resource with information about domains of knowledge

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

http://lcl.uniroma1.it/babeldomains

Already integrated into BabelNet (online interface and API)

BabelDomains



Word Sense Disambiguation

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

?

Word Sense Disambiguation

<u>Kobe</u>, 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**).

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

Word Sense Disambiguation on textual definitions

(Camacho-Collados et al., LREC 2016; LREV 2019)

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**.*

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

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

*Mano... d'**éche**...*

***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**.

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?

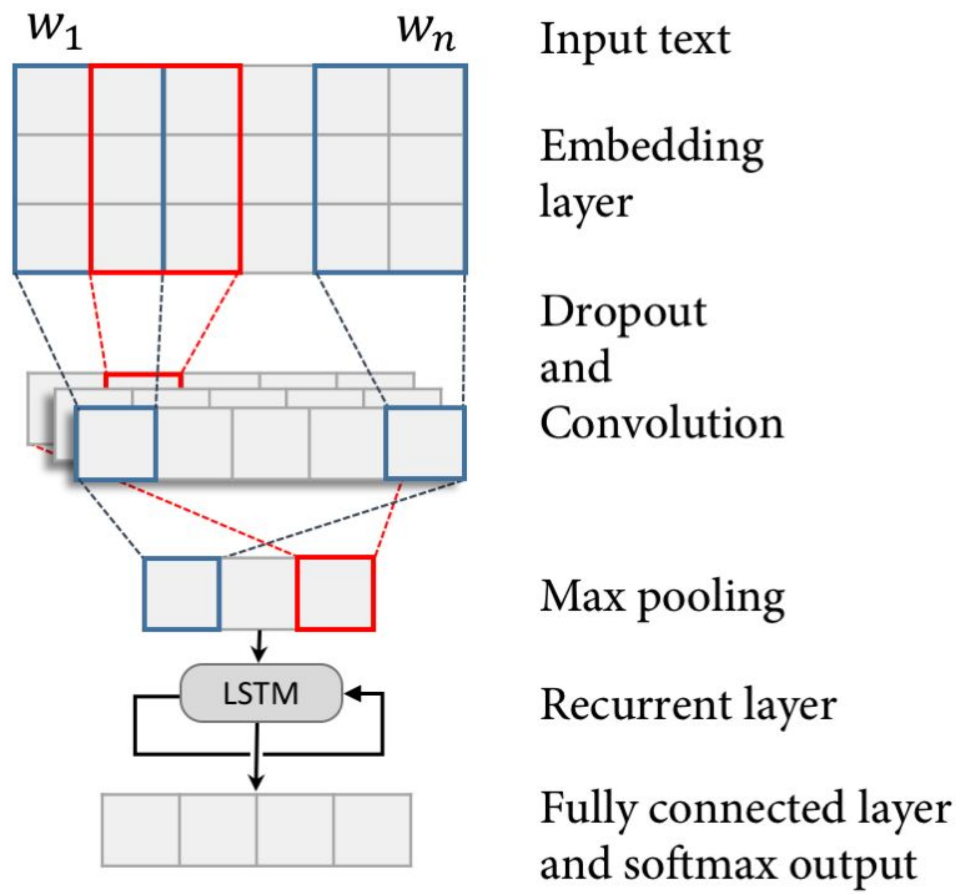
Tasks: Topic categorization and sentiment analysis (polarity detection)

Topic categorization: Given a text, assign it a topic (e.g. politics, sports, etc.).

Sentiment analysis: Predict the sentiment of the sentence/review as either positive or negative.

