NAACL HLT

New Orleans, Louisiana | June 1-June 6, 2018

Tutorial on:

The interplay between lexical resources and Natural Language Processing

CARDIFF UNIVERSITY

UNIVERSITY OF CAMBRIDGE

Google Group: goo.gl/JEazYH

Luis Espinosa Anke

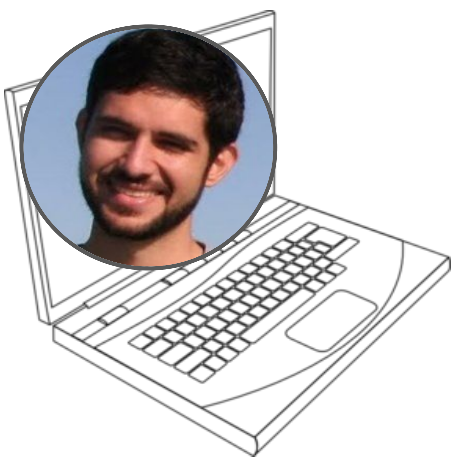
Jose Camacho-Collados

Mohammad Taher Pilehvar

Google Group: goo.gl/JEazYH

Luis Espinosa Anke (linguist)

Jose Camacho-Collados (mathematician)



Mohammad Taher Pilehvar (computer scientist) www.kahoot.it

PIN: 7024700

QUESTION 1
PIN: 7024700
www.kahoot.it

Outline

1. Introduction
2. Overview of Lexical Resources
3. **NLP for Lexical Resources**
4. **Lexical Resources for NLP**
5. Conclusion and Future Directions

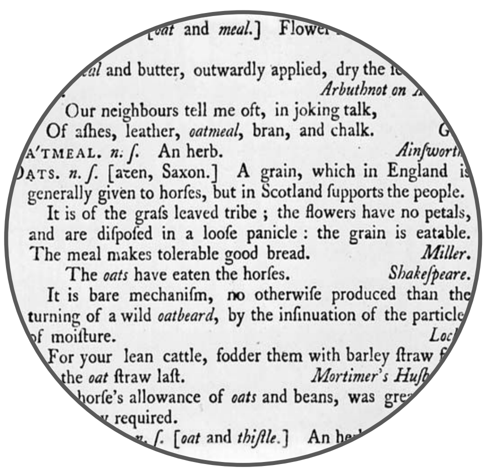
INTRODUCTION

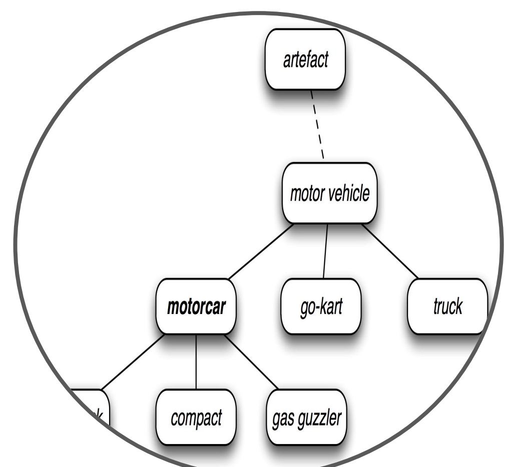
Introduction

- "A *lexical resource* (LR) is a database consisting of one or several dictionaries." (en.wikipedia.org/wiki/Lexical_resource)
- "What is *lexical resource*? In a word it is **vocabulary** and it matters for IELTS writing because ..." (dcielts.com/ielts-writing/lexical-resource)
- "The term *Language Resource* refers to a set of speech or language data and descriptions in machine readable form, used for ... " (elra.info/en/about/what-language-resource)

Introduction

- “A *lexical resource* (LR) is a database consisting of one or several dictionaries.” (en.wikipedia.org/wiki/Lexical_resource)
- “What is *lexical resource*? In a word it is **vocabulary** and it matters for IELTS writing because …” (dcielts.com/ielts-writing/lexical-resource)
- “The term *Language Resource* refers to a set of speech or language data and descriptions in machine readable form, used for … ” (elra.info/en/about/what-language-resource)





Introduction

Not straightforward definition of what a lexical resource actually is. Intuitively, a resource holding meaning of **words** (and their relations).

In addition, several commonalities such as:

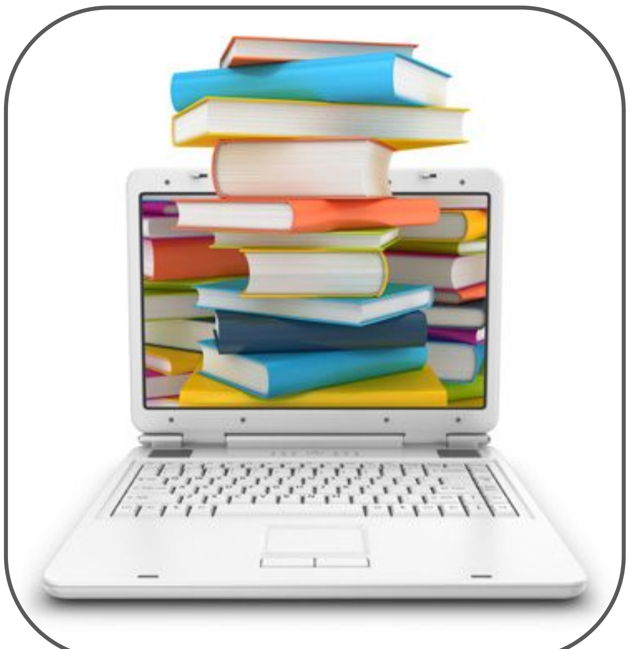
- Traditionally the result of **manual** efforts (e.g., professional lexicographers)
- Useful for **linguistic** and **world knowledge** dissemination (e.g., language learners or text books)
- Regardless of the focus (linguistic vs encyclopedic/world *knowledge*), **valuable in NLP** because they provide *high quality* data to be leveraged in downstream tasks.
- We would like to have reliable means to **create** them anew, but more importantly **extend** and **enrich** existing ones.

Introduction

- " … **renaissance of knowledge-rich approaches in AI and NLP -** namely, approaches that exploit large amounts of machine readable knowledge to perform tasks requiring human intelligence" (Hovy et al., 2013 AI)

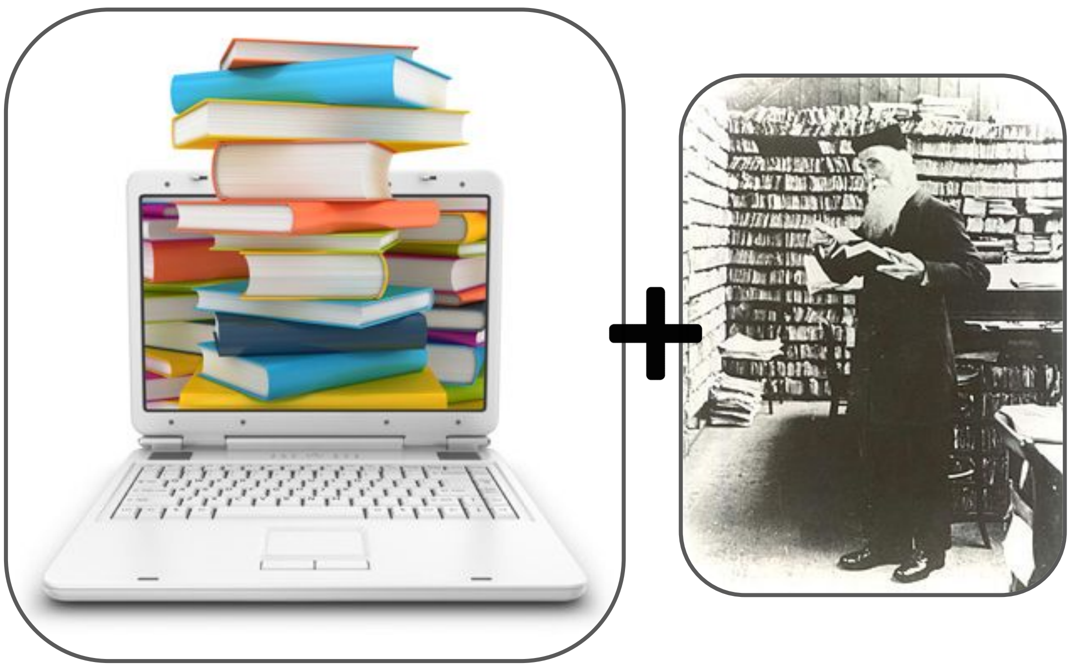
Introduction

- " ... **renaissance of knowledge-rich approaches in AI and NLP -** namely, approaches that exploit large amounts of machine readable knowledge to perform tasks requiring human intelligence" (Hovy et al., 2013 AI)



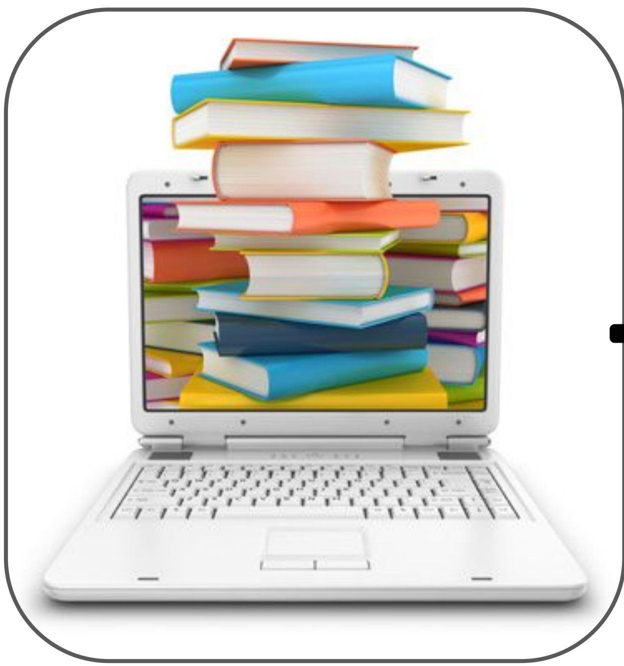
Introduction

- “ … **renaissance of knowledge-rich approaches in AI and NLP -** namely, approaches that exploit large amounts of machine readable knowledge to perform tasks requiring human intelligence” (Hovy et al., 2013 AI)

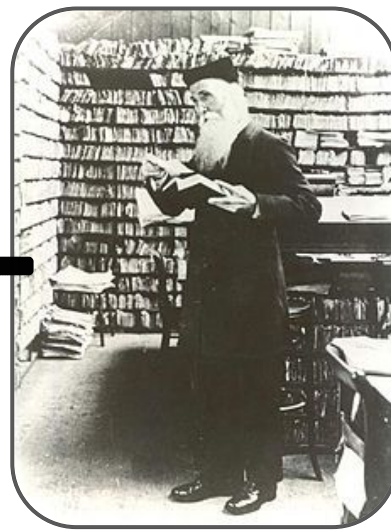


Introduction

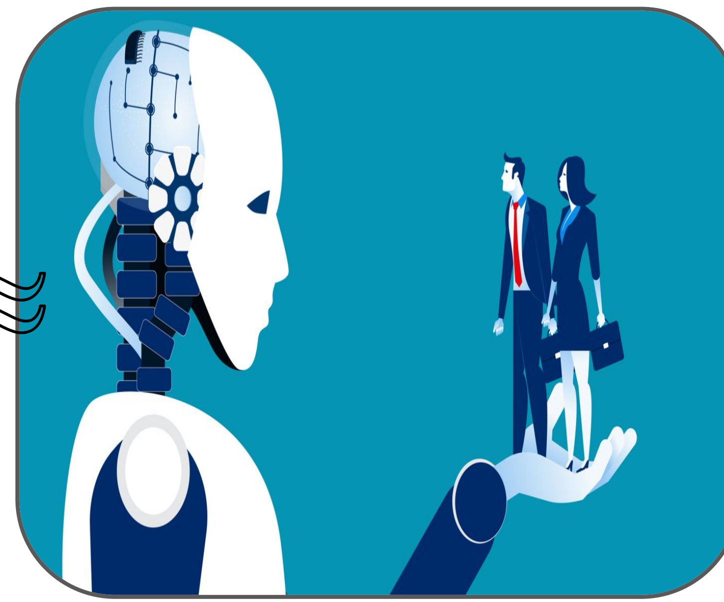
- “ … **renaissance of knowledge-rich approaches in AI and NLP -** namely, approaches that exploit large amounts of machine readable knowledge to perform tasks requiring human intelligence” (Hovy et al., 2013 AI)



+

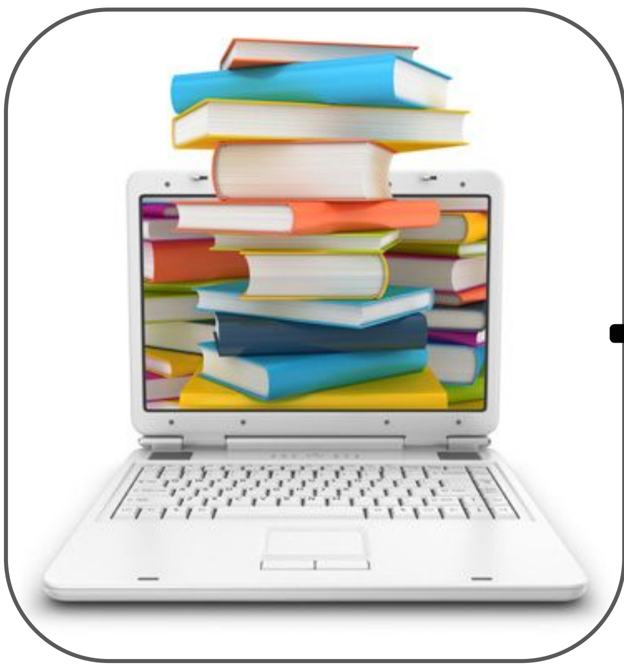


≈

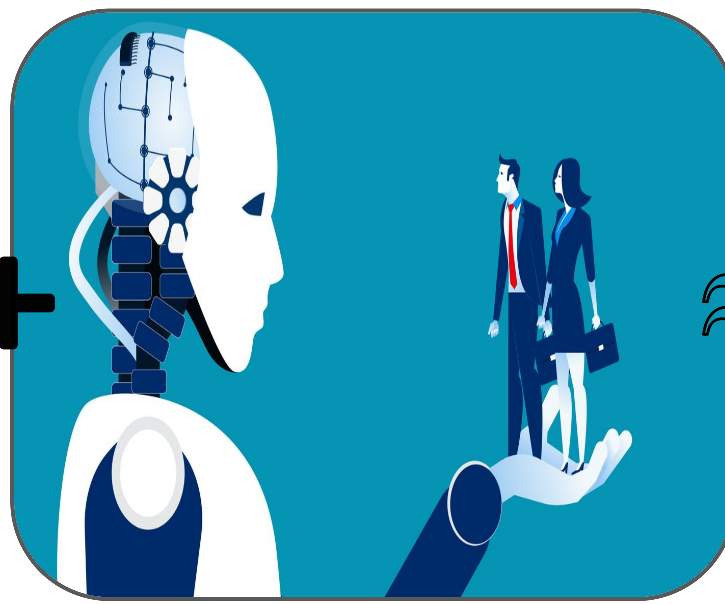


Introduction

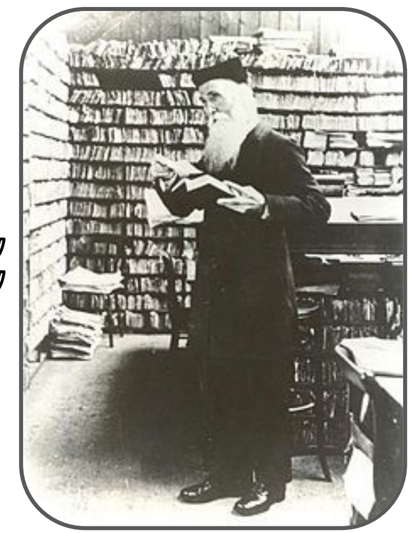
- " … **renaissance of knowledge-rich approaches in AI and NLP -** namely, approaches that exploit large amounts of machine readable knowledge to perform tasks requiring human intelligence" (Hovy et al., 2013 AI)



+

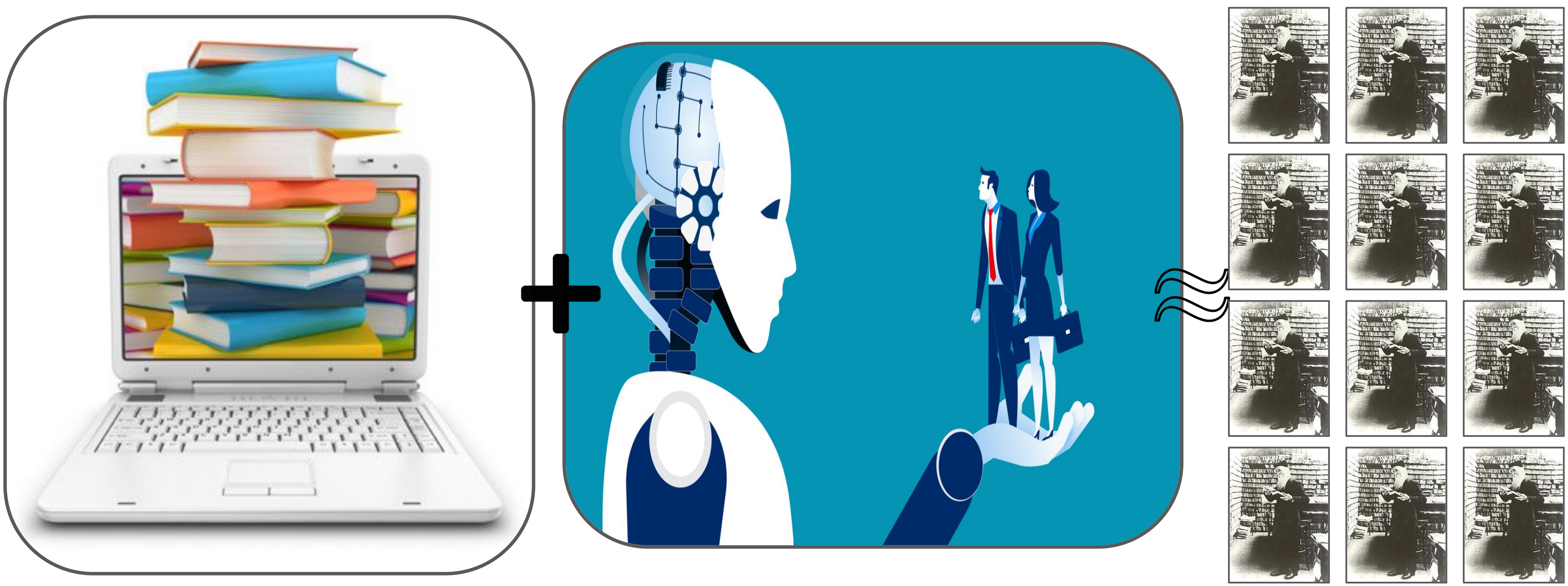


≈



Introduction

- " ... **renaissance of knowledge-rich approaches in AI and NLP** - namely, approaches that exploit large amounts of machine readable knowledge to perform tasks requiring human intelligence" (Hovy et al., 2013 AI)



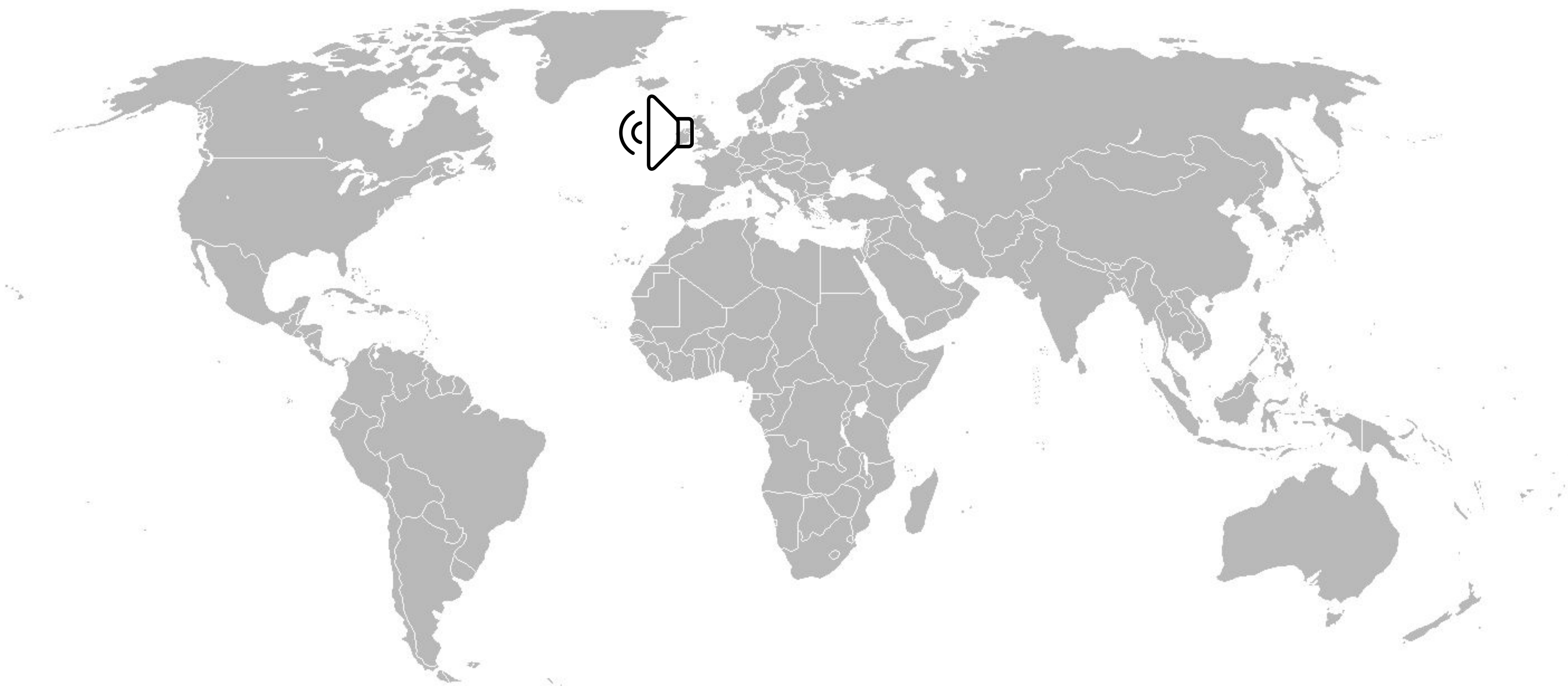

To the
point

Introduction

This tutorial

- Overview of well known lexical resources (prevalence in NLP)
  - Size, features, linguistic/knowledge complexity...
- NLP for lexical resources
  - From raw corpora to the extension (or creation from scratch) of a LR
- Lexical resources for NLP
  - WSD, knowledge-based embeddings and applications in actual NLP problems such as text classification.
- Looking ahead.
  - Future work, current and upcoming challenges (in both areas), new language problems, need for encoding different types of knowledge?

QUESTIONS 2 & 3
PIN: 7024700
www.kahoot.it



LEXICAL RESOURCES

Lexical Resources

A database (or a machine readable dictionary) that provides **structured knowledge** for words, e.g., synonyms of words, semantic and phonological relations between different words.

WordNet

BabelNet

WIKIDATA

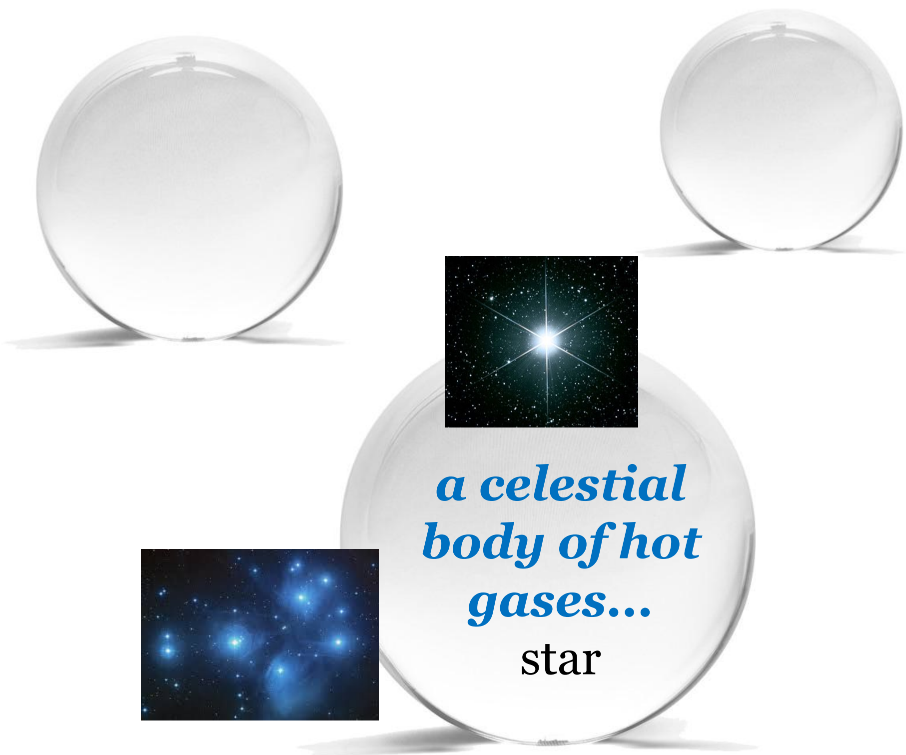
SNOMED CT
The global language of healthcare

Freebase

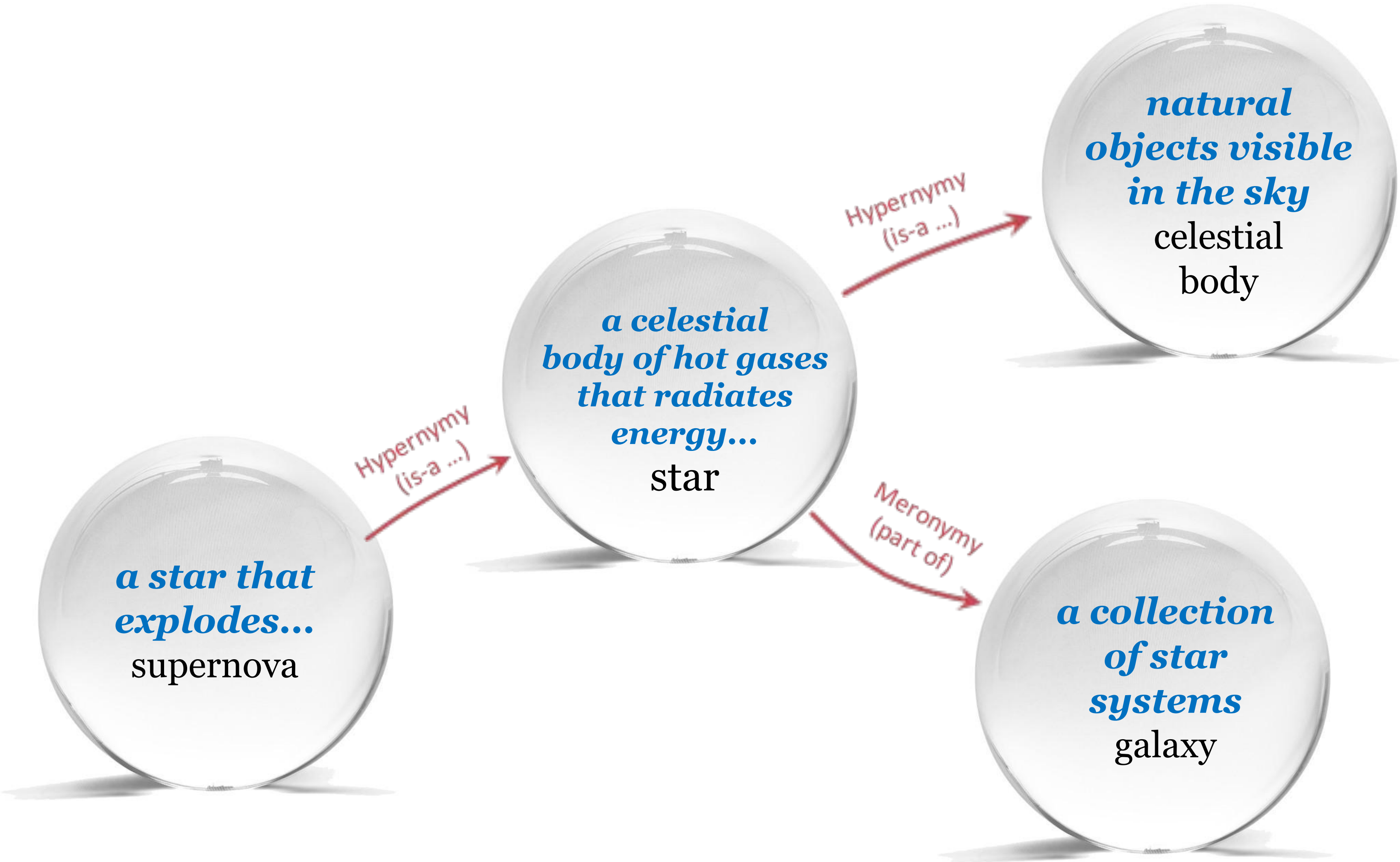
The Paraphrase Database

WordNet: the de facto standard lexical database

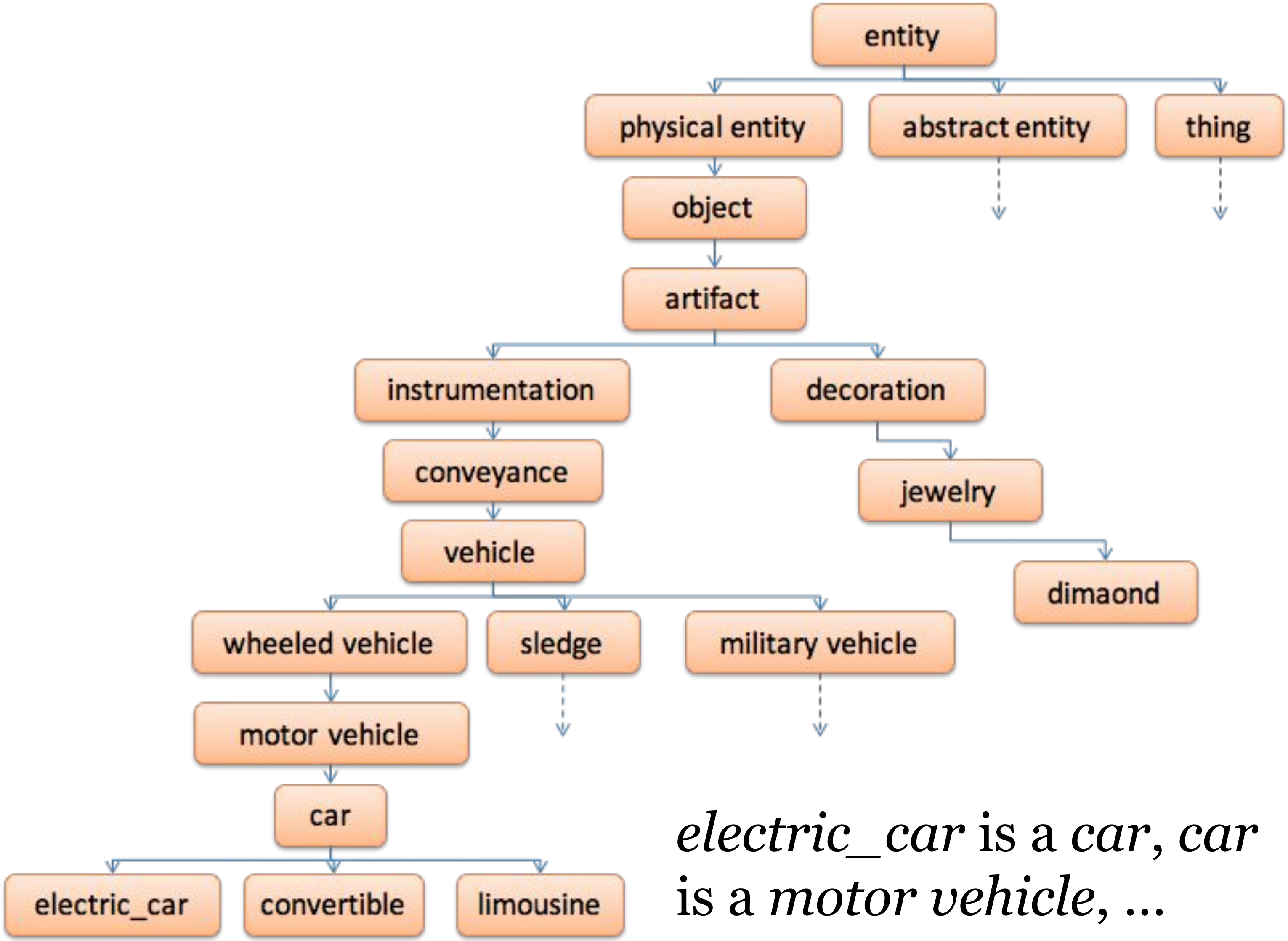
The basic constituents in WordNet are **synsets** (sets of synonymous words that correspond to a unique concept)



WordNet: semantic relations

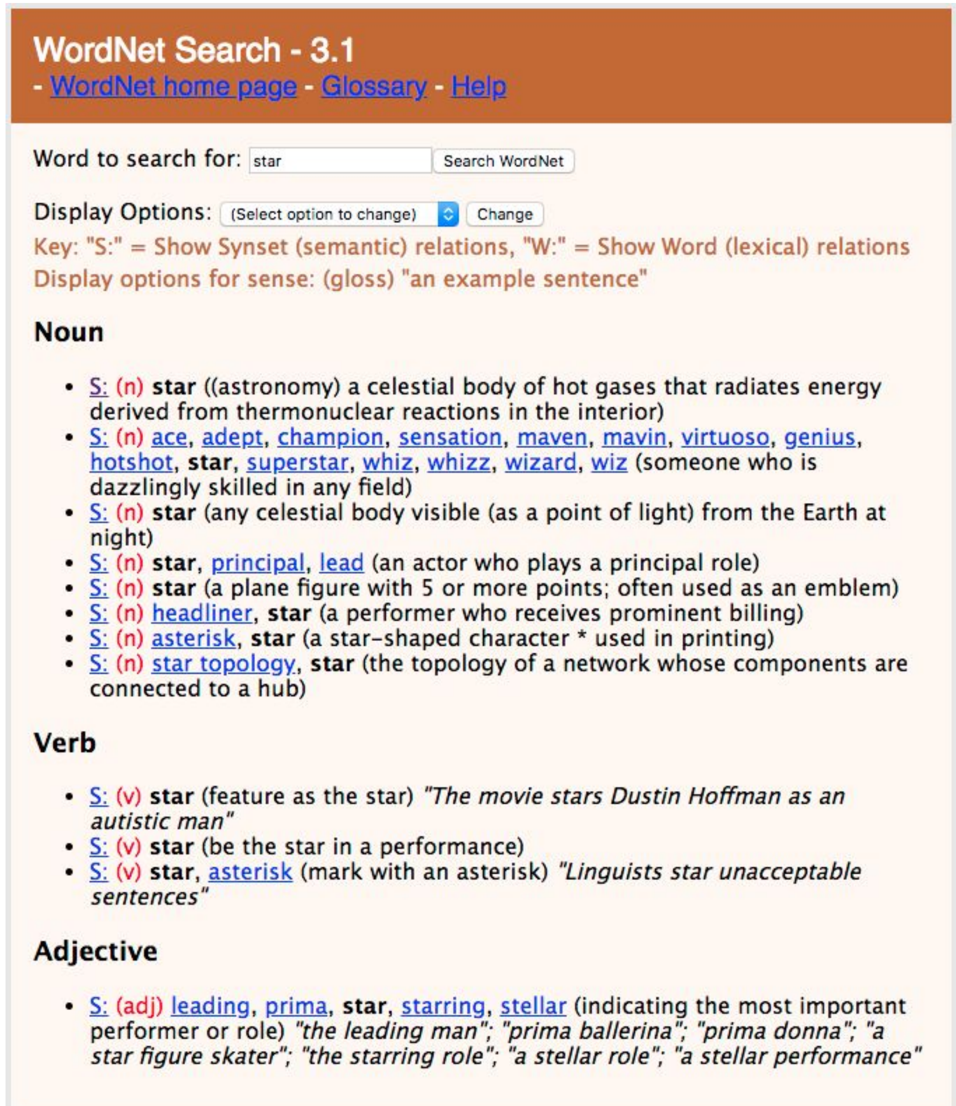


WordNet as a hypernymy hierarchy



electric_car is a *car*, *car* is a *motor vehicle*, ...

WordNet as a sense inventory



[Online browser]

WordNet: Limitations

- **Difficult to update (needs expert curation)**
  - Most recent major update (v3.0) was 10 years ago.

- **Limited vocabulary**
  - Misses many named entities and domain specific terms.

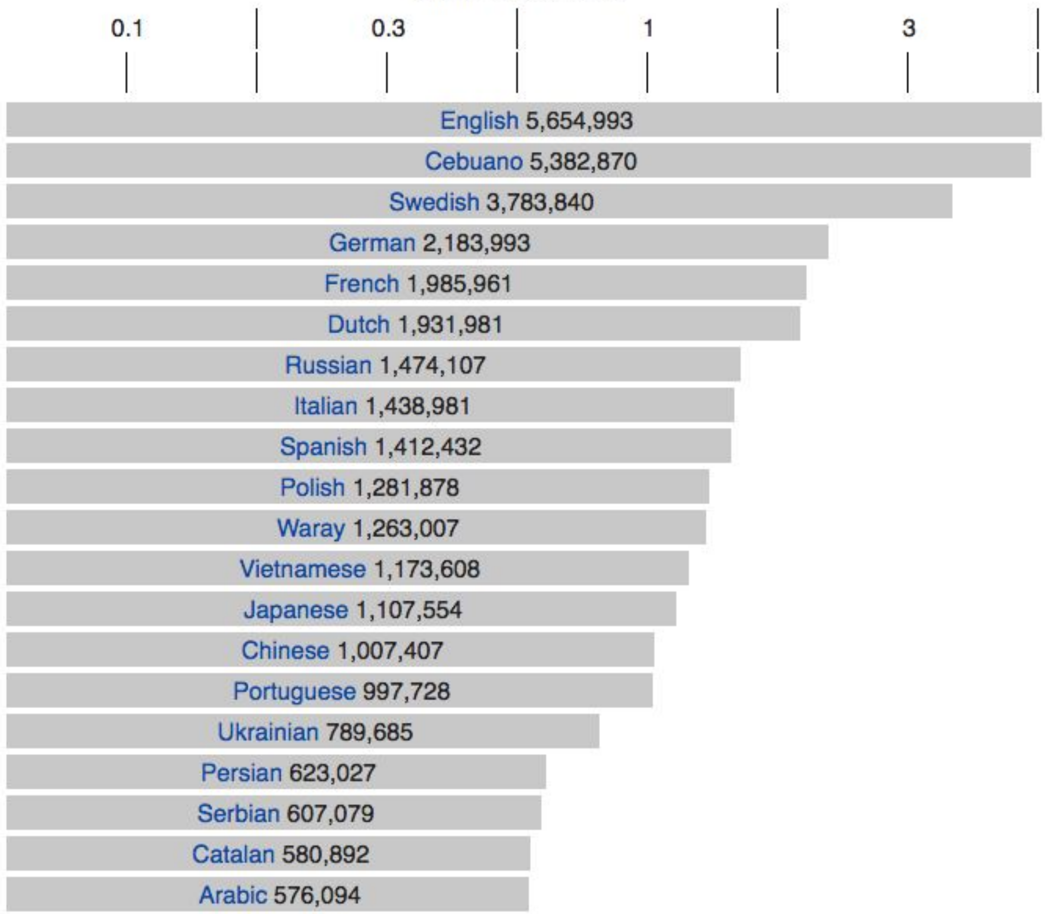
- **Monolingual**

Solution: collaborative resources

Resource diaspora: Wikipedia

Collaborative resources: Wikipedia

Logarithmic graph of the 20 largest language editions of Wikipedia
(as of 24 May 2018)[128]
(millions of articles)



Wikipedia
The Free Encyclopedia

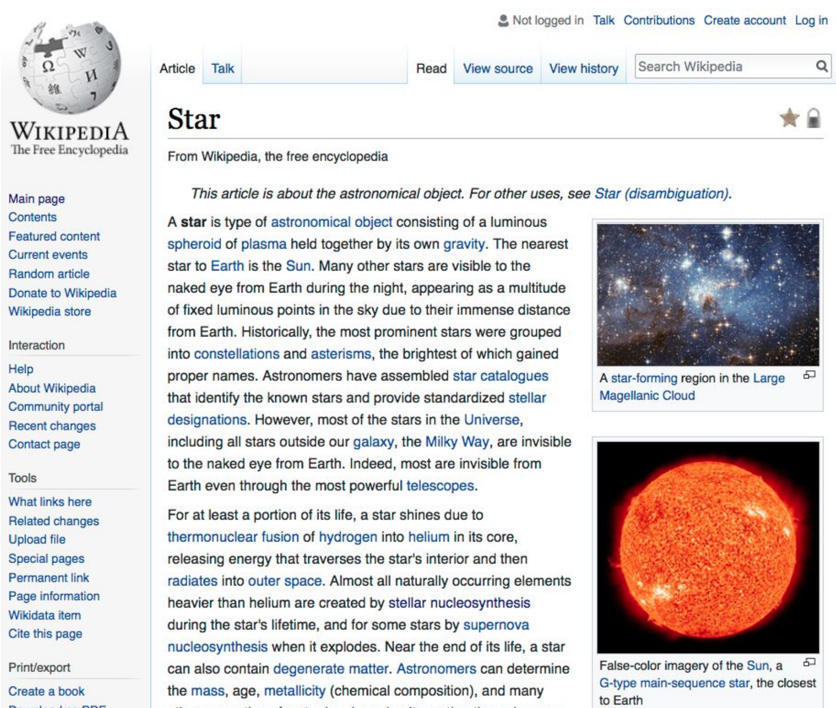
Massively multilingual

Diverse set of information

Constantly update:
Hundreds of new articles every day!

Collaborative resources: Wikipedia

Each Wikipedia article is a concept



Not logged in Talk Contributions Create account Log in

Article Talk Read View source View history Search Wikipedia

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
Download as PDF

Star

From Wikipedia, the free encyclopedia

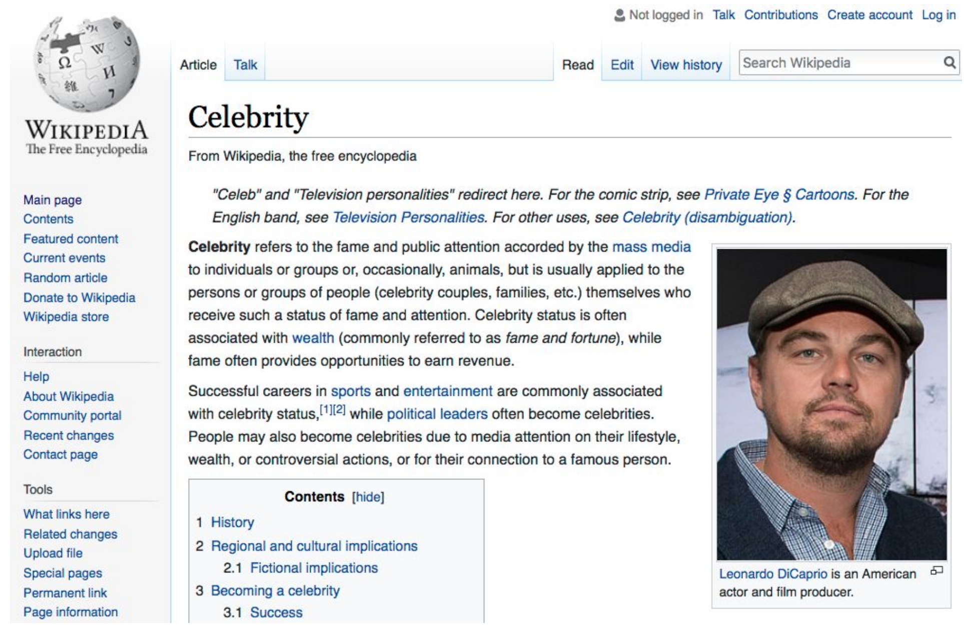
This article is about the astronomical object. For other uses, see Star (disambiguation).

A **star** is type of astronomical object consisting of a luminous spheroid of plasma held together by its own gravity. The nearest star to Earth is the Sun. Many other stars are visible to the naked eye from Earth during the night, appearing as a multitude of fixed luminous points in the sky due to their immense distance from Earth. Historically, the most prominent stars were grouped into constellations and asterisms, the brightest of which gained proper names. Astronomers have assembled star catalogues that identify the known stars and provide standardized stellar designations. However, most of the stars in the Universe, including all stars outside our galaxy, the Milky Way, are invisible to the naked eye from Earth. Indeed, most are invisible from Earth even through the most powerful telescopes.

For at least a portion of its life, a star shines due to thermonuclear fusion of hydrogen into helium in its core, releasing energy that traverses the star's interior and then radiates into outer space. Almost all naturally occurring elements heavier than helium are created by stellar nucleosynthesis during the star's lifetime, and for some stars by supernova nucleosynthesis when it explodes. Near the end of its life, a star can also contain degenerate matter. Astronomers can determine the mass, age, metallicity (chemical composition), and many

A star-forming region in the Large Magellanic Cloud

False-color imagery of the Sun, a G-type main-sequence star, the closest to Earth



Not logged in Talk Contributions Create account Log in

Article Talk Read Edit View history Search Wikipedia

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

Celebrity

From Wikipedia, the free encyclopedia

"Celeb" and "Television personalities" redirect here. For the comic strip, see Private Eye § Cartoons. For the English band, see Television Personalities. For other uses, see Celebrity (disambiguation).

Celebrity refers to the fame and public attention accorded by the mass media to individuals or groups or, occasionally, animals, but is usually applied to the persons or groups of people (celebrity couples, families, etc.) themselves who receive such a status of fame and attention. Celebrity status is often associated with wealth (commonly referred to as *fame and fortune*), while fame often provides opportunities to earn revenue.

Successful careers in sports and entertainment are commonly associated with celebrity status,[1][2] while political leaders often become celebrities. People may also become celebrities due to media attention on their lifestyle, wealth, or controversial actions, or for their connection to a famous person.

Contents [hide]

Leonardo DiCaprio is an American actor and film producer.

Collaborative resources: Wikipedia

Structured knowledge

Alan Turing

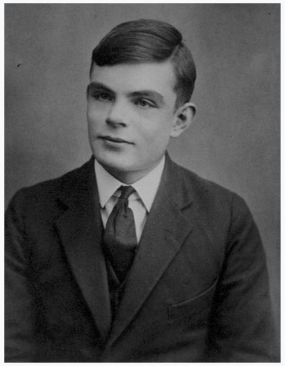
From Wikipedia, the free encyclopedia
(Redirected from Alan turing)

"Turing" redirects here. For other uses, see Turing (disambiguation).