Classification model

Standard CNN classifier inspired by Kim (2014)

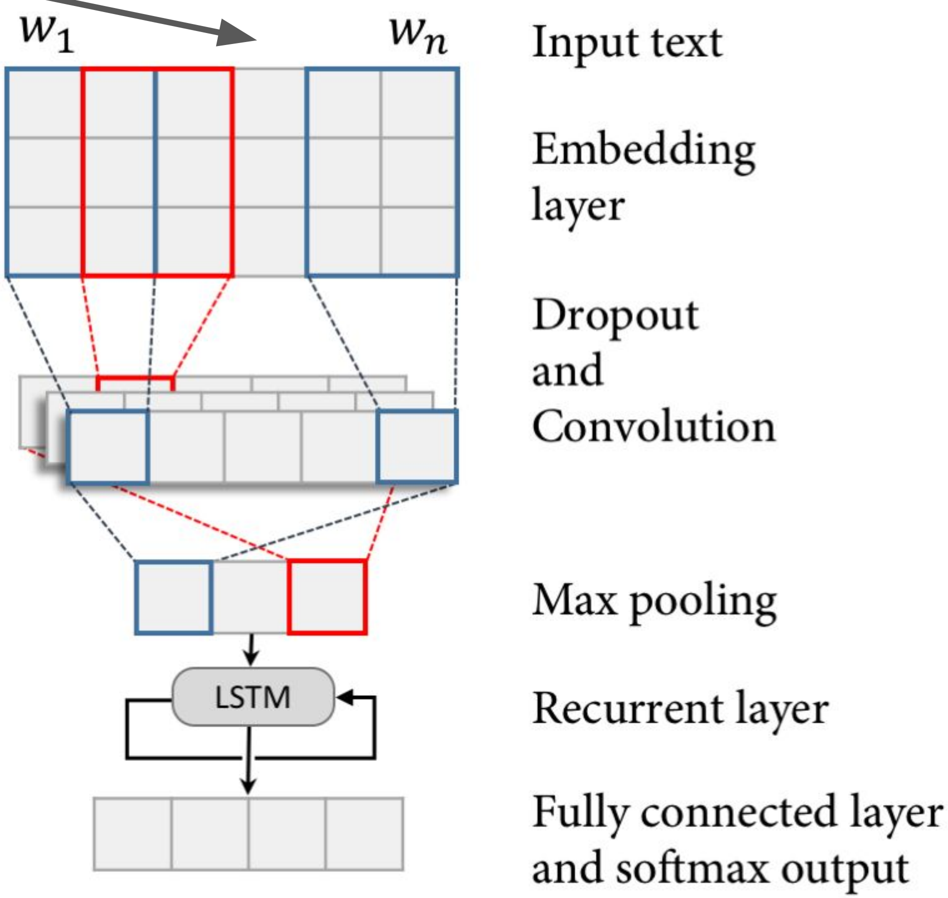


Classification model

Inject senses
embeddings as input!

Standard CNN classifier

inspired by Kim (2014)