Alan Mathison Turing OBE FRS (/ˈtjʊərɪŋ/; 23 June 1912 – 7 June 1954) was an English computer scientist, mathematician, logician, cryptanalyst and theoretical biologist. He was highly influential in the development of theoretical computer science, providing a formalisation of the concepts of algorithm and computation with the Turing machine, which can be considered a model of a general purpose computer.[2][3][4] Turing is widely considered to be the father of theoretical computer science and artificial intelligence.[5]

During the Second World War, Turing worked for the Government Code and Cypher School (GC&CS) at Bletchley Park, Britain's codebreaking centre that produced Ultra intelligence. For a time he led Hut 8, the section responsible for German naval cryptanalysis. He devised a number of techniques for speeding the breaking of German ciphers, including improvements to the pre-war Polish bombe method, an electromechanical machine that could find settings for the Enigma machine. Turing played a pivotal role in cracking intercepted coded messages that enabled the Allies to defeat the Nazis in many crucial engagements, including the Battle of the Atlantic, and in so doing helped win the war.[6][7] Counterfactual history is difficult with respect to the effect Ultra intelligence had on the length of the war,[8] but at the upper end it has been estimated that this work shortened the war in Europe by more than two years and saved over fourteen million lives.[6]

Alan Turing
OBE FRS



Turing in 1927

Born	23 June 1912 Maida Vale, London, England, United Kingdom
Died	7 June 1954 (aged 41) Wilmslow, Cheshire, England, United Kingdom Cyanide poisoning
Residence	Wilmslow, Cheshire, England
Citizenship	British
Fields	Mathematics, cryptanalysis, logic, computer science, mathematical and theoretical biology
Institutions	Victoria University of Manchester Government Code and Cypher School National Physical Laboratory
Education	Sherborne School
Alma mater	King's College, Cambridge (BA) Princeton University (PhD)
Thesis	*Systems of Logic Based on Ordinals* (1938)
Doctoral advisor	Alonzo Church[1]
Doctoral students	Robin Gandy[1]

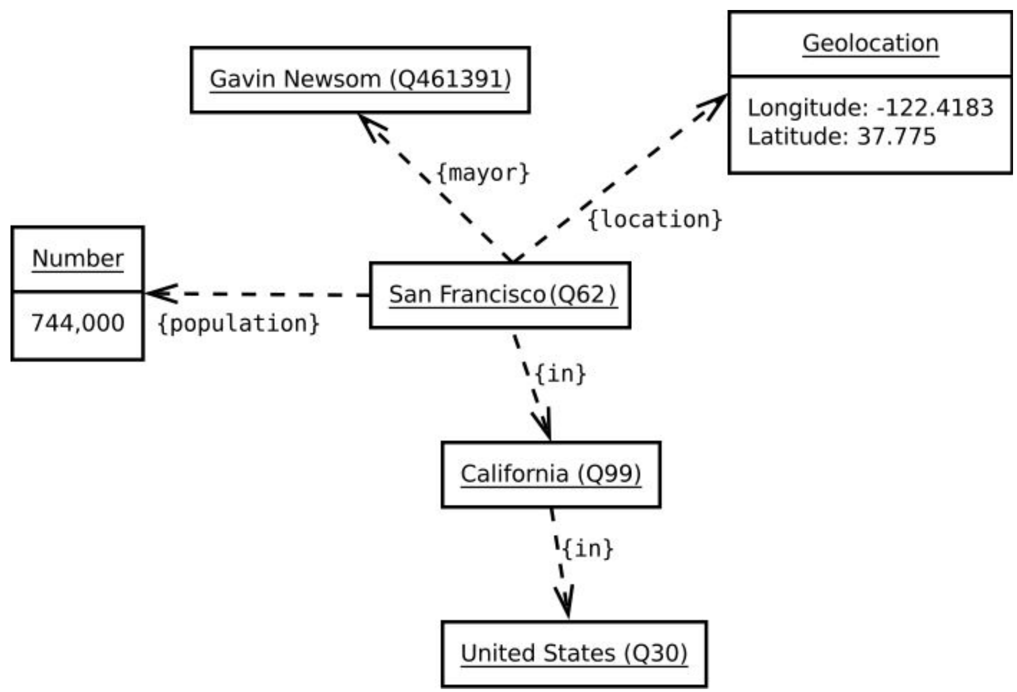
Freebase

Was a large collaborative knowledge base

Shut down and move to Wikidata (from 2015)

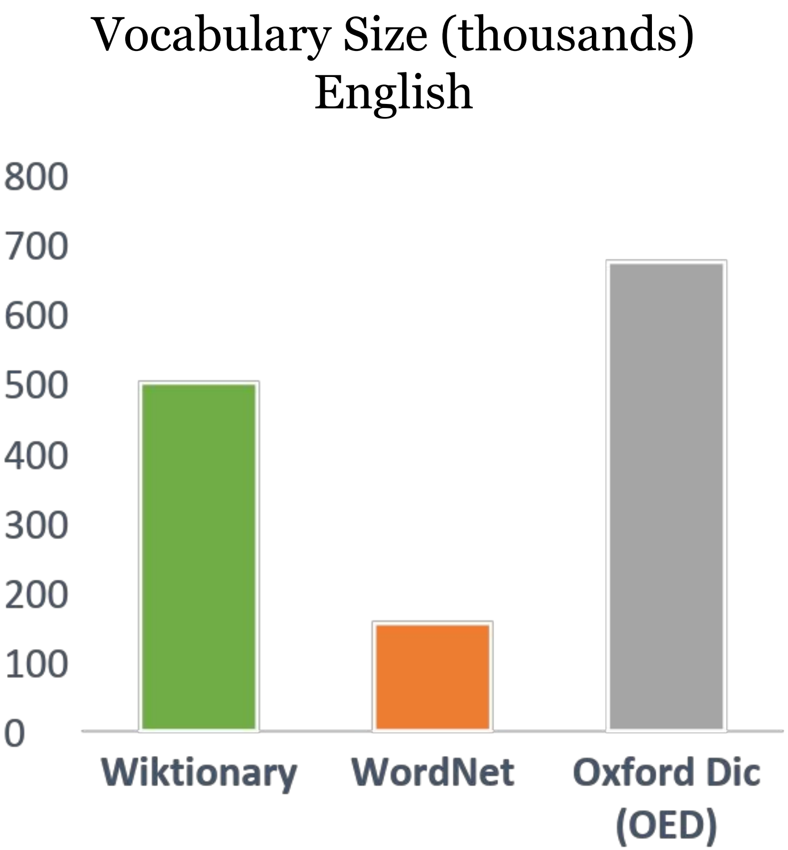
Wikidata

48,323,790 data items



Collaborative resources: Wiktionary

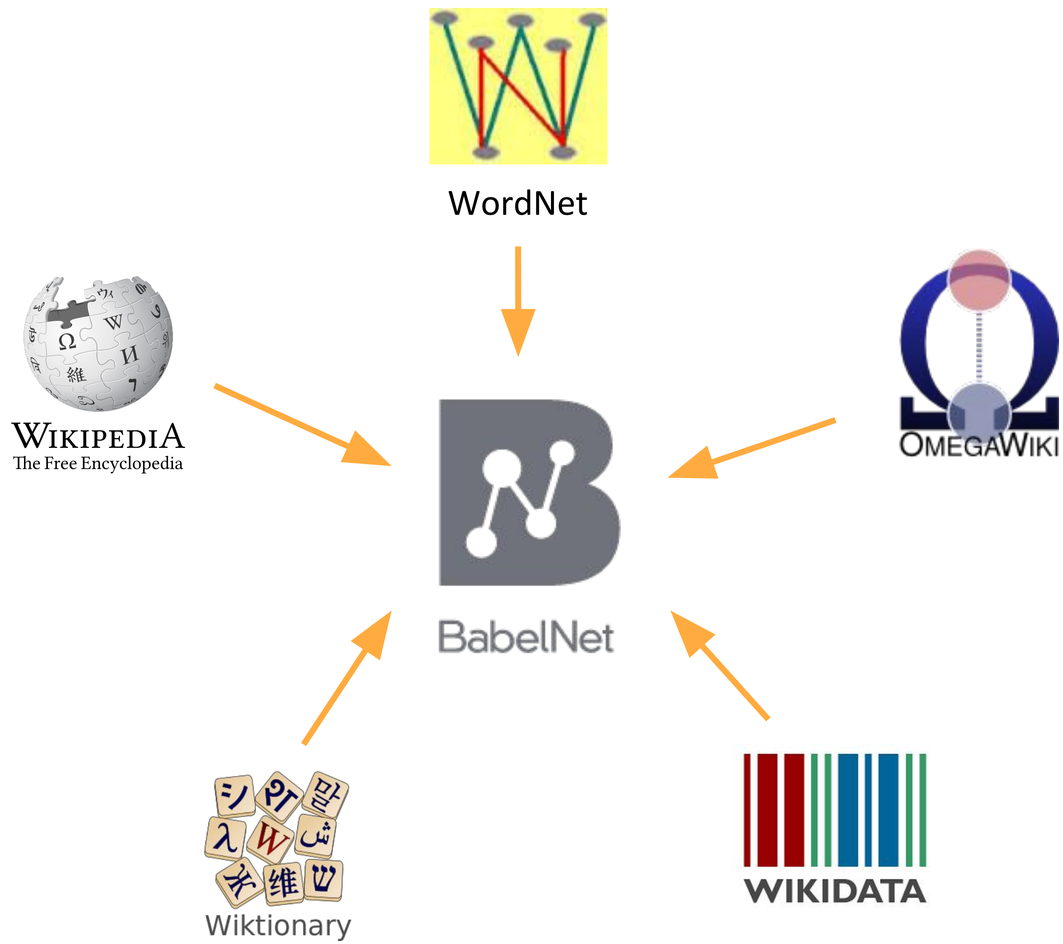
Wiktionary



Wiktionary is available in 172 languages!

Language name	Number of entries
Latin	628175
English	501171
Italian	491347
French	279926
Spanish	252154
German	111264
Finnish	110834
Esperanto	104972
Portuguese	98816
Swedish	90480
Latvian	67924

BabelNet: Multilingual encyclopedic dictionary



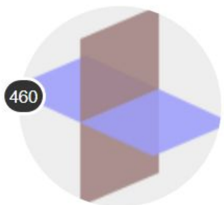
BabelNet: Multilingual sense inventory

airplane, plane, aeroplane

An aircraft that has a fixed wing and is powered by propellers or jets

ID: 00001697n | Concept

- 固定翼飛機, 飛行機, 飞龙机
- avion, aéroplane
- Flugzeug
- aereo, aeroplano, apparecchio



plane, sheet

(mathematics) an unbounded two-dimensional shape

ID: 00062766n | Concept

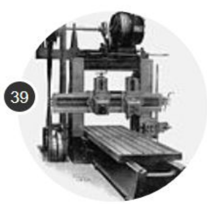
- 平面, 面
- plan
- Ebene (Mathematik)
- piano, piano geometrico

plane

A level of existence or development

ID: 00062767n | Concept

- 平 面 的 存 在
- plan
- Ebene
- piano, Spostamento della realtà, livello



planer, plane, planing machine

A power tool for smoothing or shaping wood

ID: 00062768n | Concept

- 刨床
- raboteuse, rabot
- Hobelmaschine
- piallatrice

plane, woodworking plane, carpenter's plane

A carpenter's hand tool with an adjustable blade for smoothing or shaping wood

ID: 00016196n | Concept

- 刨
- rabot, avion, appareil
- Hobel
- pialla, piana, pialletto

BabelNet: Rich and diverse knowledge

bn:00001697n • NOUN • Concept • • Categories: Aircraft configurations, American inventions, Vehicles introduced in 1903

EN **airplane** • **plane** • **aeroplane** • **fixed-wing aircraft** • **Fixed-wing aeroplane**

An aircraft that has a fixed wing and is powered by propellers or jets *WordNet* More definitions

The flight was delayed due to trouble with the airplane *WordNet*

IS A	heavier-than-air craft • Powered aircraft • fixed-wing aircraft
HAS PART	stick shaker • planform • DE Tragwerk
HAS KIND	hydroplane • Ader Avion III • Bücker Bü 134
HAS INSTANCE	tail rotor • fighter aircraft • jet
DESCRIBED BY SOURCE	Armenian Soviet Encyclopedia

EXPLORE NETWORK

BabelNet

Translations

AR طائرة, طَائِرَة, طائِرة, طَيَّارَة, مرْكبة جوِّيّة, الطائرات, الطائرة, Aircraft, Airplane, الطائرات الثابتة الجناحين, الطيارة, طائرات, طائره, طيارة, قاذفة

ZH 飞机, 固定翼飛機, 飛機, 飛行機, 飛龍機, 飞行机, 飞龙机, *Aircraft, Fixed-wing aircraft, Flugan, ✈, 固定机翼飞机, 固定翼, 固定翼机, 固定翼機, 固定翼航空器, 固定翼飞机, 定翼機, 定翼飛機*, 飞 机

EN airplane, plane, aeroplane, fixed-wing aircraft, Fixed-wing aeroplane, powered fixed-wing aircraft, *Aero-plane, Aero-planes, Aero planes, Aeroplanes, Aerplane, Air-plane, Air-planes, Air plane, Air planes, Airoplane, Airplanes, Æroplane, ✈*, □ □ , □

FR avion, aéroplane, *Avions, Fonctionnement d'un avion*, appareil

DE Flugzeug, Flieger, *3-Achs-gesteuert, 3-Achs-Steuerung, Aëroplan, Flight Control System, Flugsteuerung, Flugzeugsteuerung, Flugzeugzelle, Flächenflugzeug, Gemischtbauweise, Starrflügel, Starrflügelflugzeug, Steuerfläche, Verkehrflugzeug, ✈*, flugzeuge, flugzeugen

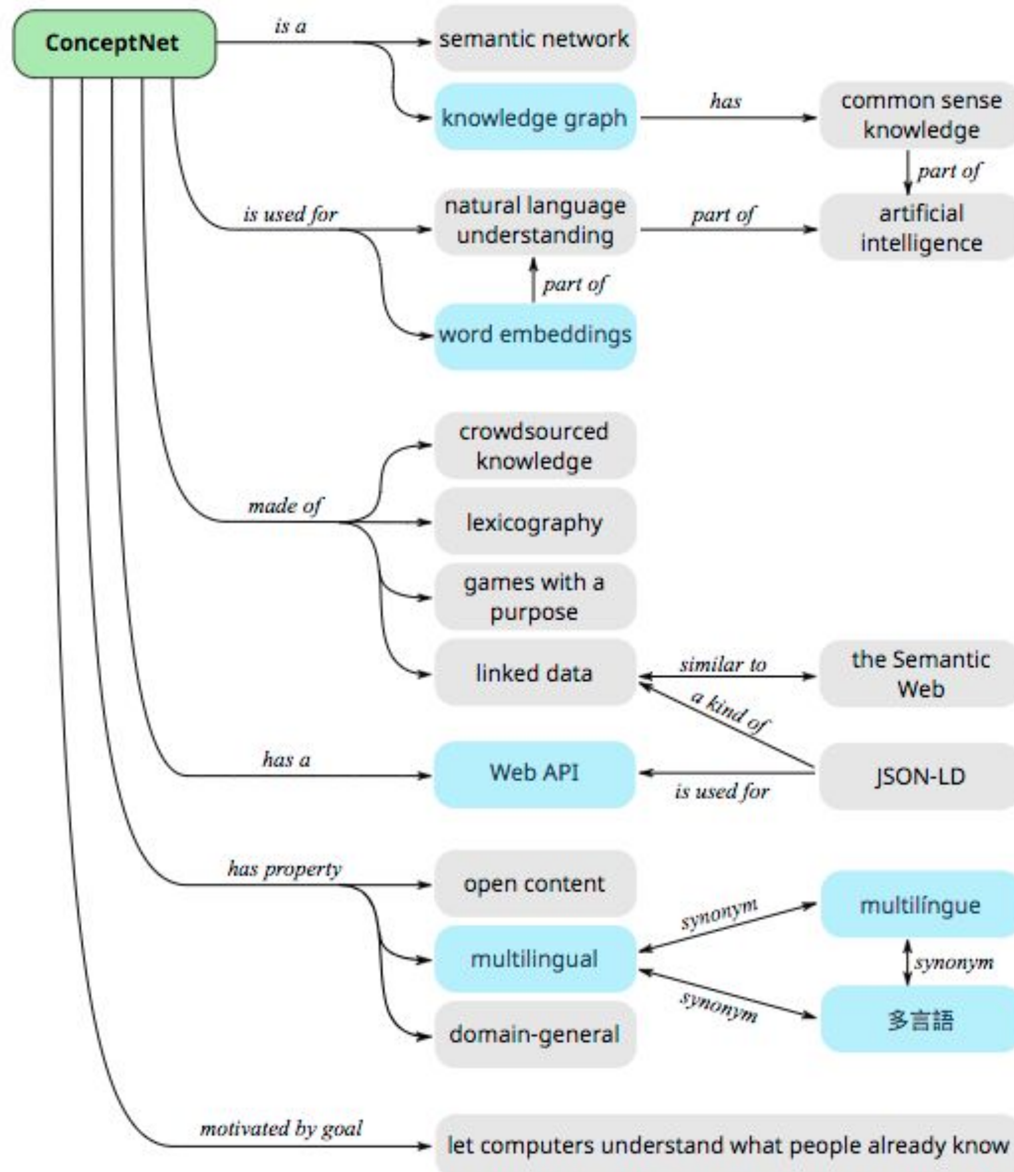
ConceptNet

A freely-available semantic network

Started as a crowdsourcing project, collecting facts from people.

But, now includes, among others:

Wiktionary DBpedia WordNet



PPDB: The Paraphrase Database

An automatically extracted database containing **millions of paraphrases** in 16 different languages.

thrown into jail ~ imprisoned

Extracted from bilingual parallel corpora through **bilingual pivoting** (Bannard and Callison-Burch, 2005)

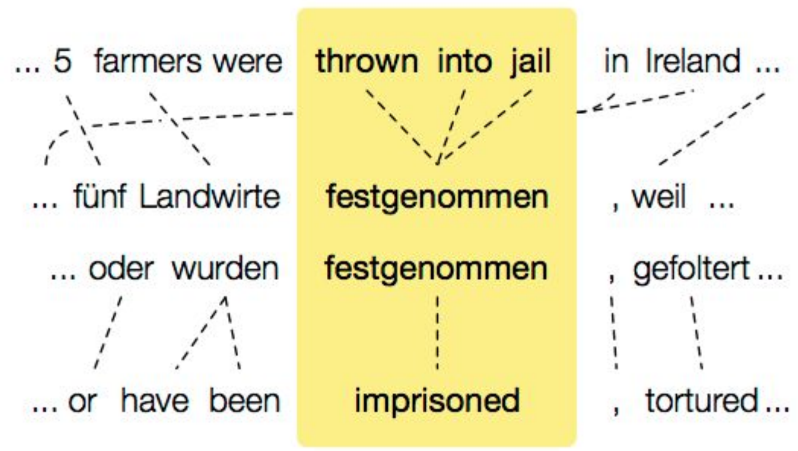


Illustration from Ganitkevitch et al (2013)

PPDB: The Paraphrase Database

Three types of paraphrases:

- Lexical - single word to single word
  - reinforced ||| strengthened
  - bibliography ||| references
- Phrasal - multiword to single/multiword
  - power plants ||| power stations
  - free trade area ||| free trade zone
- Syntactic - paraphrase rules containing non-terminal symbols
  - DT characteristic of NP ||| DT feature of NP

PPDB: The Paraphrase Database

Examples from @ppdb

Time alterations

09:00 | 9 a.m.

09:00 | 9 hours

09:00 | nine hours

09:00 | nine o 'clock

Abbreviations

sme | small and medium enterprises

unicef | united nations children 's fund

roi | return on investment

Verb particles

speed up | accelerate

blow up | explode

throw up | puke

set up | establish

speed up | expedite give

up | abandon

Comparatives

safer | more secure

denser | more dense

wetter | more humid

fairer | more just

Multilingual PPDB

Extended to 23 different languages

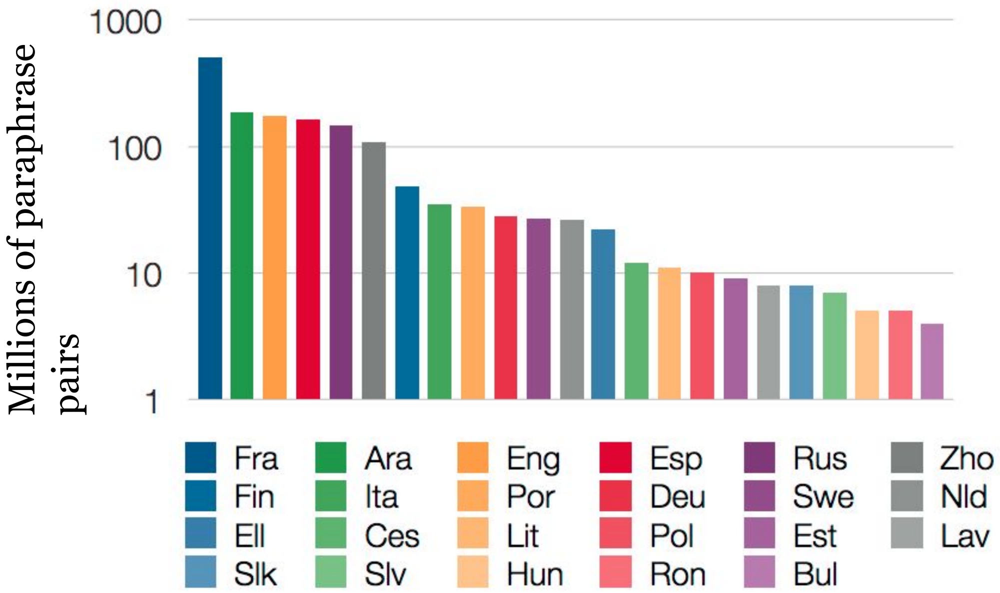


Figure from Ganitkevitch and Chris Callison-Burch (2014)

SNOMED: clinical health terminology

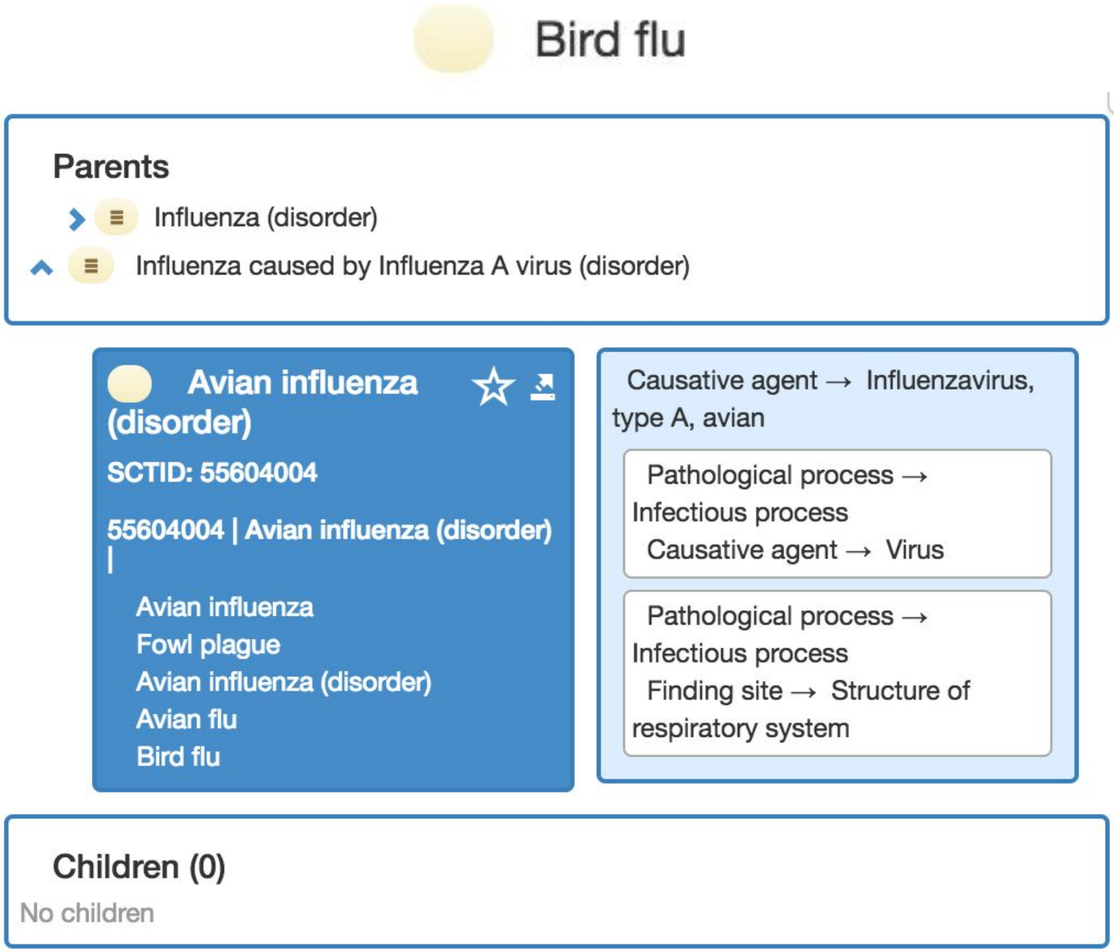
Goal: the development of a global language for health

Includes 311,000 concepts

Clinical findings, Causes of disease, Procedures, Anatomy, Observations, Products




SNOMED: concepts and descriptions



QUESTION 4

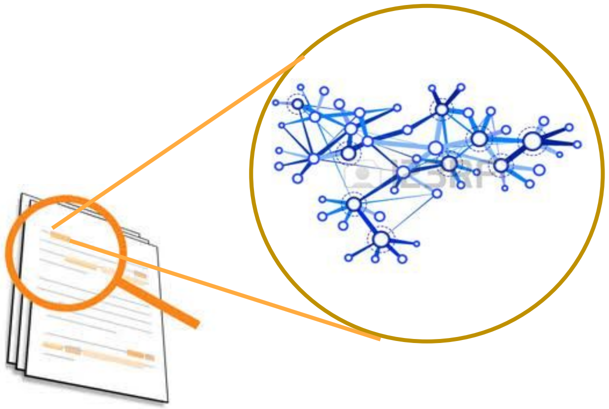
PIN: 7024700

www.kahoot.it

NLP FOR LEXICAL RESOURCES

NLP for Lexical Resources

1. Intro
2. Terminology Extraction
3. Definition Modeling
4. Dictionary Examples
5. Hypernymy and Taxonomies
6. Topical/Thematic Clustering



Introduction

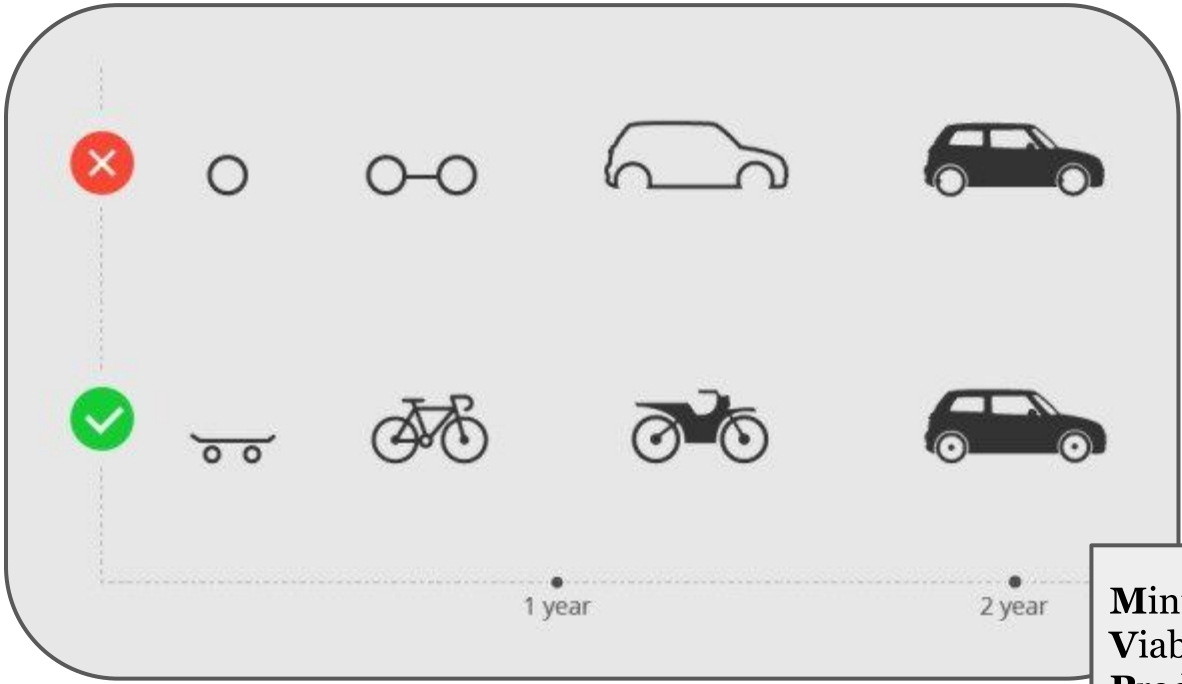
In this segment we explore existing NLP systems which model specific linguistic/lexicographic phenomena from a resource prism.

However, in addition to the inherent difficulty of NLP, there is the **MVP** issue…

Introduction

In this segment we explore existing NLP systems which model specific linguistic/lexicographic phenomena from a resource prism.

However, in addition to the inherent difficulty of NLP, there is the **MVP** issue...



Minimum
Viable
Product

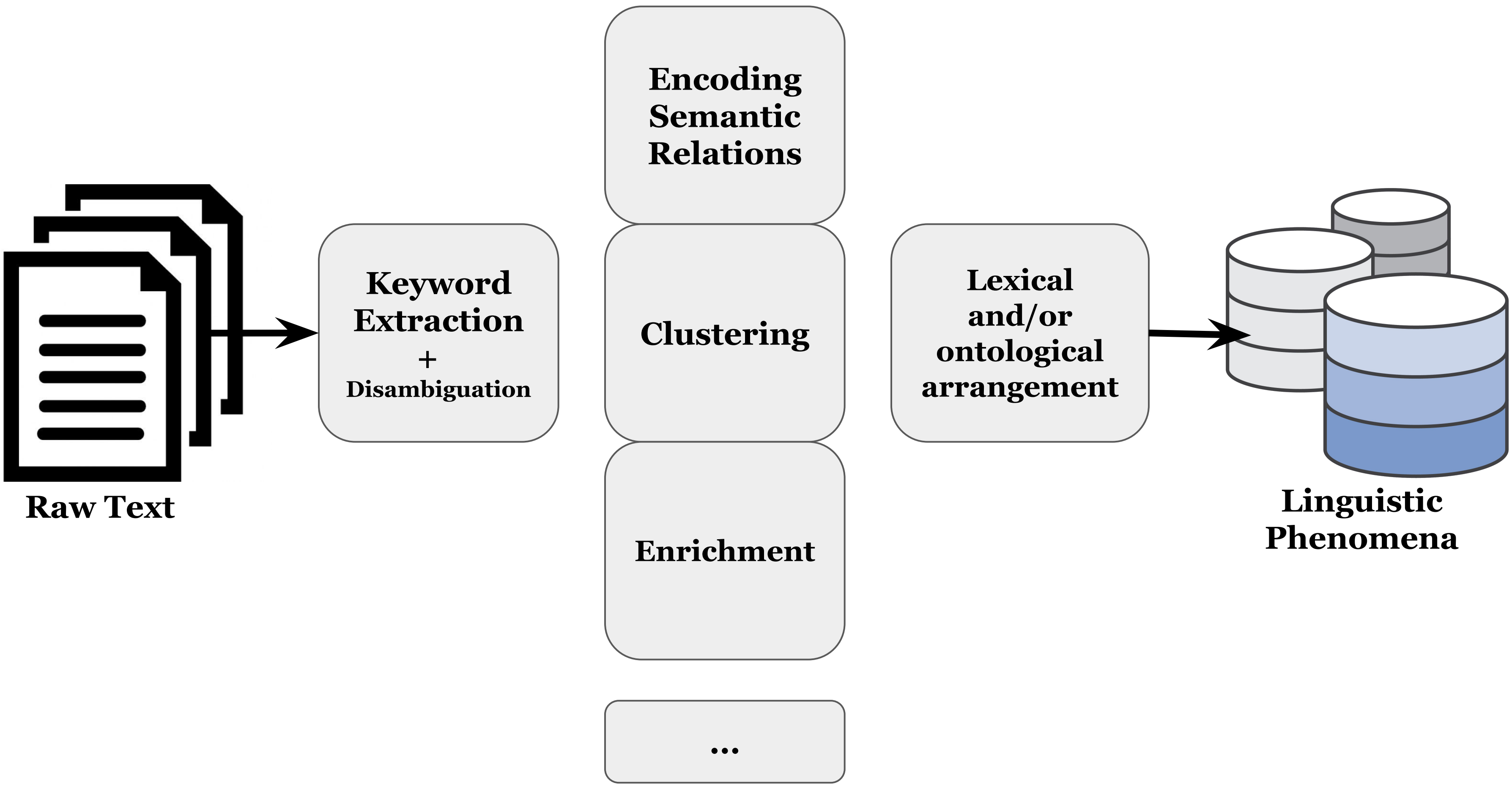
Introduction

In this segment we explore existing NLP systems which model specific linguistic/lexicographic phenomena from a resource prism.

However, in addition to the inherent difficulty of NLP, there is the **MVP** issue...

- Are there **robust** enough NLP systems to reliably transform *raw* textual data into a suitable representation for a lexical resource?
- Lexical resources are about **quality**, can we come close with automatic systems?
- Aren't we better off simply with corpus-derived **statistical models**?

Introduction



Terminology Extraction

Automatic acquisition of domain terminologies from corpora emerges as a natural *zero step* in any attempt towards enrichment of lexical resources.

- Frequency and *tf-idf*

- Lexical specificity

- Termhood measures

Frequency and *tf*idf*

- Give me the most frequent words in the whole corpus.
- Give me the most frequent words of each document, or per section, or per position, or per font formatting.
- Give me the most important words according to their relative weight in each document of the corpus.

Terminology Extraction

- Factor both raw frequency and inverse document frequency

$$\mathrm{tfidf}(t,d,D) = \mathrm{tf}(t,d) \times \mathrm{idf}(t,D)$$

- tf may be logarithmically scaled: tf(t,d) = 1 + log f(t,d)
  - ... or normalized to avoid rewarding long documents

$$\mathrm{tf}(t,d) = \frac{\mathrm{f}(t,d)}{\max\{\mathrm{f}(w,d) : w \in d\}}$$

- idf tells us how common is a term in a document collection

$$\mathrm{idf}(t,D) = \log \frac{|D|}{|\{d \in D : t \in d\}|}$$

Terminology Extraction

Lexical Specificity (Lafon, 1980)

- 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 conventional *tf-idf* (Camacho-Collados et al., AIJ 2016)

- Given a corpus of size T and a subcorpus of size F, for each word w:

$$spec(T, t, F, f) = -\log_{10} P(X \geq f)$$

$$P(X \geq f) = \sum_{i=f}^{F} P(X = i)$$

- F frequency of w in corpus
- f frequency of w in subcorpus
- P follows the hypergeometric distribution

Terminology Extraction

Termhood measures (Frantzi et al., 2000; Bonin et al., 2010)

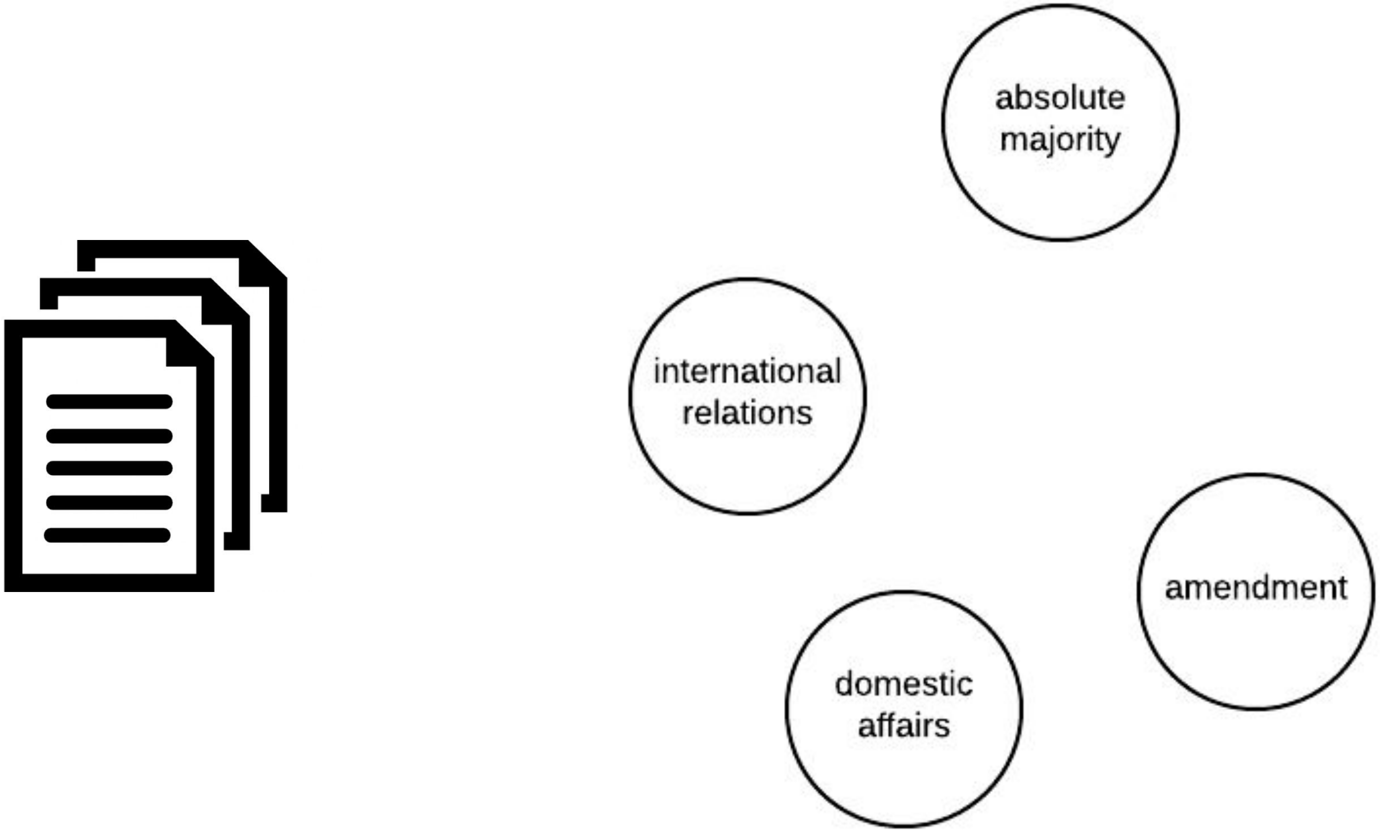
- C-Value, given a multiword term, and given one or more nested terms, define *termhood* of each individual candidate.
  - ***soft contact lens*** > {***contact lens, soft contact***} (Frantzi et al., 2000)

$$C\text{-}value(a) = \begin{cases} \log_2 |a| \cdot f(a) & a \textit{ is not nested,} \\ \log_2 |a|\left(f(a) - \frac{1}{P(T_a)} \sum_{b \epsilon T_a} f(b)\right) & \textit{otherwise} \end{cases} \quad (3)$$

where a is the candidate string, $f(.)$ is its frequency of occurrence in the corpus, T_a is the set of extracted candidate terms that contain a, $P(T_a)$ is the number of these candidate terms.

- Increasingly sophisticated variants: NValue, NCValue, NTValue, etc.

Terminology Extraction



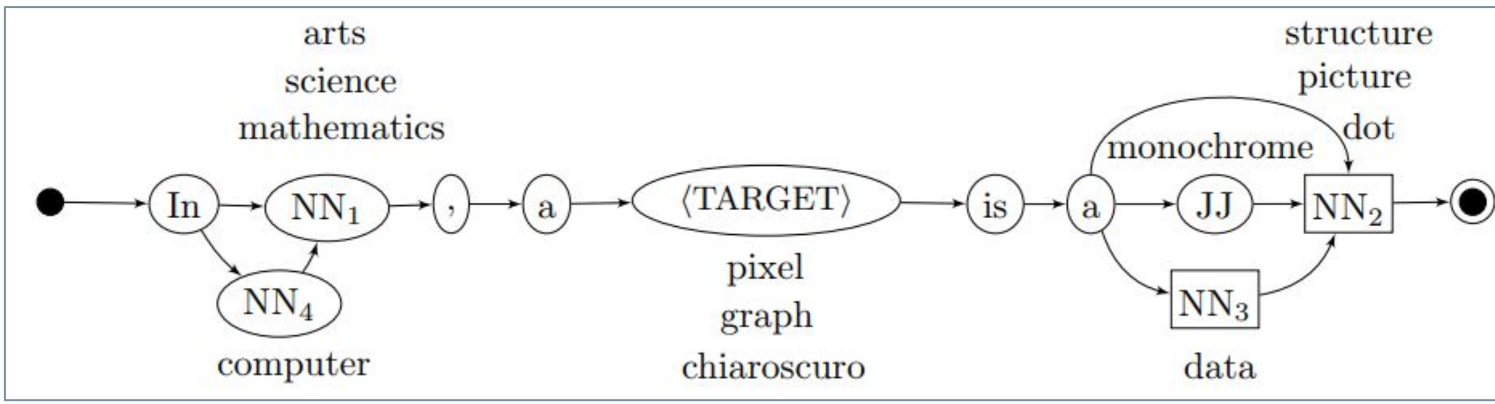
Definition Modeling

Definitions are important for seeking the **meaning** of a word (Navigli and Velardi, ACL 2010) - Also language learning, WSD and modeling OOV words.

- **Definition Extraction**
- **Definition Generation**
- **Lexical Access**

Definition Extraction

- Navigli and Velardi (ACL 2010) - Word Class Lattices

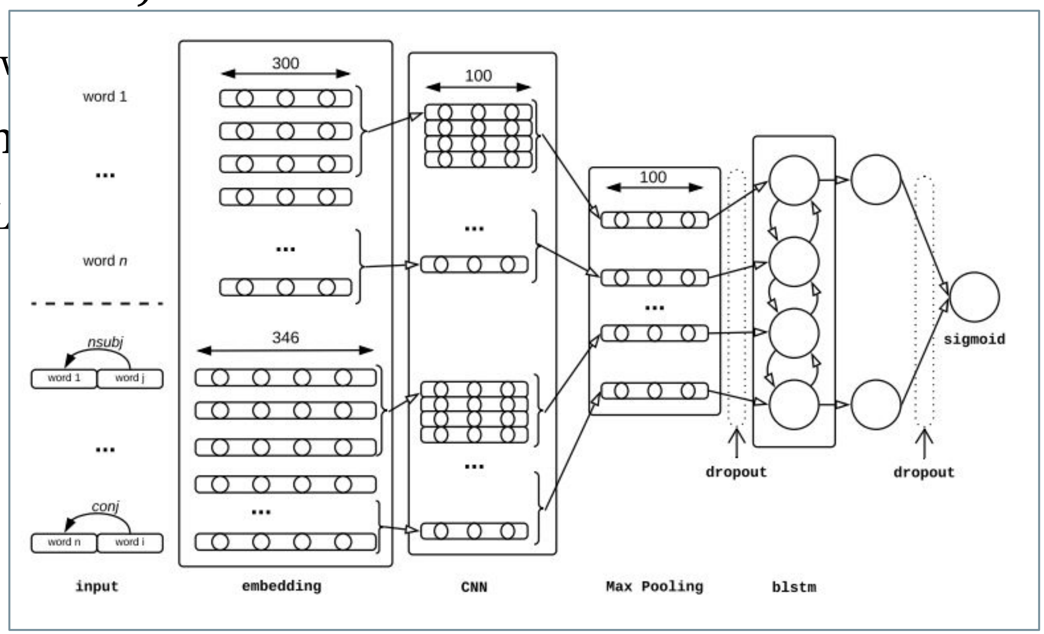


Definition Extraction

- Boella and DiCaro (ACL 2013)
  - Dependency relations and SVM classifier.
- Jin et al. (EMNLP 2013)
  - CRF system + lexical, terminological and structural features.
- Li et al. (CCL 2016)
  - Frequent words + POS of infrequent words into LSTMs.
- Espinosa-Anke and Schockaert (poster Sunday)
  - CNN + BLSTM + Syntactic dependencies

Definition Extraction

- Boella and DiCaro (ACL 2013)
  - Dependency relations and SVM classifier.
- Jin et al. (EMNLP 2013)
  - CRF system + lexical, terminological and structural features.
- Li et al. (CCL 2016)
  - Frequent
- Espinosa-An
  - CNN + BL



Definition Modeling

- Reverse Dictionary / Lexical Access (Hill et al., TACL 2016)

Concept lookup: given a definition, find the corresponding word

Tip-of-the-tongue problem (Zock and Bilac, 2004)

Useful for writers or translators, when they are are unsure how to express an idea they want to convey or cannot recall the word in time

Example: https://www.onelook.com

Takes WordNet as lexical resource; an LSTM network that encodes the definition to the corresponding word embedding

Multiple experiments: (1) recall seen definitions, (2) generalise to unseen definitions from the same resource, and (3) generalize to unseen out-of-domain definitions

Definition Generation

- Definition Generation: Learning to Define Word Embeddings in Natural Language (Noraset et al., AAAI 2017)

Model	creek	feminine	mathematical
Random Emb	to make a loud noise	to make a mess of	of or pertaining to the middle
NE	any of numerous bright translucent organic pigments	a gender that refers chiefly but not exclusively to males or to objects classified as male	of or pertaining to algebra
Seed	a small stream of water	of or pertaining to the fox	of or pertaining to the science of algebra
S+I	a small stream of water	of or pertaining to the human body	of or relating to or based in a system
S+H	a stream of water	of or relating to or characteristic of the nature of the body	of or relating to or characteristic of the science
S+G	a narrow stream of water	having the nature of a woman	of or pertaining to the science
S+G+CH	a narrow stream of water	having the qualities of a woman	of or relating to the science of mathematics
S+G+CH+HE	a narrow stream of water	having the character of a woman	of or pertaining to the science of mathematics

- https://github.com/websail-nu/torch-defseq

Definition Generation

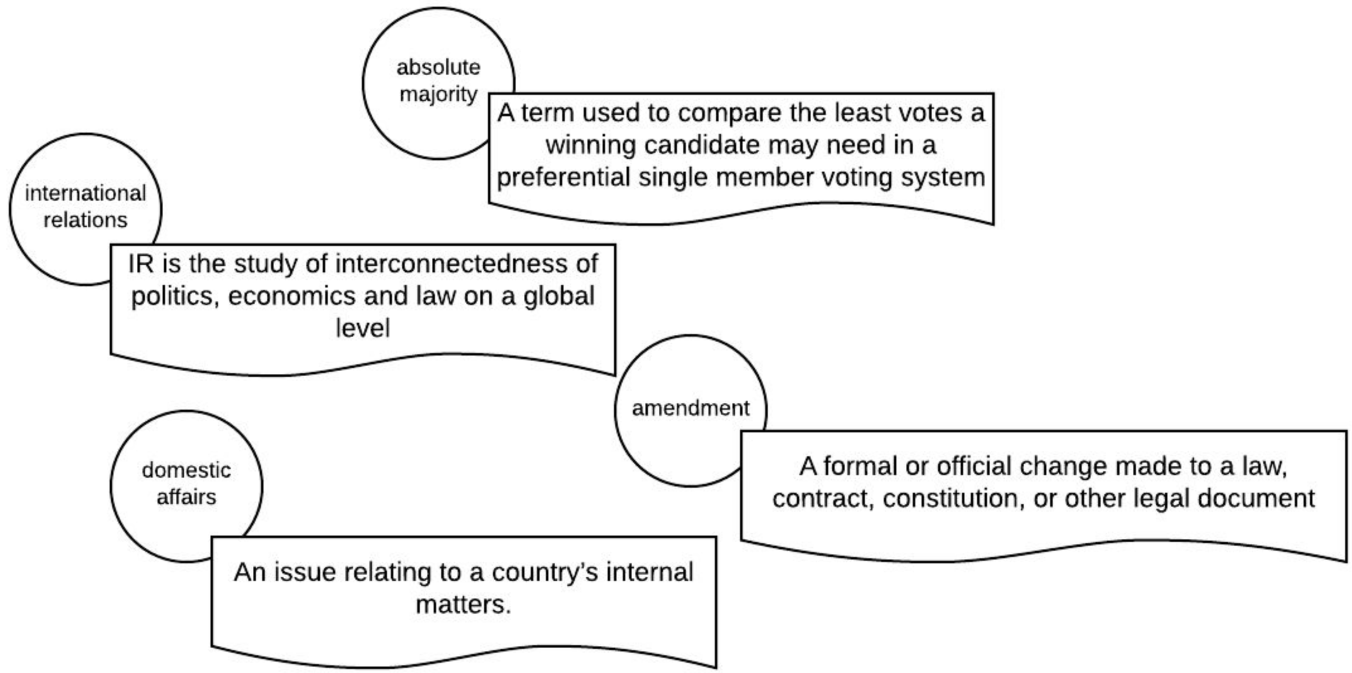
- Definiti… ural Langua…

Word	Definition
(1) Redundancy and overusing common phrases: 4.28%	
propane	a volatile flammable gas that is used to burn gas
(2) Self-reference: 7.14%	
precise	to make a precise effort
(3) Wrong part-of-speech: 4.29%	
accused	to make a false or unethical declaration of
(4) Under-specified: 30.00%	
captain	a person who is a member of a ship
(5) Opposite: 8.57%	
inward	not directed to the center
(6) Close semantics: 22.86%	
adorable	having the qualities of a child
(7) Incorrect: 32.14%	
incase	to make a sudden or imperfect sound

Table 9: Error types and examples.

- https://github.com/websail-nu/torch-defseq

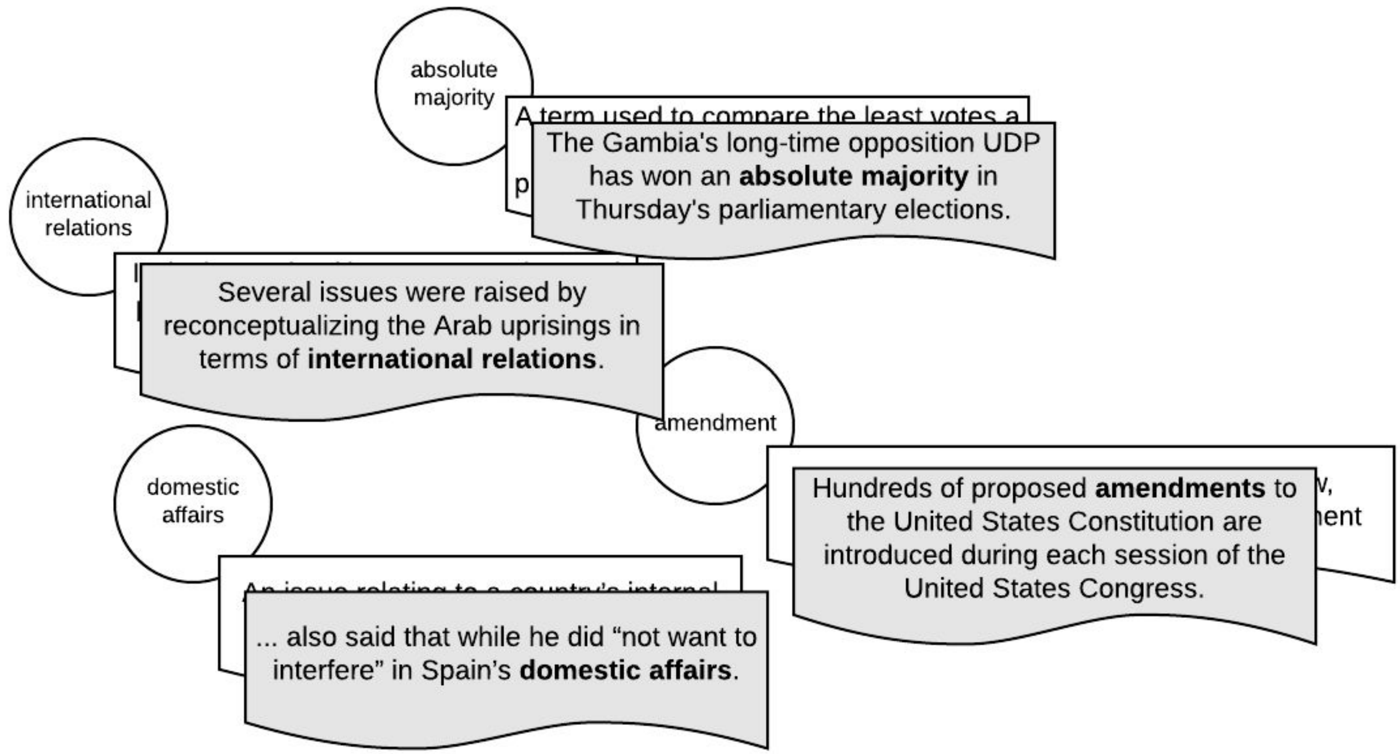
Definition Modeling



Dictionary Examples Acquisition

- GDEX: Automatically finding good dictionary examples in a corpus (Kilgariff et al., EURALEX 2008)
- A good dictionary example must be:
  - typical, exhibiting frequent and **well-dispersed patterns of usage**
  - **informative**, helping to elucidate the definition
  - **intelligible** to learners, **avoiding gratuitously difficult lexis** and structures, puzzling or distracting names, anaphoric references or other deictics which cannot be understood without access to the wider context. We call this its "**readability**".
- How?
  - Sentences between 10 and 25 words, using frequent words, etc.
- Is it successful?
  - "In sum: yes it worked, but we have an agenda for making it work better."