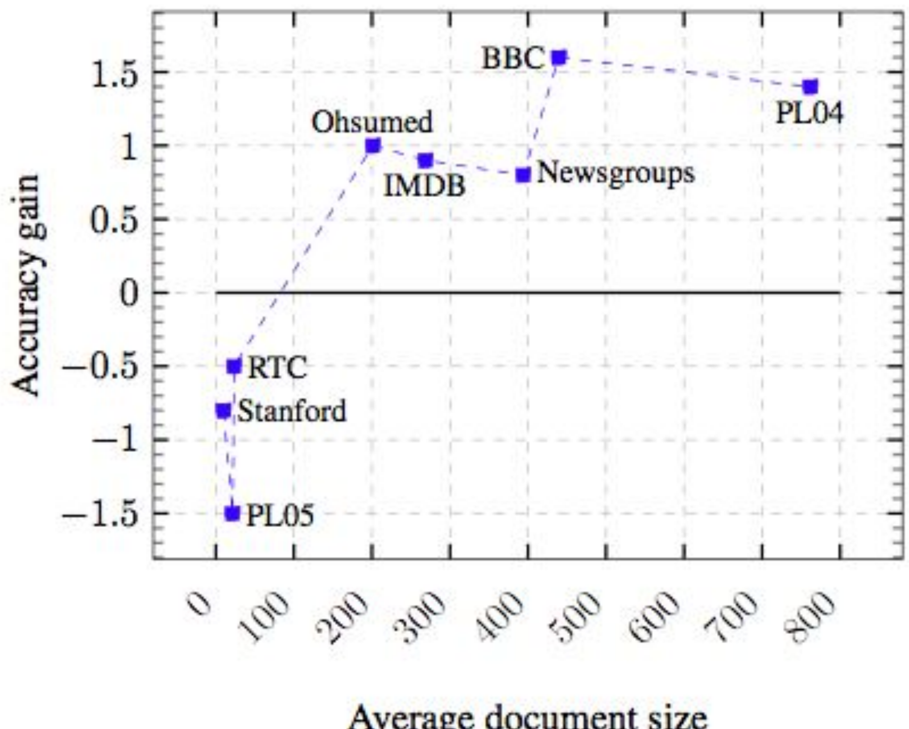


Sense-based vs. word-based: Conclusions

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**:



Contextualized word embeddings

ELMo

BERT

Peters et al.
(NAACL 2018)

Devlin et al.
(NAACL 2019)

Contextualized word embeddings

ELMo

Peters et al.
(NAACL 2018)

Based on LSTMs

BERT

Devlin et al.
(NAACL 2019)

Based on Transformers

Contextualized word embeddings

ELMo

Peters et al.
(NAACL 2018)

Based on LSTMs

BERT

Devlin et al.
(NAACL 2019)

Based on Transformers

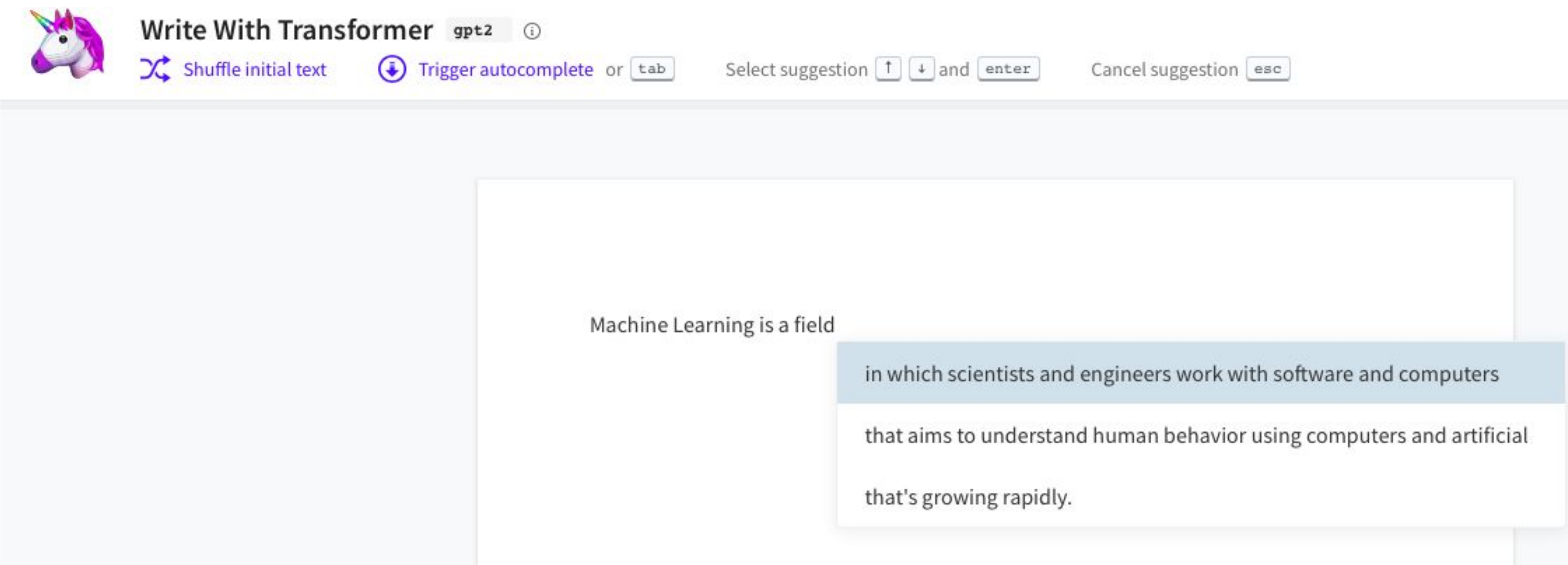
More successful nowadays

Contextualized word embeddings
ELMo/BERT

New AI fake text generator may be too dangerous to release, say creators

The Elon Musk-backed nonprofit company OpenAI declines to release research publicly for fear of misuse

Play with Transformers
Text generation



https://transformer.huggingface.co

Contextualized word embeddings ELMo/BERT

As word embeddings, learned by leveraging language models on **massive amounts of text corpora**.