Dictionary Examples Acquisition



(Knowledge-based) Information Extraction

Extract *truth-bearers, beliefs, facts* in the form of n-ary relations involving a relation and a set of arguments. < Dante, wrote, Divine Comedy >

- **Open IE solves the problem of missing target relations** by identifying relation phrases, i.e., phrases that denote relations in English sentences (Banko et al., IJCAI 2007; Fader et al., EMNLP 2011).
- Typically address extraction at **surface form level**
  - Ambiguity, difficult to integrate in reference inventories, etc.
- *Knowledge-based* OIE reconciles data-driven OIE with high quality curated knowledge.

(Knowledge-based) Information Extraction

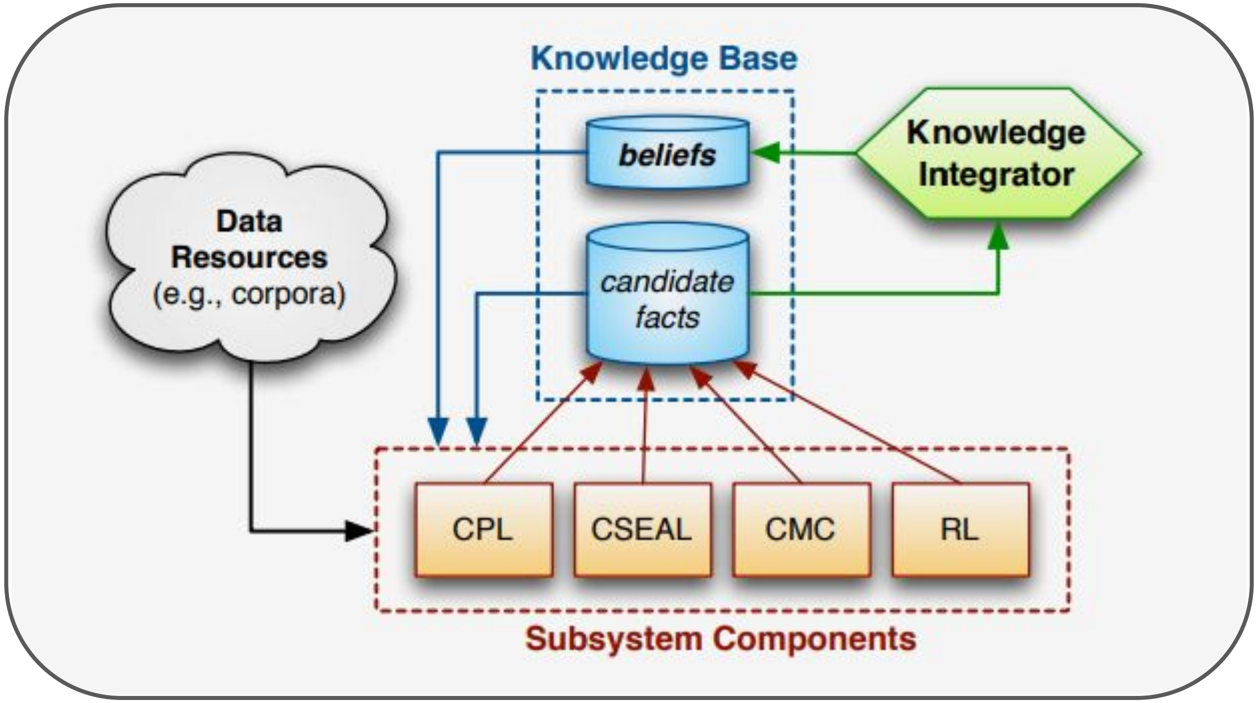
- **NELL** (Carlson et al., 2010)

- **PATTY** (Nakashole et al., 2012)

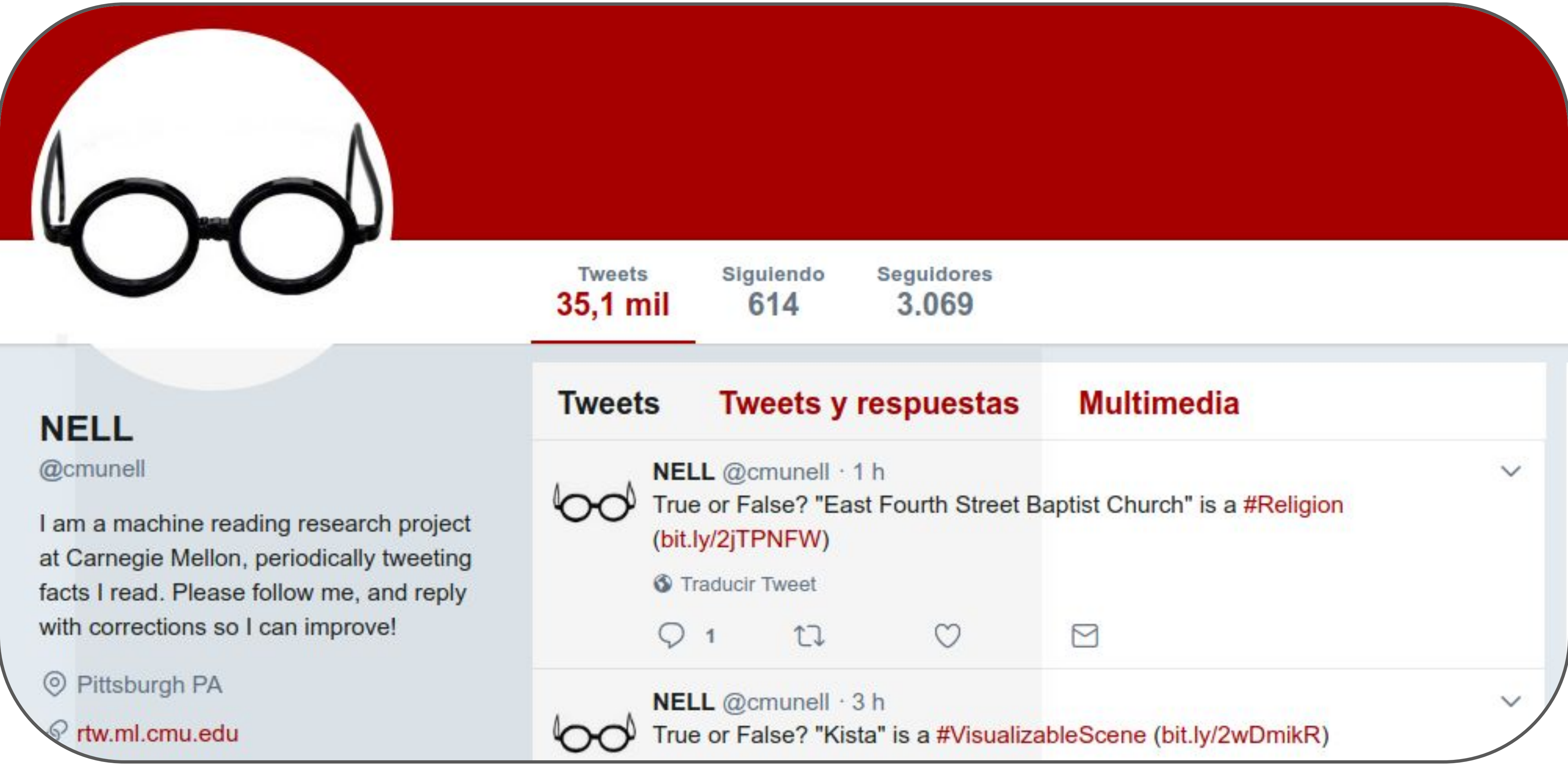
- **DefIE** and **KB-UNIFY** (Delli Bovi et al., 2015a, 2015b)

(Knowledge-based) Information Extraction

OIE with a semantic pivot: **NELL** (AAAI Carlson et al., 2010).



(Knowledge-based) Information Extraction



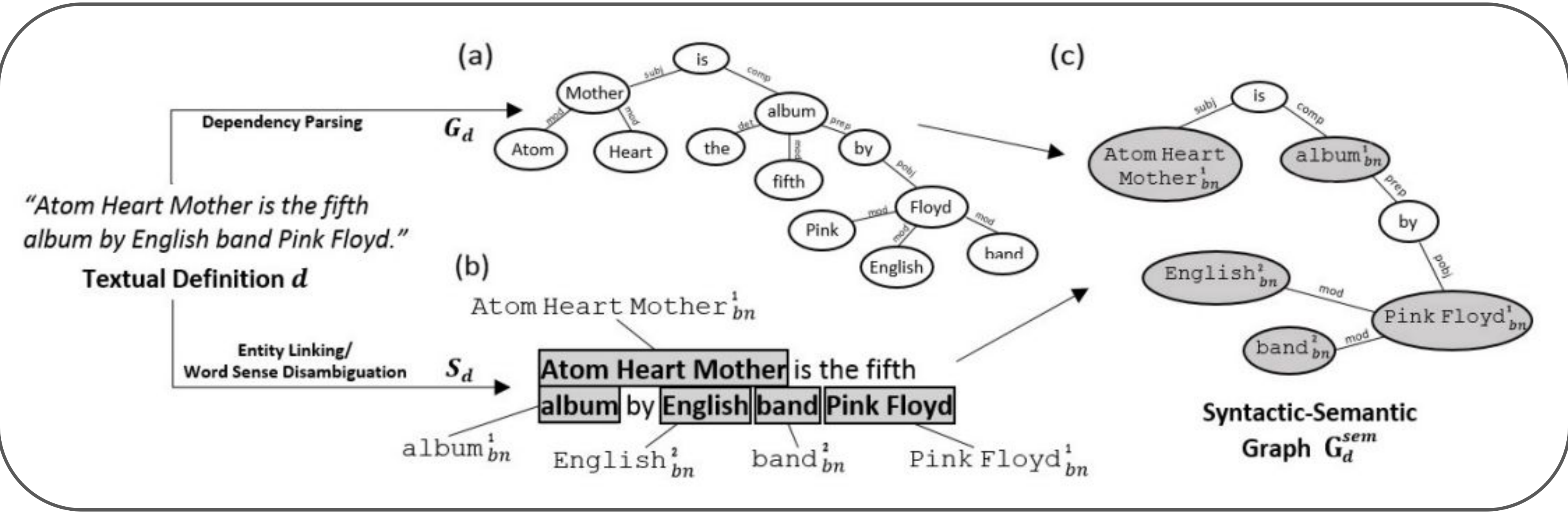
(Knowledge-based) Information Extraction

- PATTY (Nakashole et al., EMNLP 2012) introduces relation synsets

Relation	Paraphrases	Precision	Sample Paraphrases
DBPedia/artist	83	0.96±0.03	[adj] studio album of, [det] song by ...
DBPedia/associatedBand	386	0.74±0.11	joined band along, plays in ...
DBPedia/doctoralAdvisor	36	0.558±0.15	[det] student of, under * supervision ...
DBPedia/recordLabel	113	0.86±0.09	[adj] artist signed to, [adj] record label ...
DBPedia/riverMouth	31	0.83±0.12	drains into, [adj] tributary of ...
DBPedia/team	1,108	0.91±0.07	be * traded to, [prp] debut for ...
YAGO/actedIn	330	0.88±0.08	starred in * film, [adj] role for ...
YAGO/created	466	0.79±0.10	founded, 's book ...
YAGO/isLeaderOf	40	0.53±0.14	elected by, governor of ...
YAGO/holdsPoliticalPosition	72	0.73±0.10	[prp] tenure as, oath as ...

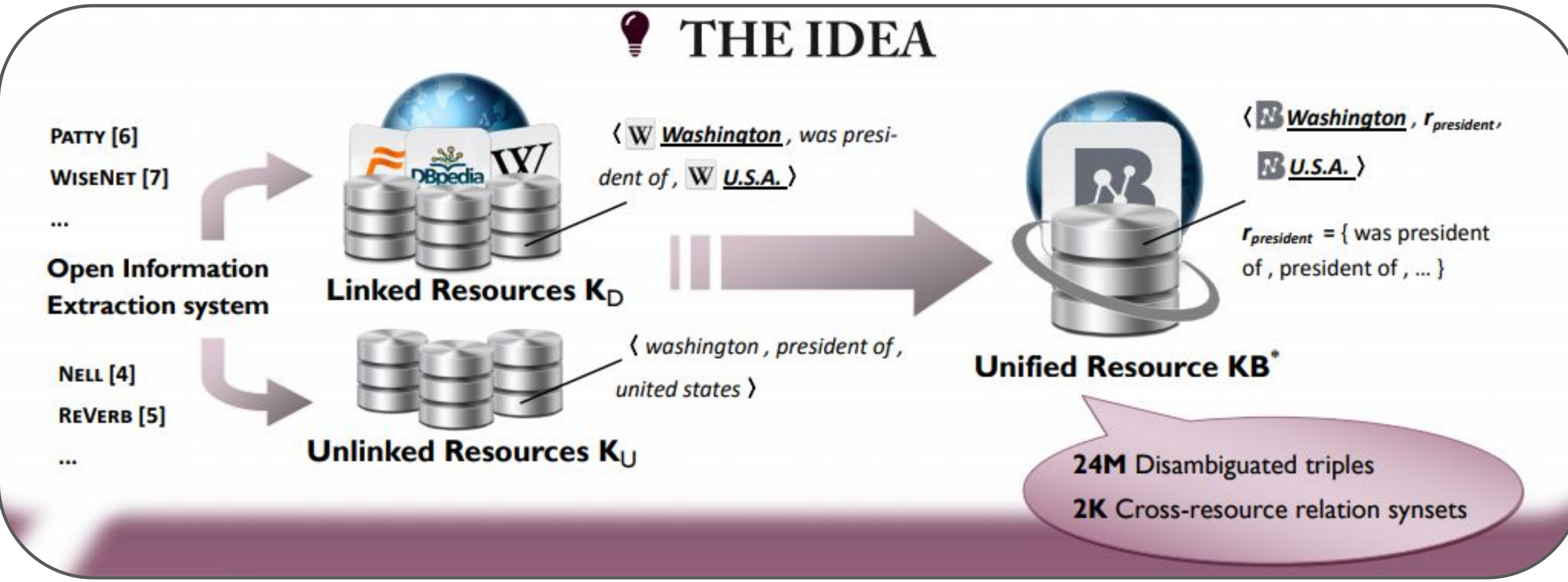
(Knowledge-based) Information Extraction

- DefIE (Delli Bovi et al., TACL 2015)
  - http://lcl.uniroma1.it/defie/

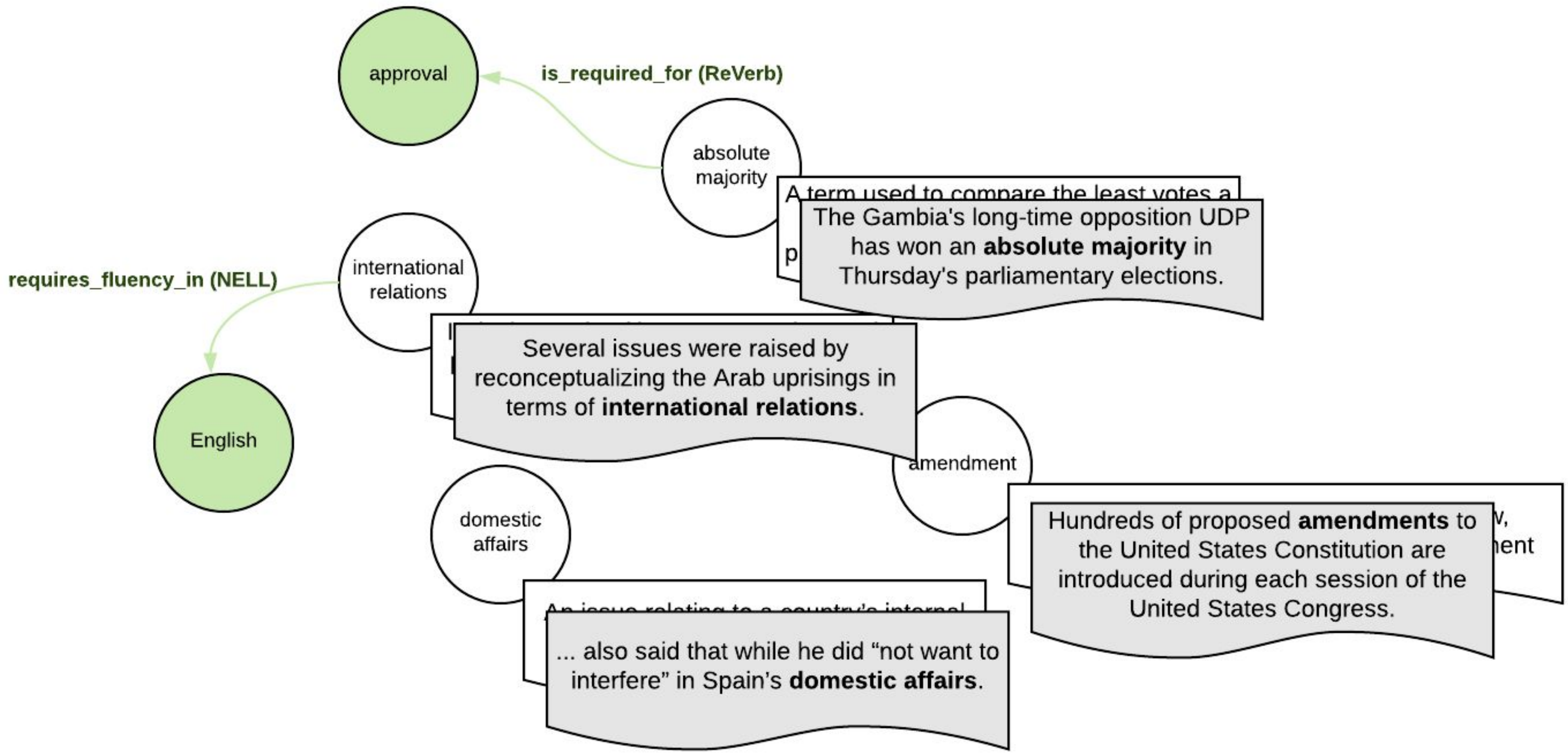


(Knowledge-based) Information Extraction

- KB-UNIFY (Delli Bovi et al., EMNLP 2015)
  - http://lcl.uniroma1.it/kb-unify/



LR Extension via OIE



Hypernymy

- Important phenomenon, **backbone relation** in taxonomies and ontologies.
- In NLP, **different sub-tasks**, e.g., *hypernym detection*, *extraction*, *discovery*, *taxonomy learning*, etc.
- Natural **applications** in semantic search, machine translation, semantic similarity, disambiguation, and even useful to explore bias in AI/ML.

Fruit

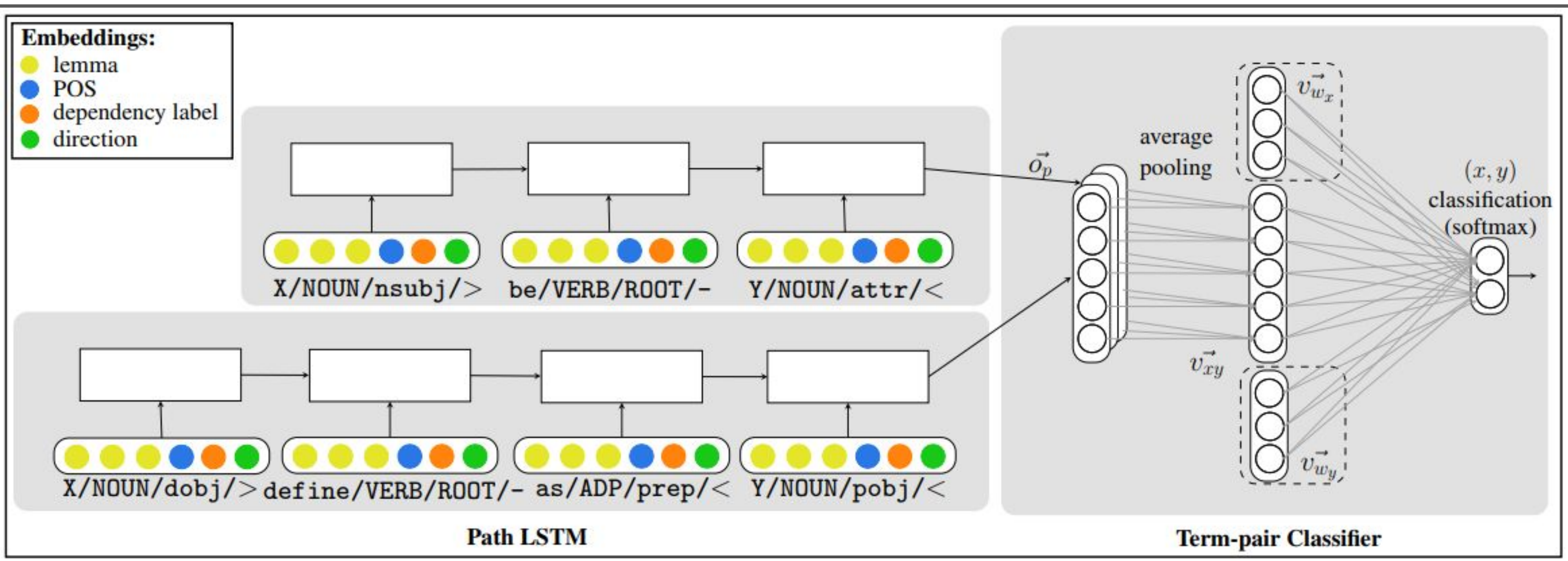
is a

Apple

SemEval 2018 Shared Task on Hypernym Discovery

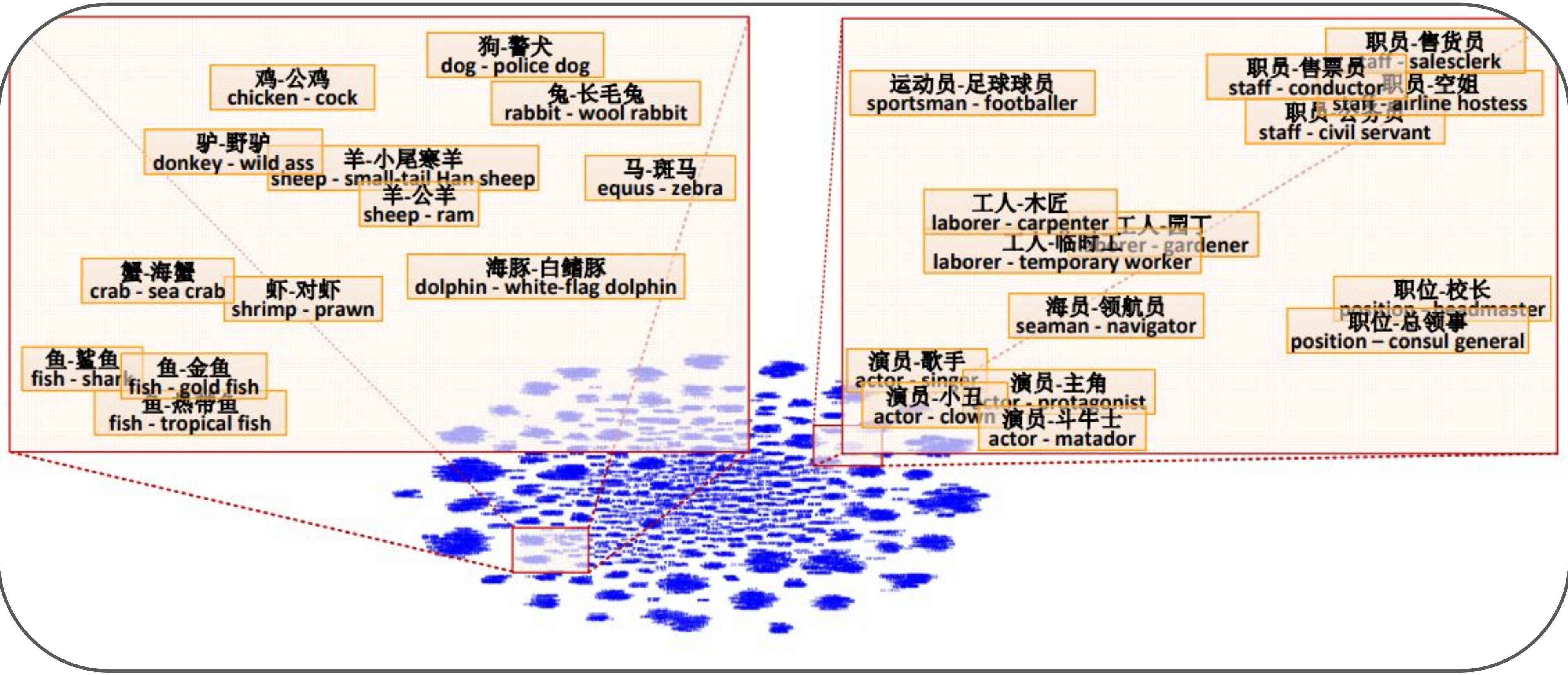
Hypernymy

- Hypernym Detection
- Example from Shwartz et al. (ACL 2016)



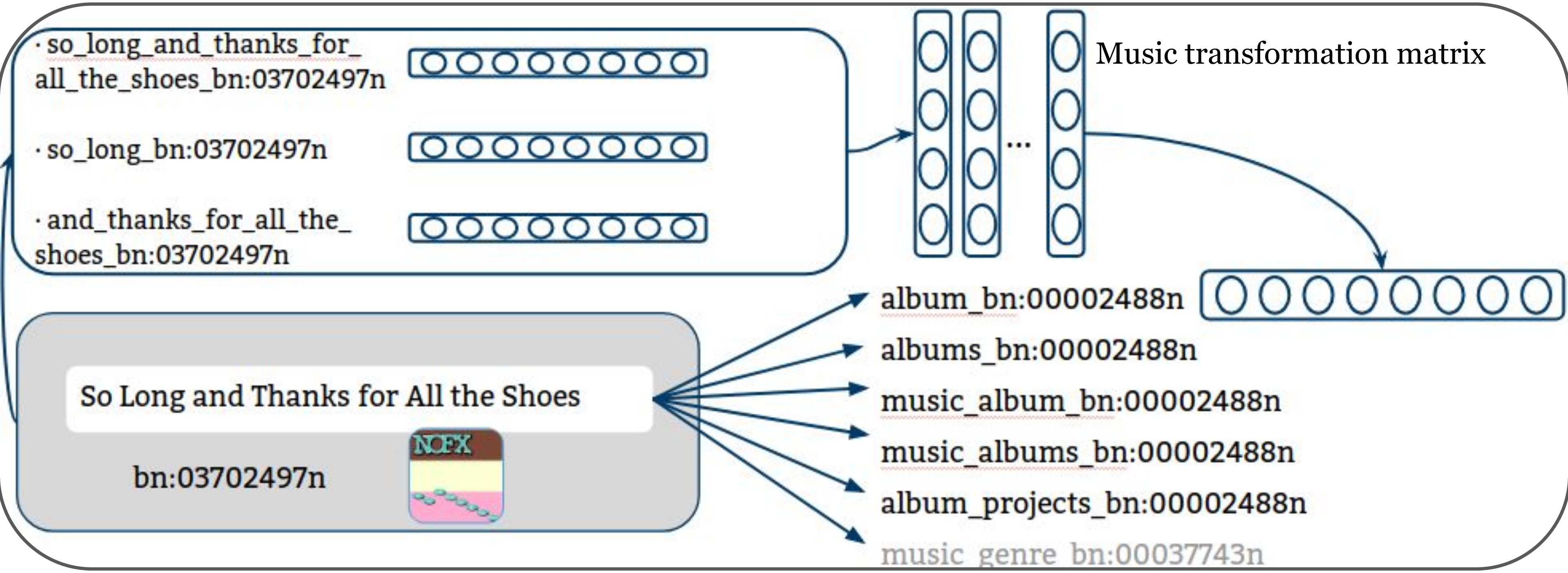
Hypernymy

- Hypernym Detection ~
- Example from Fu et al. (ACL 2014)

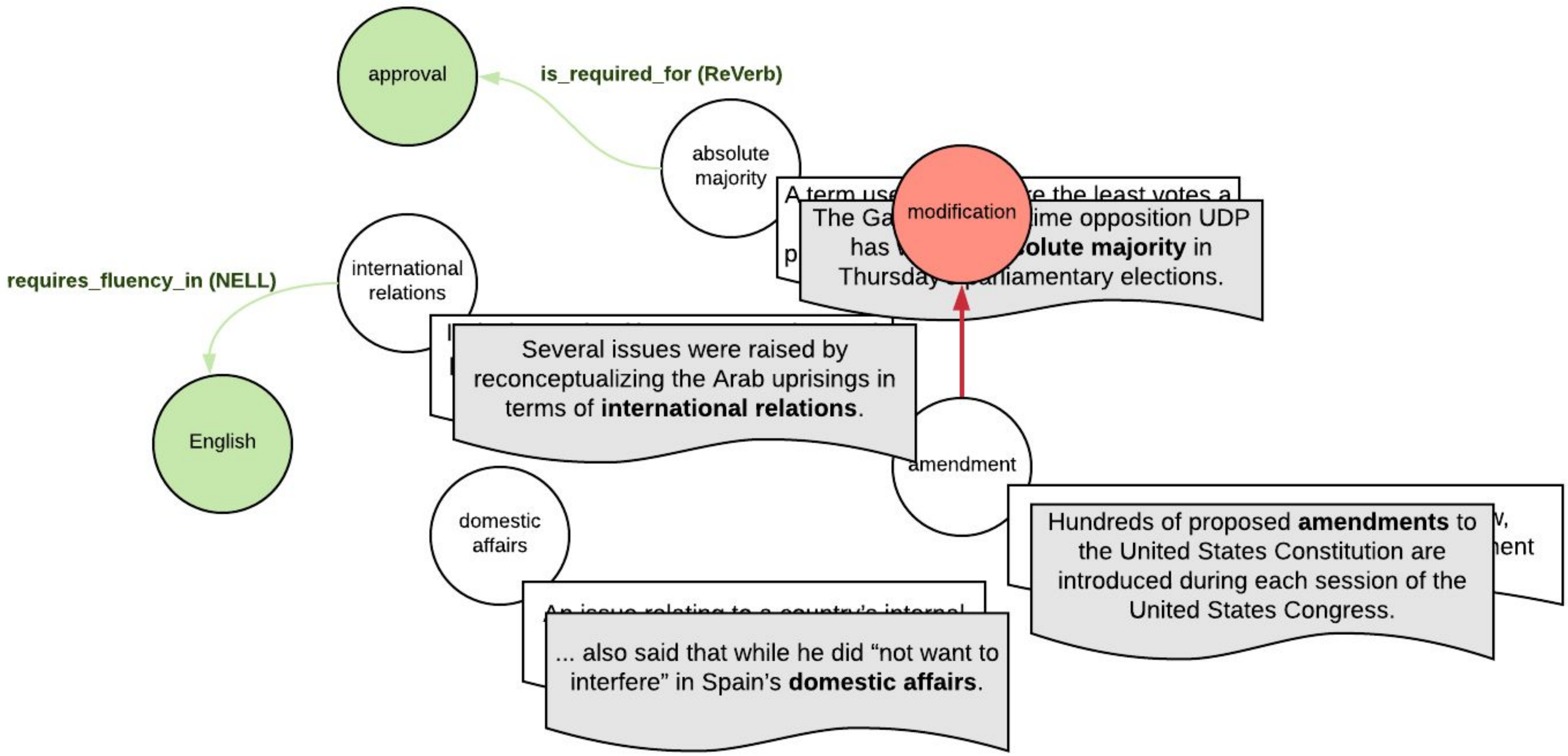


Hypernymy

- Hypernym Discovery
- Example from Espinosa-Anke et al. (EMNLP 2016)



LR Extension via Hypernymy Modeling



Topic/Domain Clustering

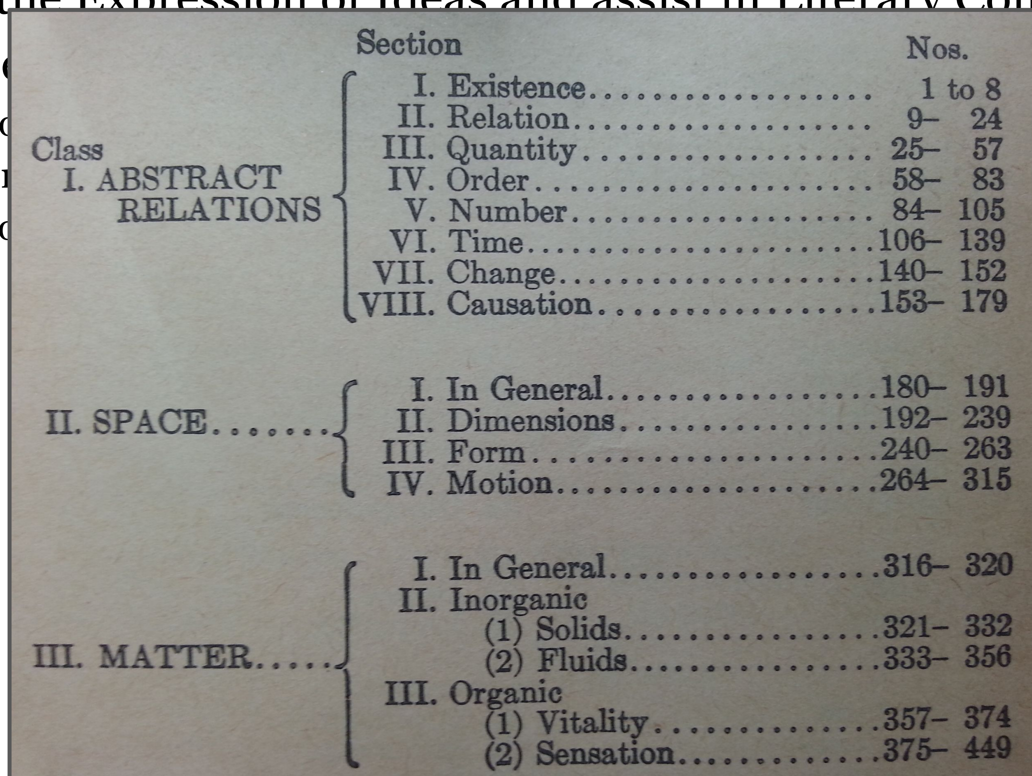
- Roget's Thesaurus

- From Wikipedia:
  - (...) tree containing over a thousand branches for individual "meaning clusters" or semantically linked words. Although these words are not strictly synonyms, they can be viewed as colours or connotations of a meaning or as a spectrum of a concept

- "Dr Roget's Thesaurus of English Words and Phrases: classified and arranged to facilitate the Expression of Ideas and assist in Literary Composition" (1852)

Topic/Domain Clustering

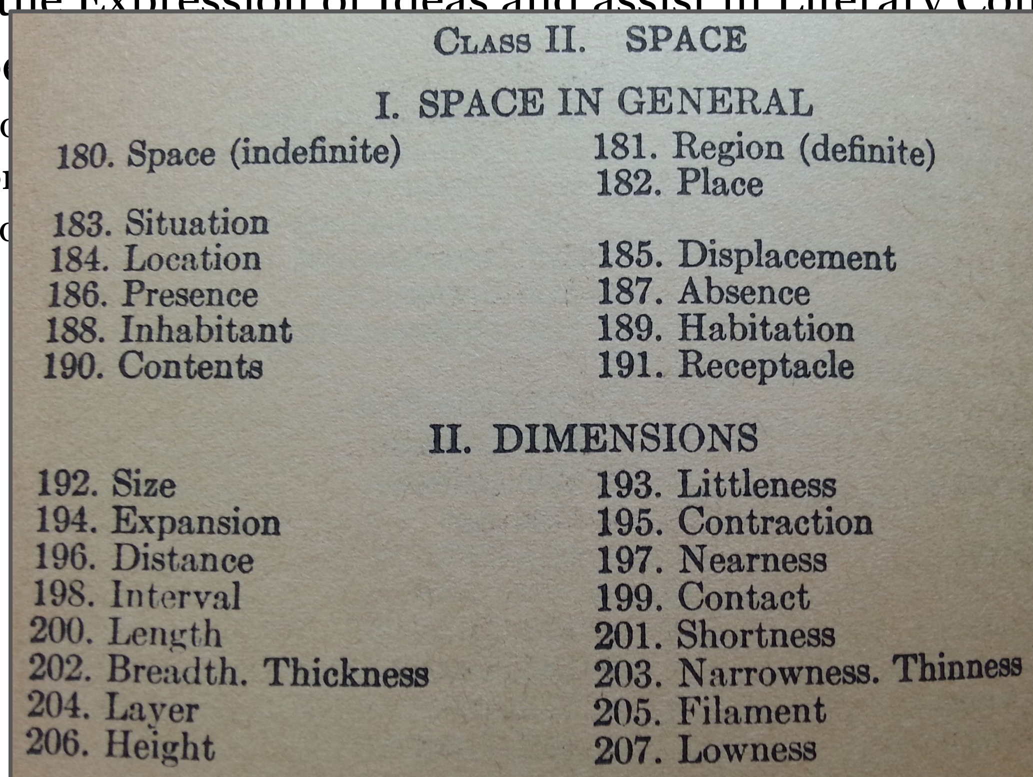
- Roget’s Thesaurus
- “Dr Roget's Thesaurus of English Words and Phrases classified and arranged to facilitate the Expression of Ideas and assist in Literary Composition” (1852)
- From Wikipe
  - (...) tree co sters" or semantically linked wor e viewed as colours or connotatio



Class	Section	Nos.
I. ABSTRACT RELATIONS	I. Existence	1 to 8
	II. Relation	9– 24
	III. Quantity	25– 57
	IV. Order	58– 83
	V. Number	84– 105
	VI. Time	106– 139
	VII. Change	140– 152
	VIII. Causation	153– 179
II. SPACE	I. In General	180– 191
	II. Dimensions	192– 239
	III. Form	240– 263
	IV. Motion	264– 315
III. MATTER	I. In General	316– 320
	II. Inorganic (1) Solids	321– 332
	(2) Fluids	333– 356
	III. Organic (1) Vitality	357– 374
	(2) Sensation	375– 449

Topic/Domain Clustering

- Roget’s Thesaurus
- “Dr Roget's Thesaurus of English Words and Phrases classified and arranged to facilitate the Expression of Ideas and assist in Literary Composition” (1852)
- From Wikipe
  - (...) tree co sters" or semantically linked wor e viewed as colours or connotatic



CLASS II. SPACE

I. SPACE IN GENERAL

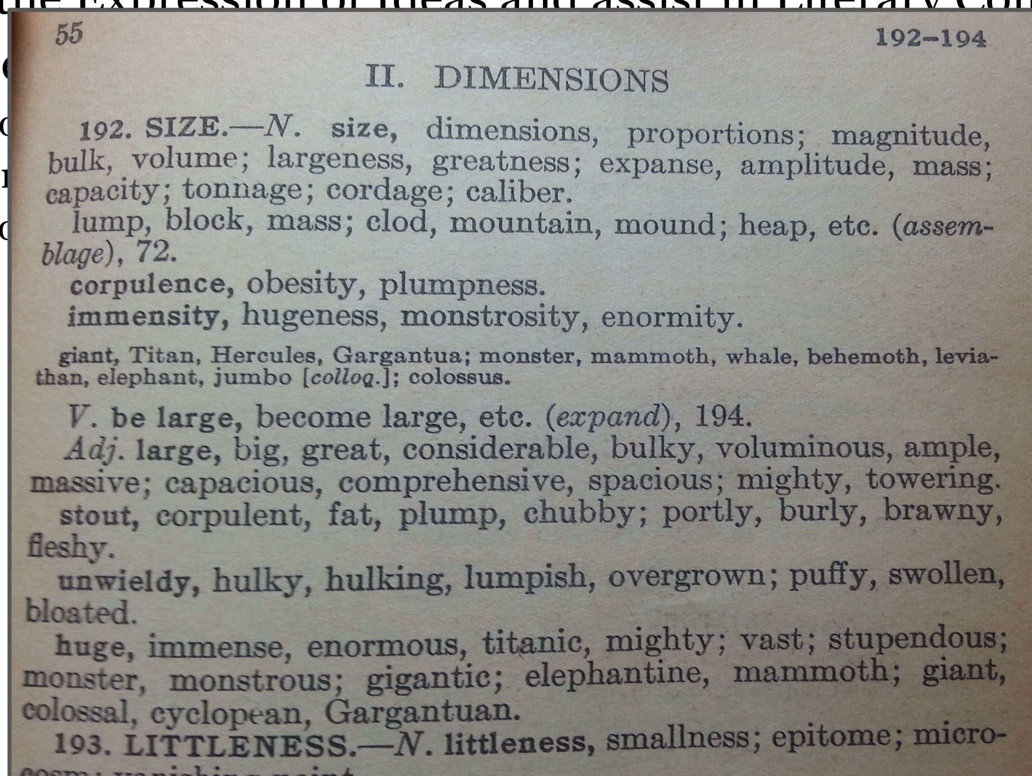
180. Space (indefinite)
181. Region (definite)
182. Place
183. Situation
184. Location
185. Displacement
186. Presence
187. Absence
188. Inhabitant
189. Habitation
190. Contents
191. Receptacle

II. DIMENSIONS

192. Size
193. Littleness
194. Expansion
195. Contraction
196. Distance
197. Nearness
198. Interval
199. Contact
200. Length
201. Shortness
202. Breadth. Thickness
203. Narrowness. Thinness
204. Layer
205. Filament
206. Height
207. Lowness

Topic/Domain Clustering

- Roget’s Thesaurus
- “Dr Roget's Thesaurus of English Words and Phrases classified and arranged to facilitate the Expression of Ideas and assist in Literary Composition” (1852)
- From Wikipe
    - (...) tree co sters" or semantically linked wor e viewed as colours or connotatio



55 192–194

II. DIMENSIONS

192. SIZE.—*N.* **size,** dimensions, proportions; magnitude, bulk, volume; largeness, greatness; expanse, amplitude, mass; capacity; tonnage; cordage; caliber.

lump, block, mass; clod, mountain, mound; heap, etc. (*assemblage*), 72.

corpulence, obesity, plumpness.

immensity, hugeness, monstrosity, enormity.

giant, Titan, Hercules, Gargantua; monster, mammoth, whale, behemoth, leviathan, elephant, jumbo [*colloq.*]; colossus.

V. **be large,** become large, etc. (*expand*), 194.

Adj. **large,** big, great, considerable, bulky, voluminous, ample, massive; capacious, comprehensive, spacious; mighty, towering.

stout, corpulent, fat, plump, chubby; portly, burly, brawny, fleshy.

unwieldy, hulky, hulking, lumpish, overgrown; puffy, swollen, bloated.

huge, immense, enormous, titanic, mighty; vast; stupendous; monster, monstrous; gigantic; elephantine, mammoth; giant, colossal, cyclopean, Gargantuan.

193. LITTLENESS.—*N.* **littleness,** smallness; epitome; microcosm; vanishing point

Domain labeling

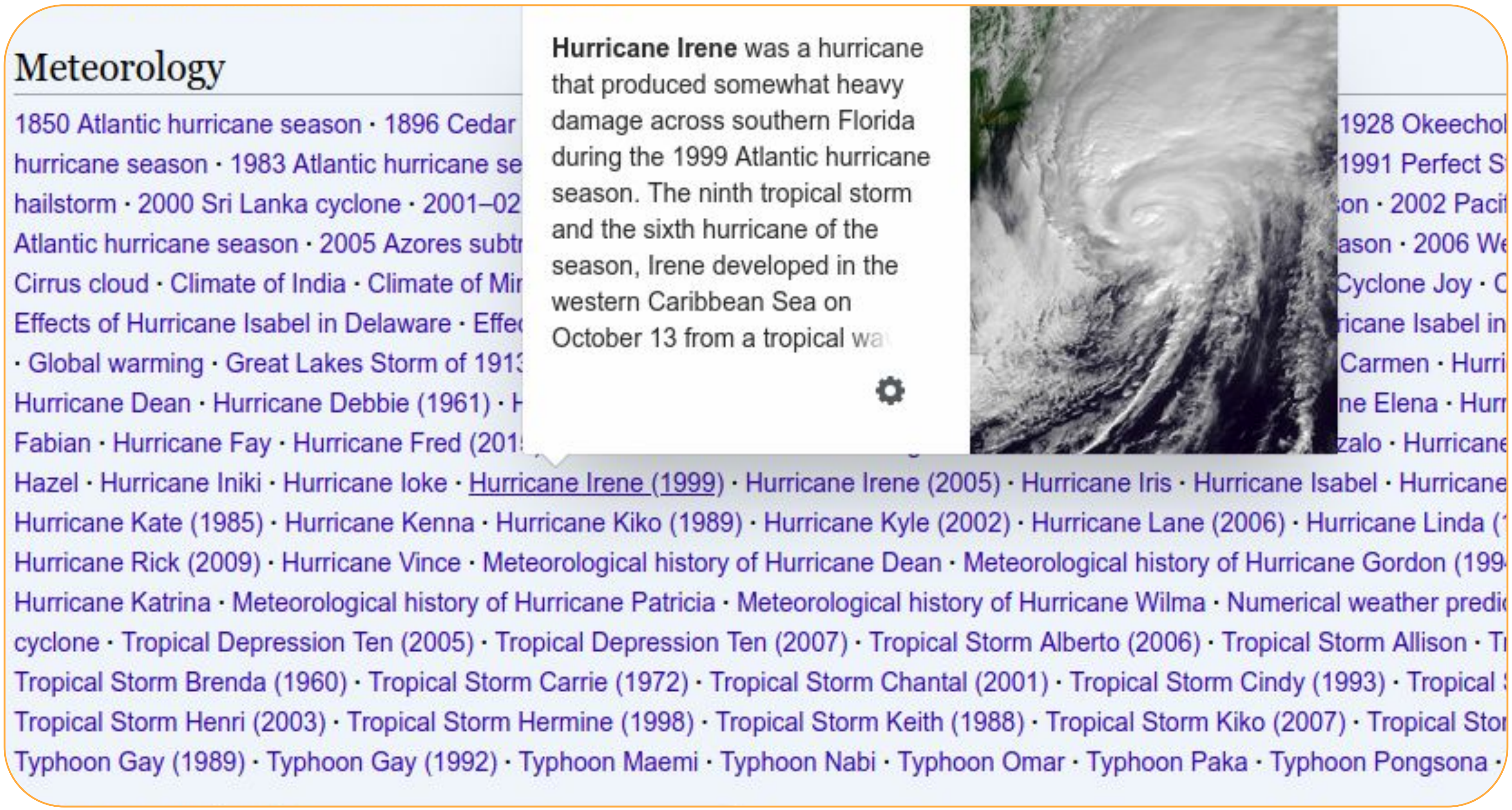
(Camacho-Collados and Navigli, EACL 2017)

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: BabelDomains

http://lcl.uniroma1.it/babeldomains/

How to associate a concept with a domain?

- **A knowledge-based vector** (to be explained in the second part of the tutorial) for the concatenation of all Wikipedia pages associated with a given domain.