Contextualized word embeddings
ELMo/BERT

As word embeddings, learned by leveraging language models on **massive amounts of text corpora**.

New: each word vector depends on the context. It is **dynamic**.

Contextualized word embeddings
ELMo/BERT

As word embeddings, learned by leveraging language models on **massive amounts of text corpora**.

New: each word vector depends on the context. It is **dynamic**.

Important **improvements in many NLP tasks**.

Contextualized word embeddings ELMo/BERT (examples)

He withdrew money from the ***bank****.*

The ***bank*** *remained closed yesterday.*

We found a nice spot by the ***bank*** *of the river.*

Contextualized word embeddings ELMo/BERT (examples)

0.25, 0.32, -0.1 ….

*He withdrew money from the **bank**.*

0.22, 0.30, -0.08 ….

*The **bank** remained closed yesterday.*

-0.8, 0.01, 0.3 ….

*We found a nice spot by the **bank** of the river.*

Contextualized word embeddings ELMo/BERT (examples)

Similar vectors

0.25, 0.32, -0.1 ….

*He withdrew money from the **bank**.*

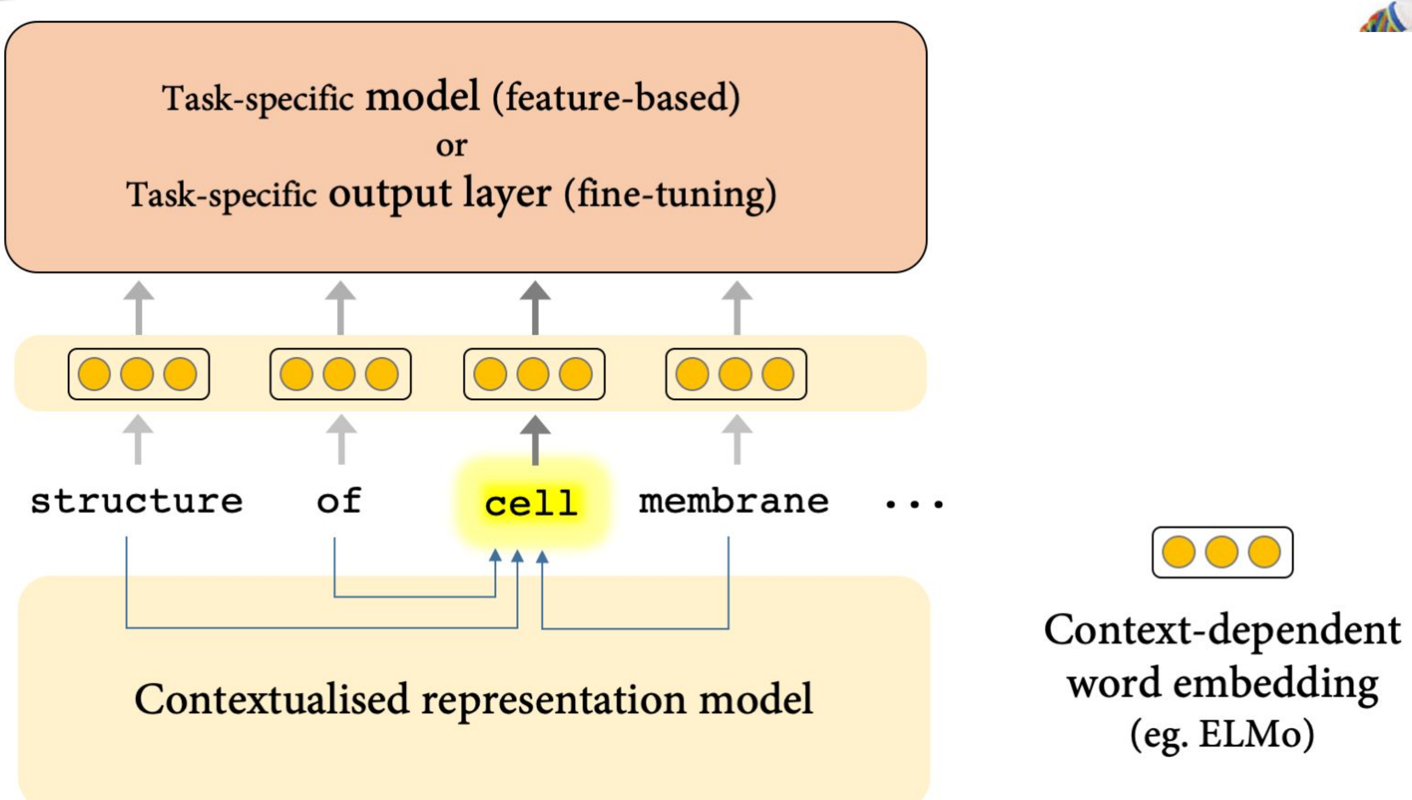
0.22, 0.30, -0.08 ….