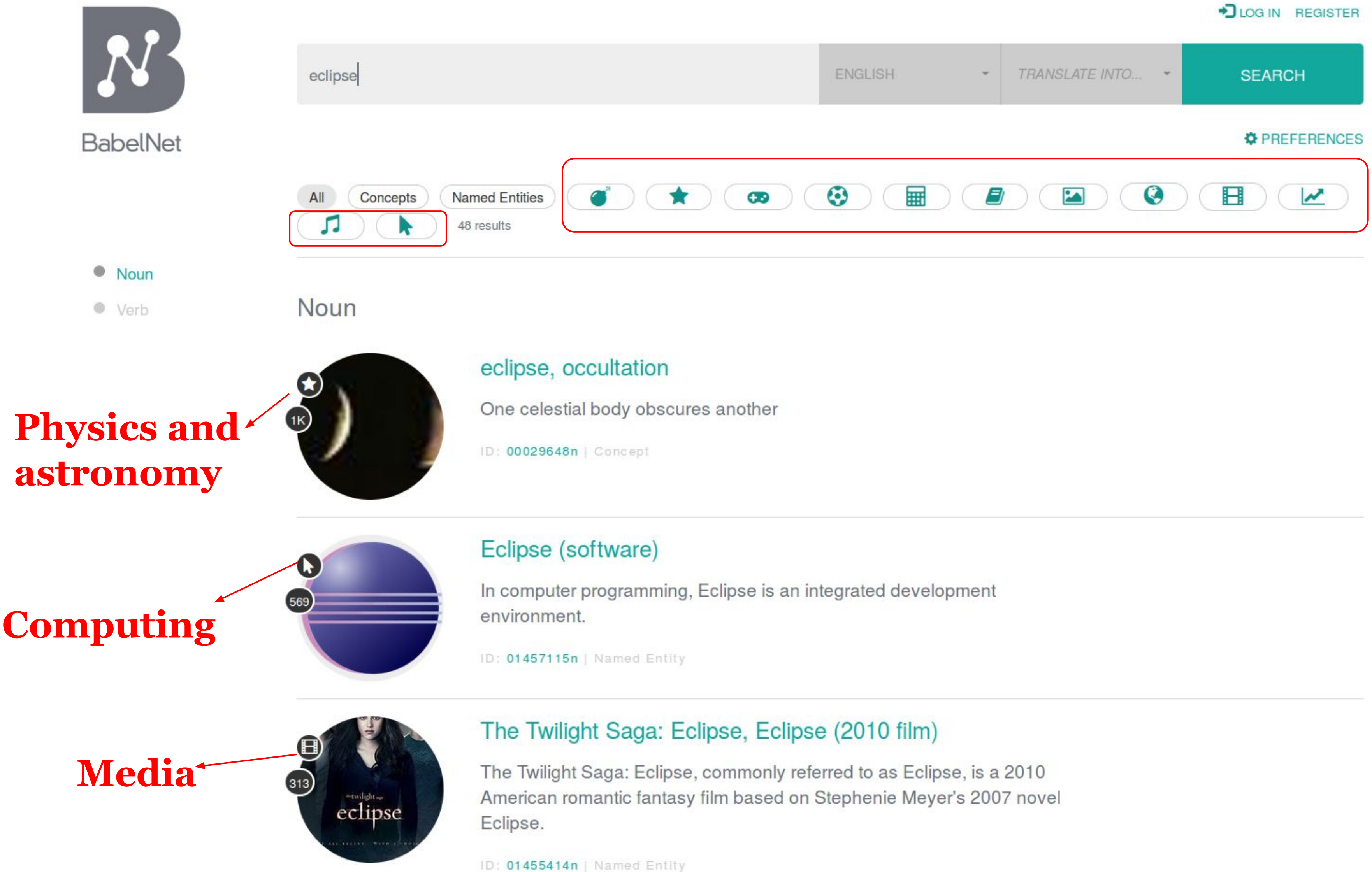
- Exploit the **semantic similarity** between knowledge-based vectors and **graph properties** of the lexical resources.

Domain labeling: BabelDomains

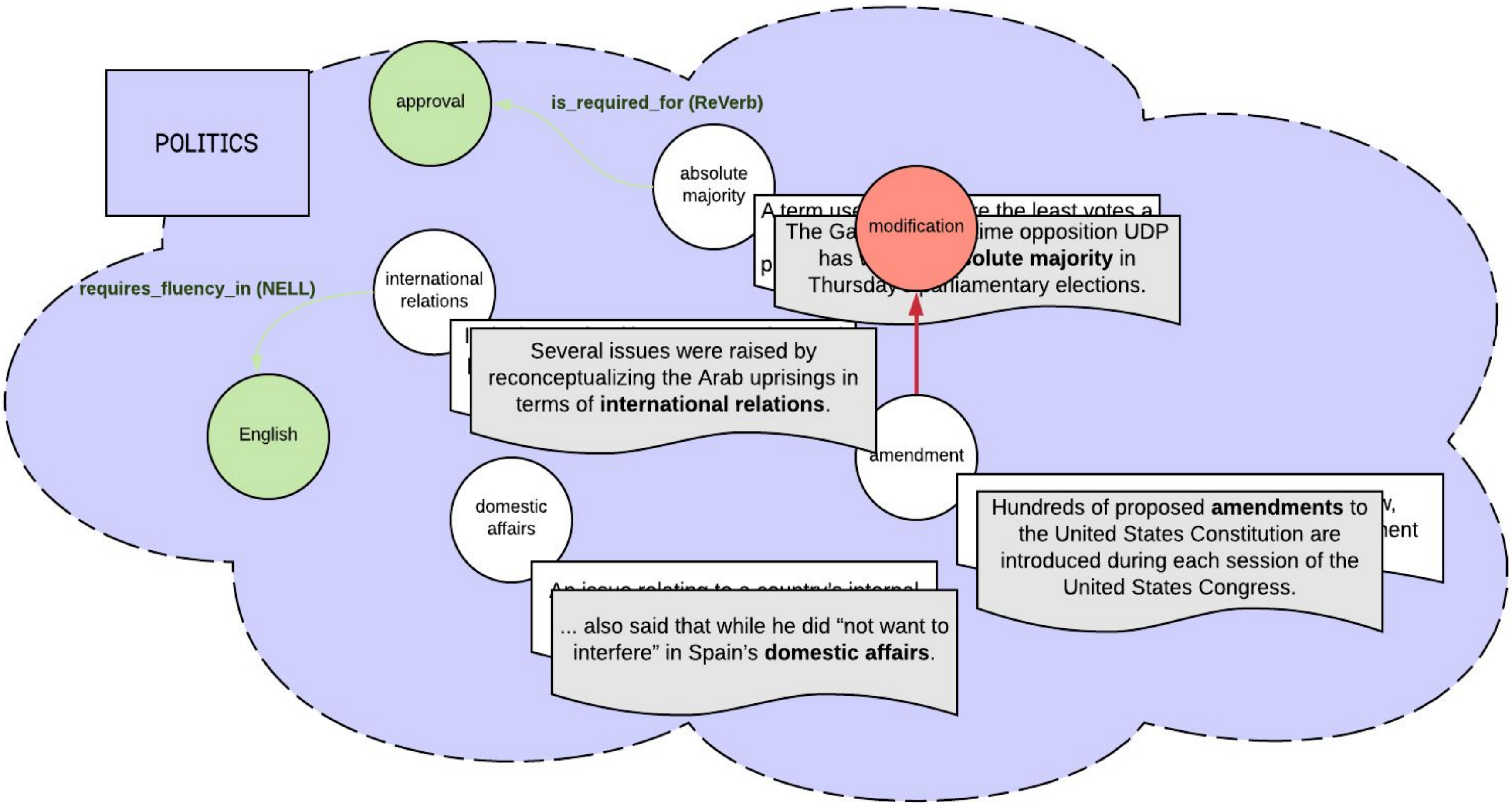
This results in **over 2.5M concepts and entities** associated with a domain of knowledge, including Wikipedia and WordNet.

This domain information has already been integrated into BabelNet.

Domain labeling: BabelDomains



Thematic / Topical Clustering



QUESTION 5
PIN: 7024700
www.kahoot.it

LEXICAL RESOURCES FOR NLP

Can lexical resources improve end-to-end models?

Can lexical resources improve end-to-end models?

YES!

Useful background knowledge that can be leveraged in complex tasks.

Introduction

End-to-end models easy to break. For example, lexical entailment models fail to capture sentence which require lexical and world knowledge (Glockner et al. ACL 2018).

- *James lives in fifth avenue.*

-> **ENTAILMENT**
OR
CONTRADICTION?

- *James lives in 6th avenue.*

Introduction

End-to-end models easy to break. For example, lexical entailment models fail to capture sentence which require lexical and world knowledge (Glockner et al. ACL 2018).

- *James lives in fifth avenue.*

-> **CONTRADICTION**

- *James lives in 6th avenue.*

Introduction

End-to-end models easy to break. For example, lexical entailment models fail to capture sentence which require lexical and world knowledge (Glockner et al. ACL 2018).

Knowledge from WordNet

Dominant Label	Category	Instances	Example Words	Decomposable Attention	ESIM	Residual Encoders	WordNet Baseline	KIM
	total	8,193		51.9%	65.6%	62.2%	85.8%	83.5%

End to end

Lexical Resources for NLP

- Word Sense Disambiguation (Entity Linking)

- Knowledge-based Sense and Concept Embeddings

- Integration of lexical resources into NLP downstream applications

Word Sense Disambiguation (Entity Linking)

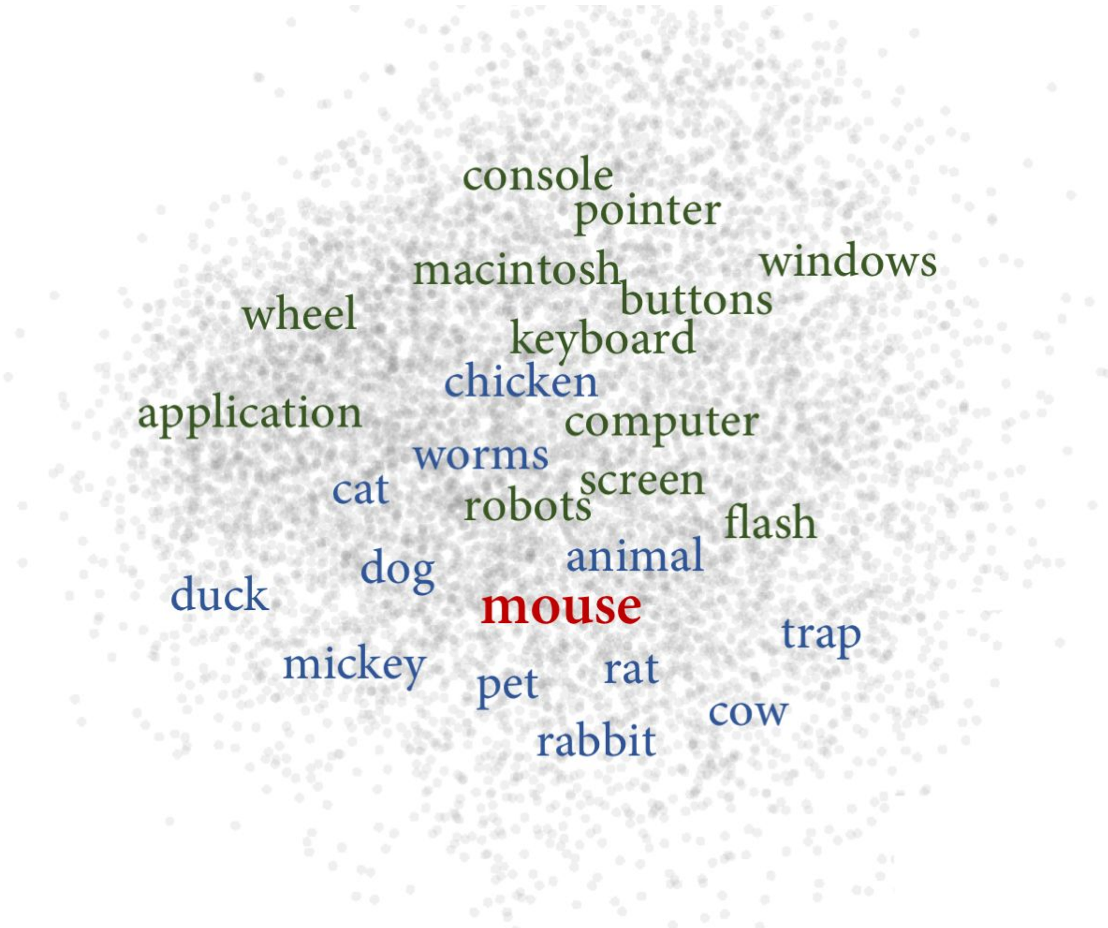
Word Sense Disambiguation (WSD)

Given the word in context, associate it with its most appropriate sense from a given sense inventory (e.g. WordNet)

- The **mouse** ate the cheese.

- A **mouse** consists of an object held in one's hand, with one or more buttons.

Word Sense Disambiguation (WSD)



Named Entity Disambiguation

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

?

Named Entity Disambiguation

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

Named Entity Disambiguation

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

Word Sense Disambiguation

- **Knowledge-based** (no sense-annotated data required)

- **Supervised** (use sense-annotated training data)

Word Sense Disambiguation

- **Knowledge-based**
  - Lesk-extended (Banerjee and Pedersen, 2003)
  - Lesk+emb (Basile et al., 2014)
  - UKB (Agirre et al., 2014)
  - Babelfy (Moro et al., 2014)

- **Supervised**

Knowledge-based WSD systems

Lesk (Lesk, 1986)

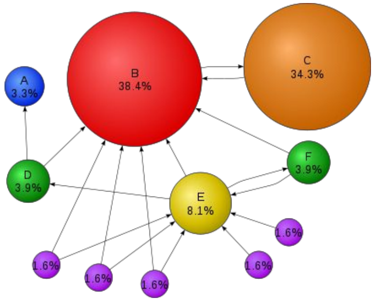
- Based on the **overlap between the definitions of a given sense and the context of the target word**. Two configurations:

- ***Lesk_extended*** (Banerjee and Pedersen, 2003): it includes related senses and tf-idf for word weighting.

- **Lesk+emb** (Basile et al., 2014): enhanced version of Lesk in which similarity between definitions and the target context is computed via word embeddings.

Knowledge-based WSD systems

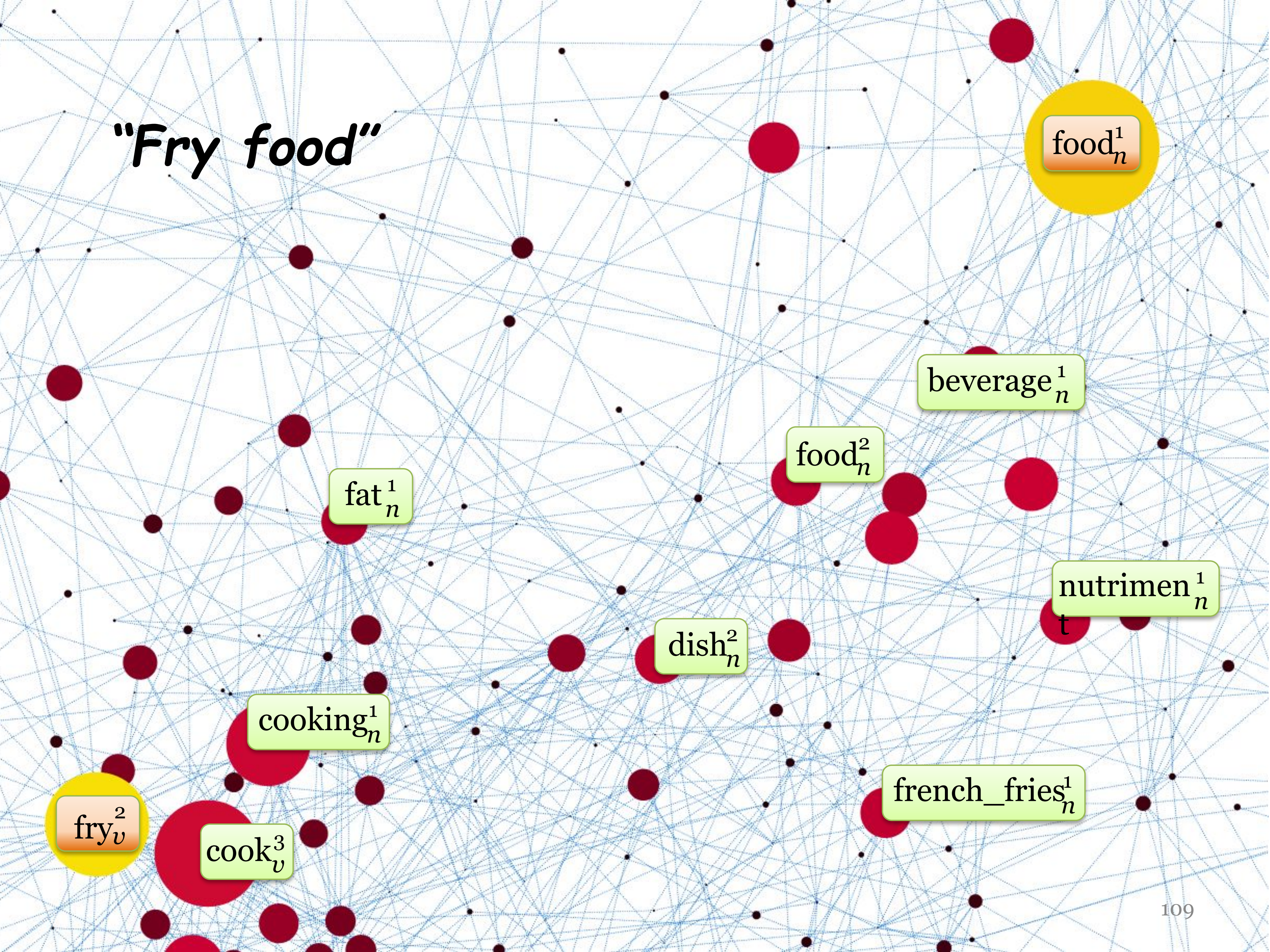
UKB (Agirre et al., CL 2014)



Graph-based system which exploits **random walks over a semantic network**, using Personalized PageRank.

It uses the standard WordNet graph plus disambiguated glosses as connections.

“Fry food”



food$_n^1$
beverage$_n^1$
food$_n^2$
fat$_n^1$
nutrimen$_n^1$
t
dish$_n^2$
cooking$_n^1$
french_fries$_n^1$
fry$_v^2$
cook$_v^3$

Knowledge-based WSD systems

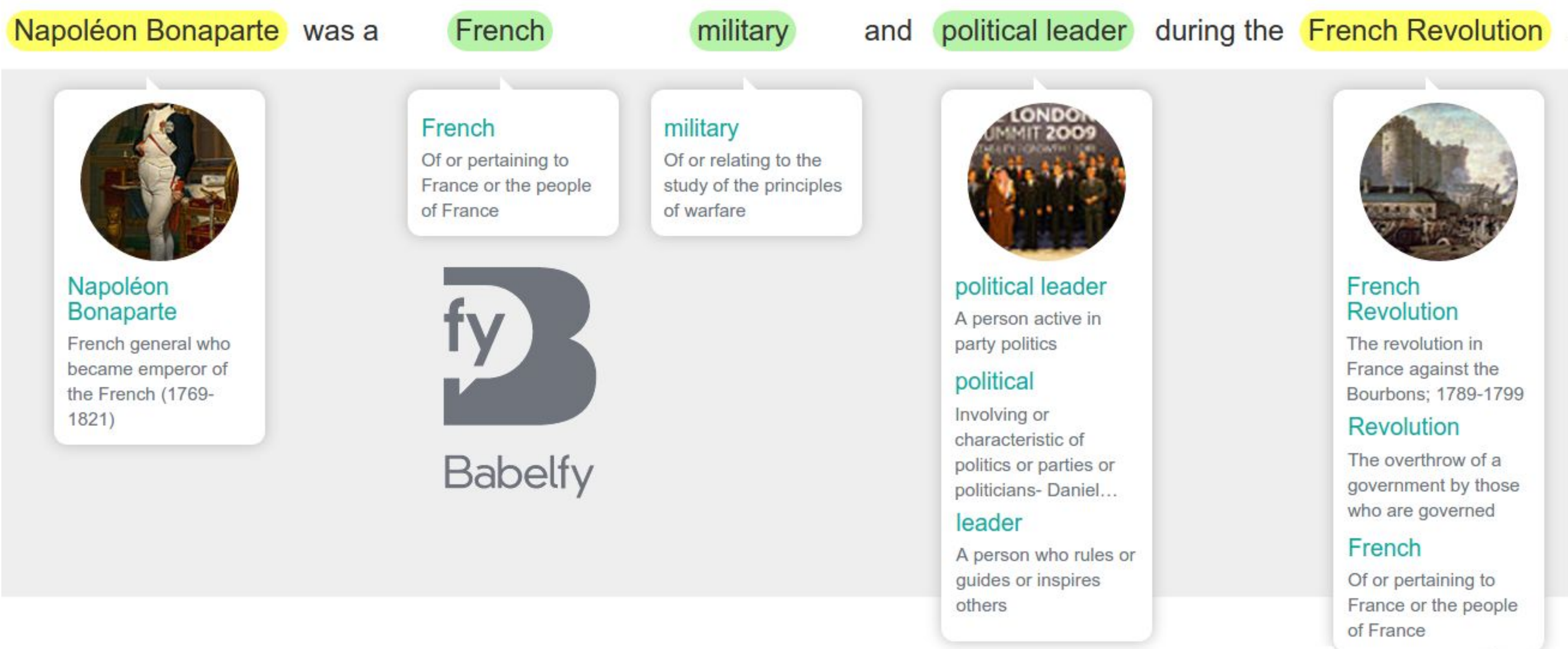
Babelfy (Moro et al., TACL 2014)

Graph-based system that uses **random walks with restart** over a semantic network, creating high-coherence semantic interpretations of the input text.

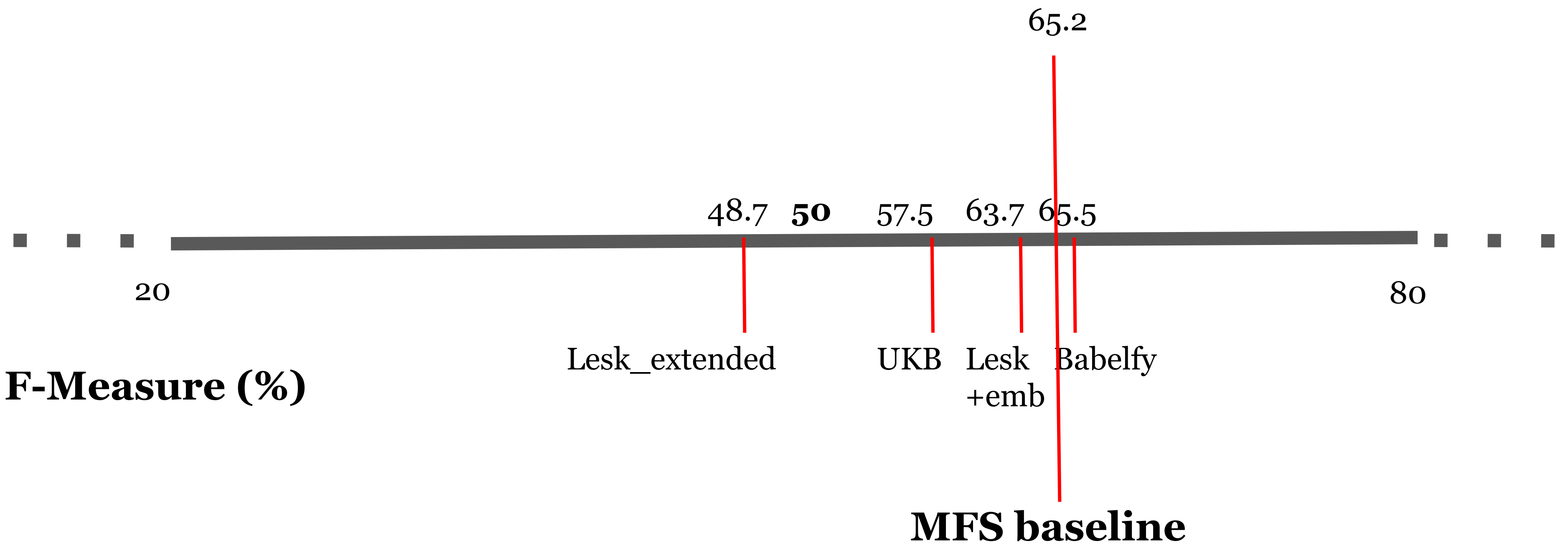
BabelNet as semantic network. BabelNet provides a large set of connections coming from Wikipedia and other resources.