*The **bank** remained closed yesterday.*

-0.8, 0.01, 0.3 ….

*We found a nice spot by the **bank** of the river.*

Contextualized word embeddings in applications



How well do these models capture “meaning”?

(Bouraoui et al. AAAI 2020)

Sentence	BERT
The color of the banana is ___.	yellow
The color of the avocado is ___.	yellow
The color of the carrot is ___.	yellow
The color of the tomato is ___.	white
The color of the kiwi is ___.	white
The capital of Japan is ___.	tokyo
The capital of France is ___.	paris
The capital of Australia is ___.	canberra
The capital of the US is ___.	washington
The capital of Brazil is ___.	santos
Recessions are caused by ___.	inflation
Recessions are often caused by ___.	stress
Hangovers are caused by ___.	stress
I took my umbrella because it was ___.	warm
He didn’t go to school because it was a ___.	secret
I like to have ___ for breakfast.	them
Her favorite subject in school was ___.	english
His favorite day of the week is ___.	christmas
They saw lots of scary animals such as ___.	bears
He likes ___ and most other vegetables.	potatoes

How well do these models capture “meaning”?

(Bouraoui et al. AAAI 2020)

Sentence	BERT
The color of the banana is ___.	yellow
The color of the avocado is ___.	yellow
The color of the carrot is ___.	yellow
The color of the tomato is ___.	white
The color of the kiwi is ___.	white
The capital of Japan is ___.	tokyo
The capital of France is ___.	paris
The capital of Australia is ___.	canberra
The capital of the US is ___.	washington
The capital of Brazil is ___.	santos
Recessions are caused by ___.	inflation
Recessions are often caused by ___.	stress
Hangovers are caused by ___.	stress
I took my umbrella because it was ___.	warm
He didn’t go to school because it was a ___.	secret
I like to have ___ for breakfast.	them
Her favorite subject in school was ___.	english
His favorite day of the week is ___.	christmas
They saw lots of scary animals such as ___.	bears
He likes ___ and most other vegetables.	potatoes

How well do these models capture “meaning”?

(Bouraoui et al. AAAI 2020)

Sentence	BERT
The color of the banana is ___.	yellow
The color of the avocado is ___.	yellow
The color of the carrot is ___.	yellow
The color of the tomato is ___.	white
The color of the kiwi is ___.	white
The capital of Japan is ___.	tokyo
The capital of France is ___.	paris
The capital of Australia is ___.	canberra
The capital of the US is ___.	washington
The capital of Brazil is ___.	santos
Recessions are caused by ___.	inflation
Recessions are often caused by ___.	stress
Hangovers are caused by ___.	stress
I took my umbrella because it was ___.	warm
He didn’t go to school because it was a ___.	secret
I like to have ___ for breakfast.	them
Her favorite subject in school was ___.	english
His favorite day of the week is ___.	christmas
They saw lots of scary animals such as ___.	bears
He likes ___ and most other vegetables.	potatoes

How well do these models capture “meaning”?

Good enough for many applications.

How well do these models capture “meaning”?

Good enough for many applications.

GLUE
Language understanding benchmark

Human baselines!

	Rank	Name	Model	URL	Score
	1	ERNIE Team - Baidu	ERNIE		90.1
	2	Microsoft D365 AI & MSR AI & GATECH	MT-DNN-SMART		89.9
	3	T5 Team - Google	T5		89.7
+	4	王玮	ALICE v2 large ensemble (Alibaba DAMO NLP)		89.5
	5	XLNet Team	XLNet (ensemble)		89.5
	6	ALBERT-Team Google Language	ALBERT (Ensemble)		89.4
	7	Microsoft D365 AI & UMD	FreeLB-RoBERTa (ensemble)		88.8
	8	Facebook AI	RoBERTa		88.5
	9	Junjie Yang	HIRE-RoBERTa		88.3
+	10	Microsoft D365 AI & MSR AI	MT-DNN-ensemble		87.6
	11	GLUE Human Baselines	GLUE Human Baselines		87.1

How well do these models capture “meaning”?

Good enough for many applications.

Room for improvement. For example, in SuperGLUE:

➢ *Winograd Schema Challenge*: BERT ~65% **vs** Humans ~95%

➢ *Word-in-Context (WiC) Challenge*: BERT ~68% **vs** Humans ~80%

How well do these models capture “meaning”?

Good enough for many applications.

Room for improvement. For example, in SuperGLUE:

- ➢ *Winograd Schema Challenge*: BERT ~65% **vs** Humans ~95%

  requires commonsense reasoning

- ➢ *Word-in-Context (WiC) Challenge*: BERT ~68% **vs** Humans ~80%

  requires abstracting the notion of sense

Word-in-Context (WiC) Challenge

(Pilehvar and Camacho-Collados, NAACL 2019)

Task: Identify the most suitable meaning of a word in context
Framed as binary classification (True/False)

Examples:

There's a lot of trash on the ***bed*** *of the river*
I keep a glass of water next to my ***bed*** *when I sleep* → **False**

He cashed a check at the ***bank***
The ***bank*** *is on the corner of Nassau and Witherspoon* → **True**

WiC competition open in CodaLab:
https://competitions.codalab.org/competitions/20010

WiC Challenge

But still…

WiC leaderboard

System	Implementation	Acc.
SenseBERT	Levine et al. (2019)	72.1
KnowBERT-W+W	Peters et al (2019)	70.9
RoBERTa	Liu et al. (2019)	69.9
BERT-large	SuperGLUE baseline	69.6
LMMS-WSD (BERT+)	Loureiro and Jorge (2019)	67.7
Ensemble (BERT+USE+ELMo)	Garí Soler et al. (2019)	66.7
BERT-large	WiC baseline	65.5

BERT is everywhere!

For more information on meaning representations (embeddings):

- ❖ ACL 2016 Tutorial on **“Semantic representations of word senses and concepts”:**http://josecamachocollados.com/slides/Slides_ACL16Tutorial_SemanticRepresentation.pdf
- ❖ EACL 2017 workshop on **“Sense, Concept and Entity Representations and their Applications”**: https://sites.google.com/site/senseworkshop2017/
- ❖ NAACL 2018 Tutorial on **“Interplay between lexical resources and NLP”:** https://bitbucket.org/luisespinosa/lr-nlp/
- ❖ Blog post on **“Word, Sense and Contextualized Embeddings”:** https://medium.com/@josecamachocollados/how-to-represent-meaning-in-natural-language-processing-word-sense-and-contextualized-embeddings-bbe31bdab84a
- ❖ **“From Word to Sense Embeddings: A Survey on Vector Representations of Meaning”** (JAIR, Dec 2018): https://www.jair.org/index.php/jair/article/view/11259

Thank you!

Questions please!

camachocolladosj@cardiff.ac.uk

[@CamachoCollados](https://twitter.com/CamachoCollados)

josecamachocollados.com