Babelfy (Moro et al. TACL 2014)

Disambiguation and Entity Linking

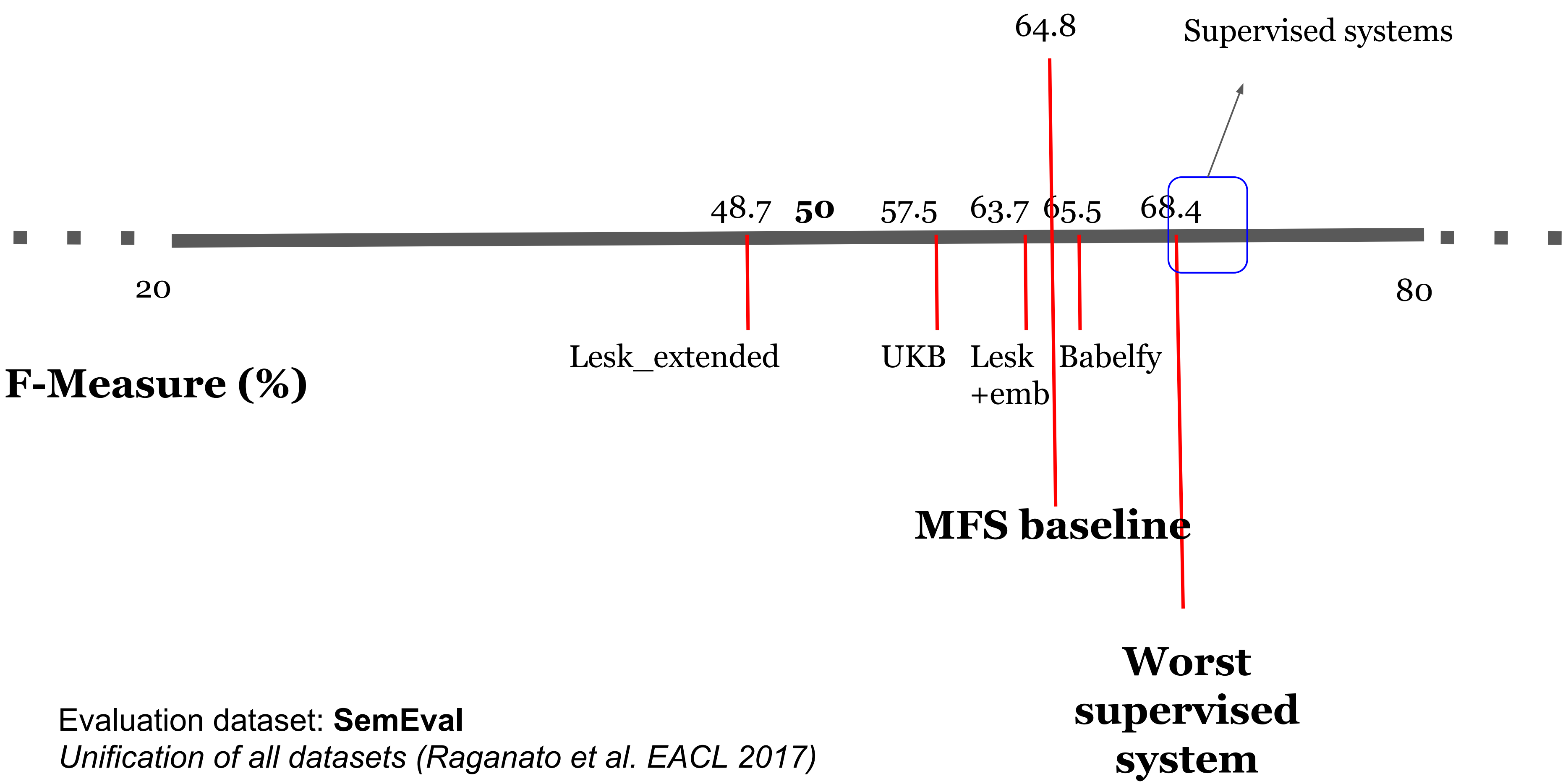


Knowledge-based WSD systems



Evaluation dataset: **SemEval**
Unification of all datasets (Raganato et al. EACL 2017)

Knowledge-based WSD systems



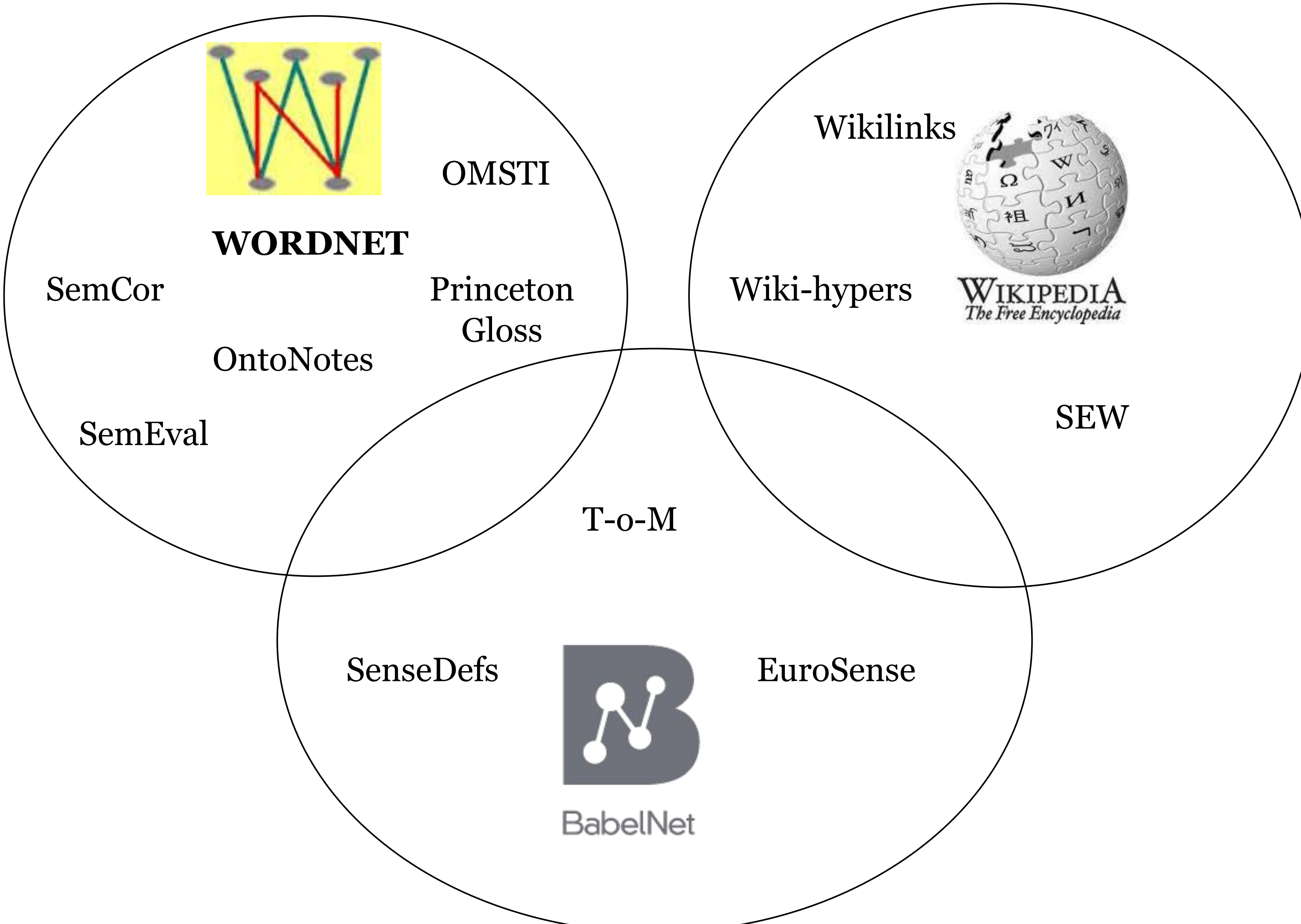
Evaluation dataset: **SemEval**
Unification of all datasets (Raganato et al. EACL 2017)

Word Sense Disambiguation

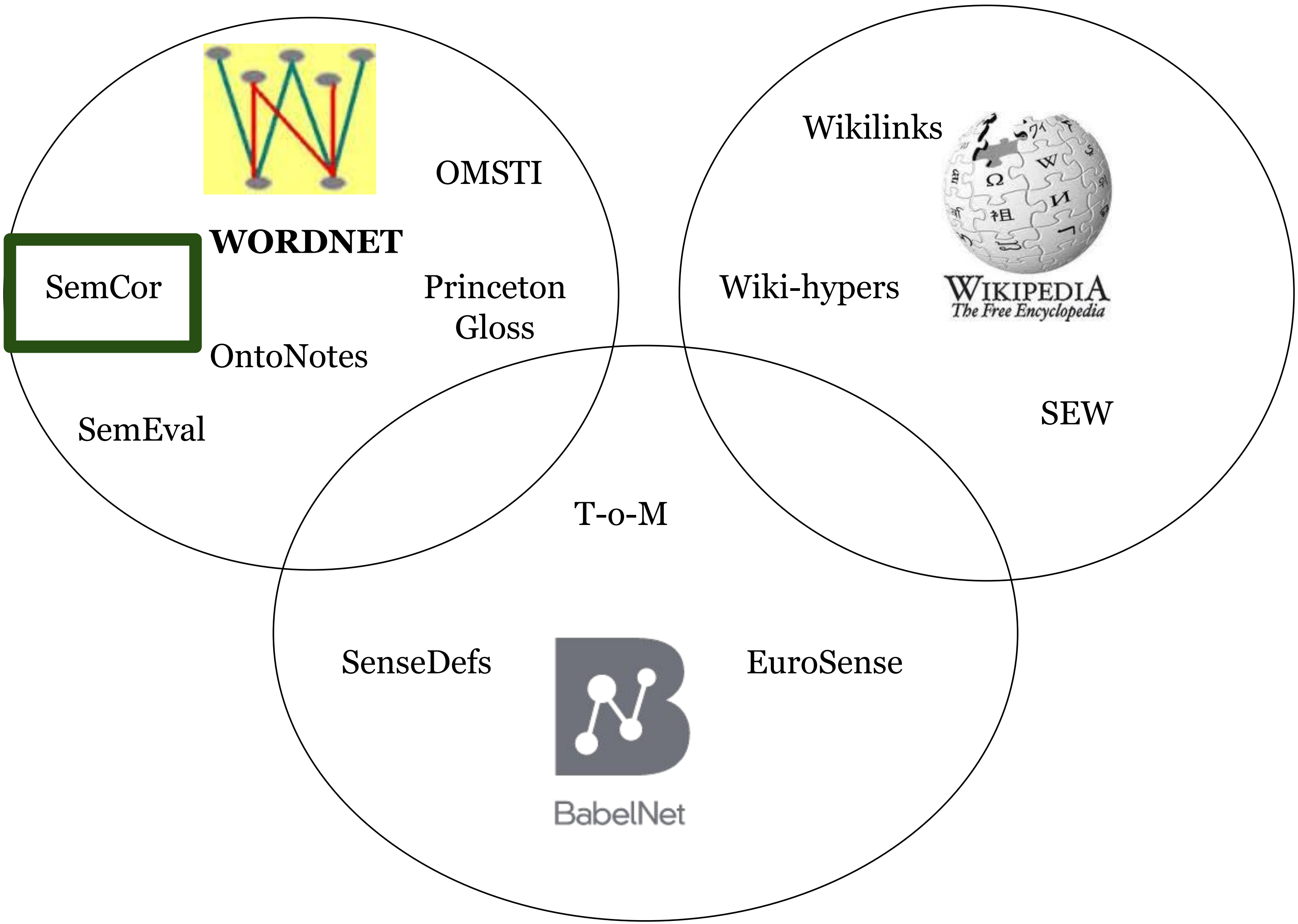
- **Knowledge-based**

- **Supervised**
  - IMS (Zhong and Ng, 2010)
  - IMS+emb (Iacobacci et al. 2016)
  - Context2Vec (Melamud et al., 2016)

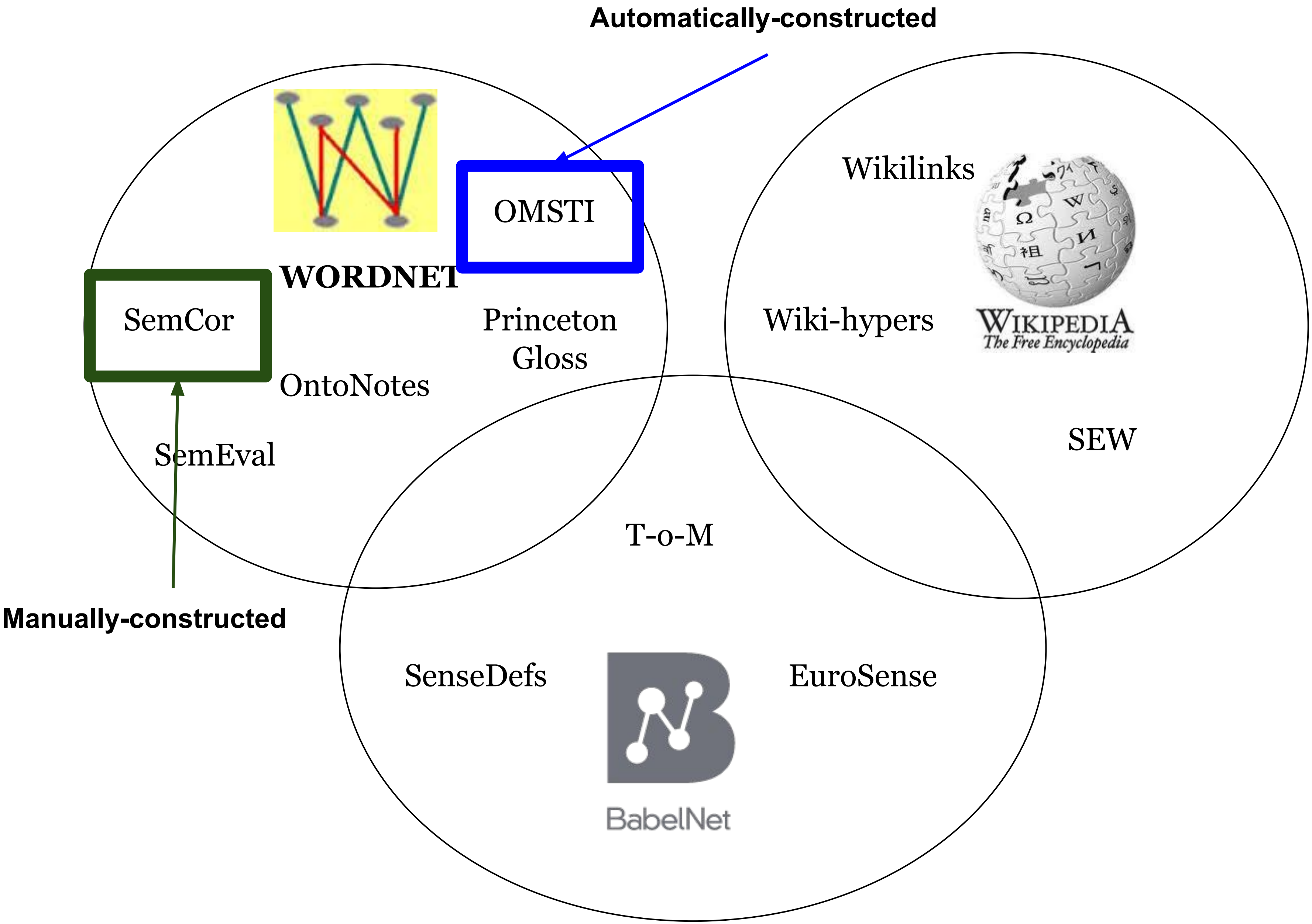
Training data



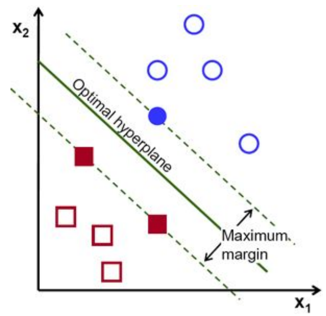
Training data



Training data



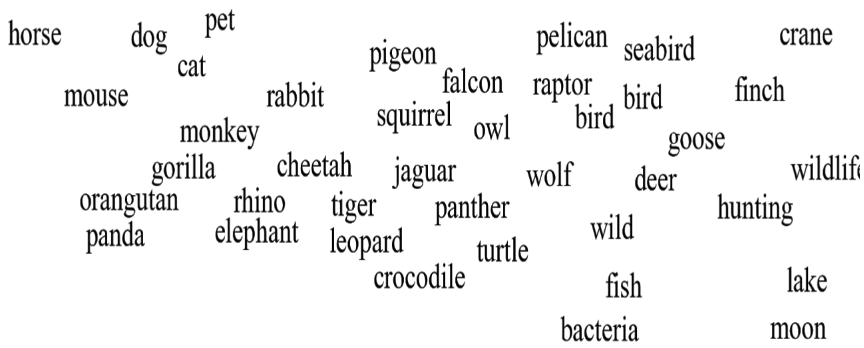
Supervised WSD systems



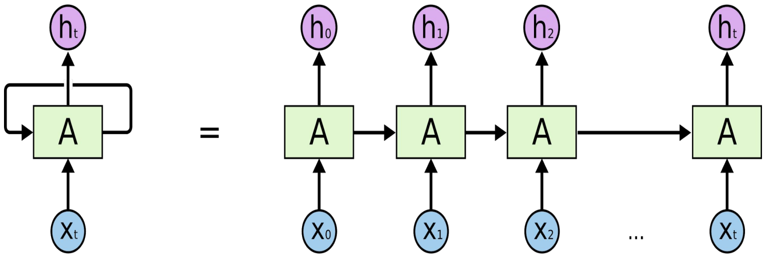
IMS (Zhong and Ng, ACL 2010)

SVM classifier over a set of conventional features: surroundings words, PoS tags and local collocations.

Improvements integrating **word embeddings** as an additional feature (Taghipour and Ng, 2015; Rothe and Schütze, 2015; Iacobacci et al. 2016) -> IMS+emb.



Supervised WSD systems



Context2Vec (Melamud et al., CoNLL 2016)

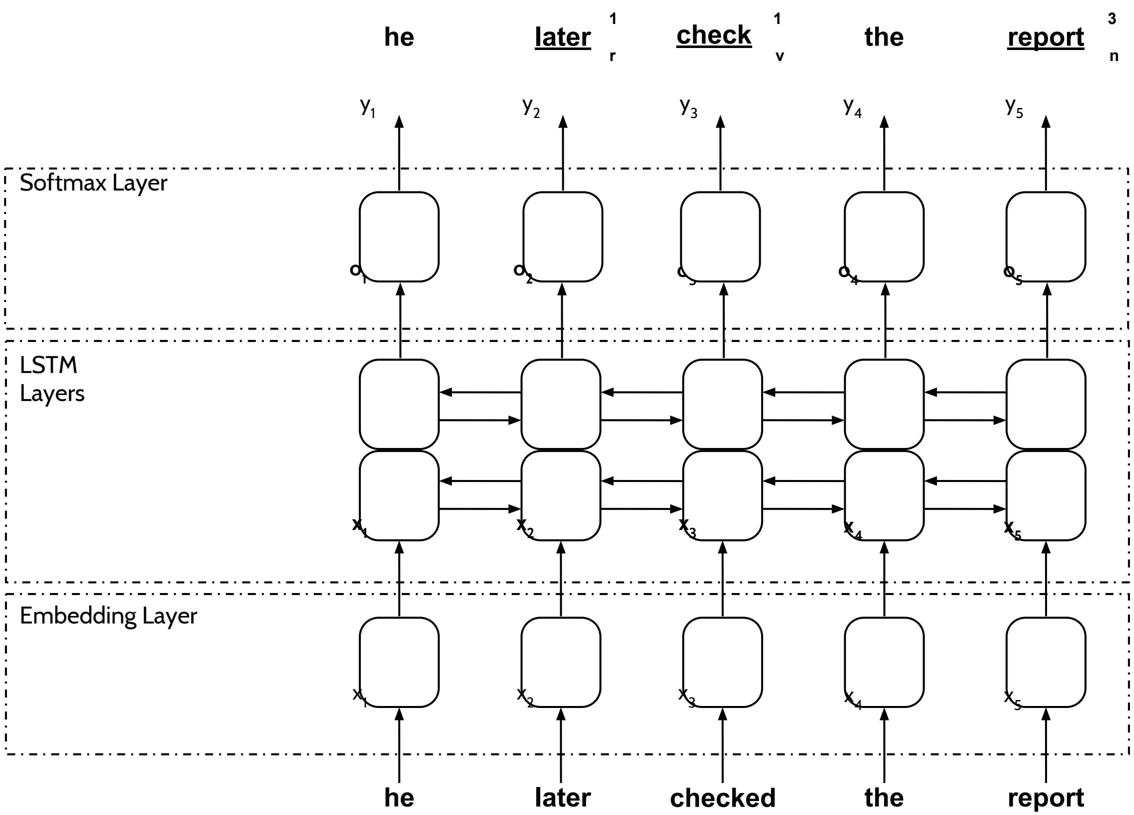
(Picture taken from Colah's blog)

Three steps:

- First, a **bidirectional LSTM** is trained on an unlabeled corpus.
- Then, this model is used to **learn an output (context) vector for each sense annotation** in the sense-annotated training corpus.
- Finally, the **sense annotation whose context vector is closer to the target word's context vector** is selected as the intended sense.

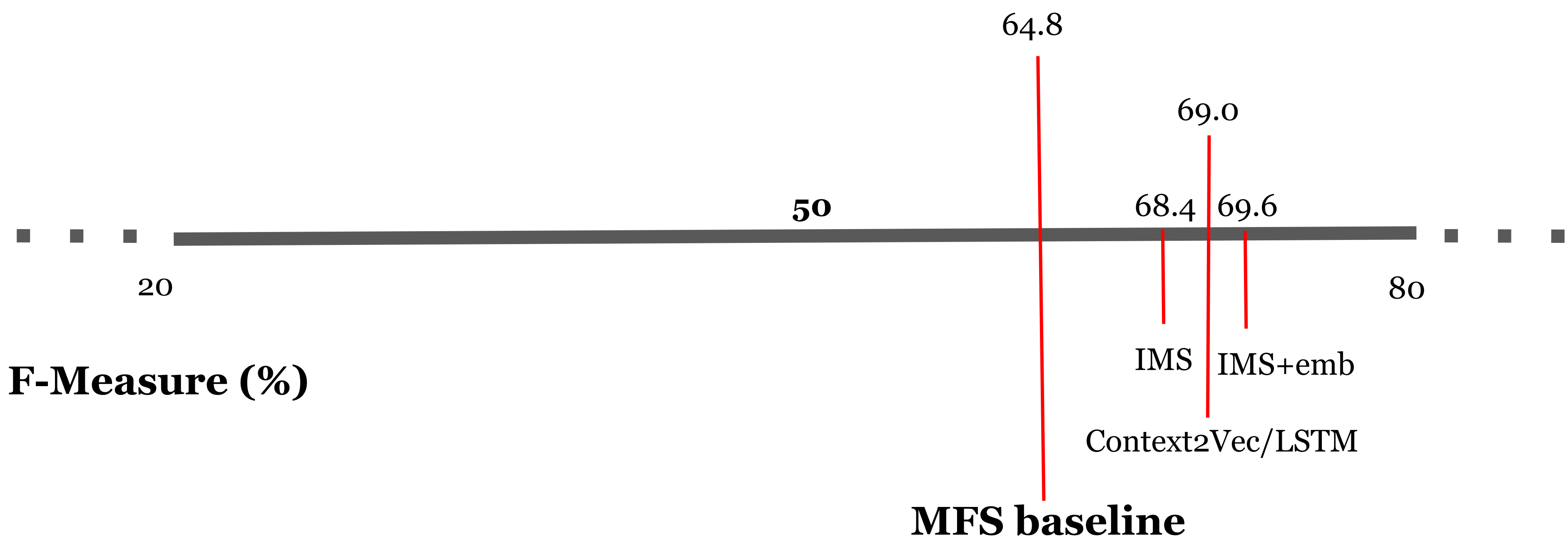
Supervised WSD systems

Neural sequence labeling (Raganato et al., EMNLP 2017)

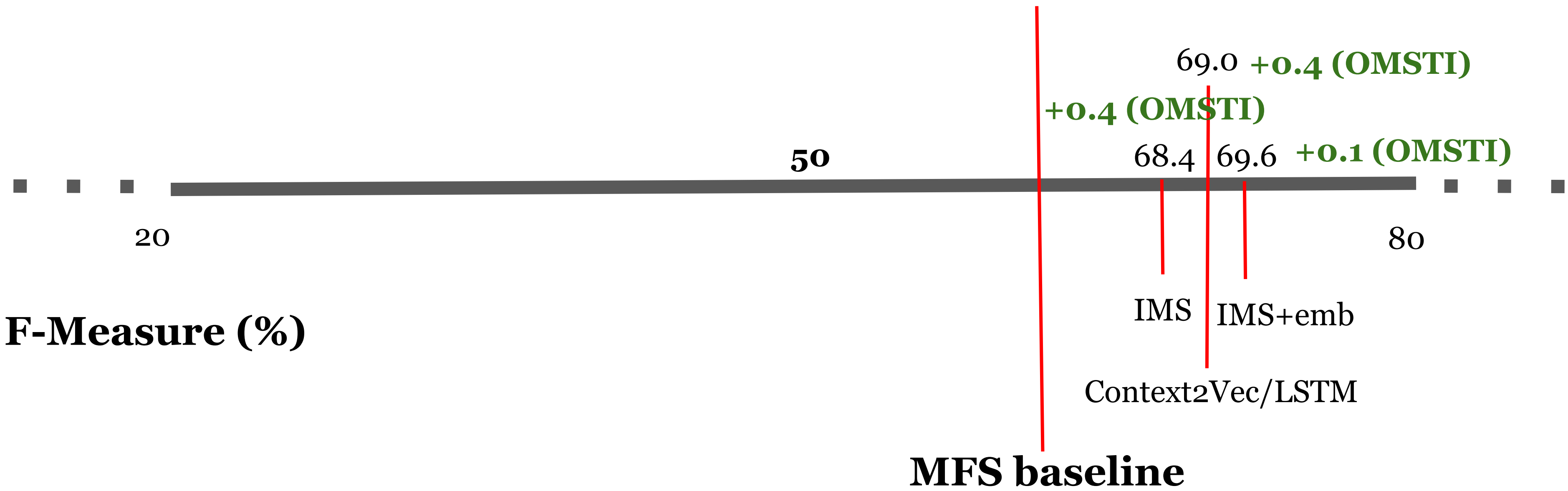


+ **Multitask learning** (PoS+supersense tagging) using **attention**

Supervised WSD systems



Supervised WSD systems



QUESTION 6
PIN: 7024700
www.kahoot.it

Knowledge-based Sense Vector Representations

Partially based on the slides of the [ACL 2016 Tutorial on Semantic Representation of Word Senses and Concepts]

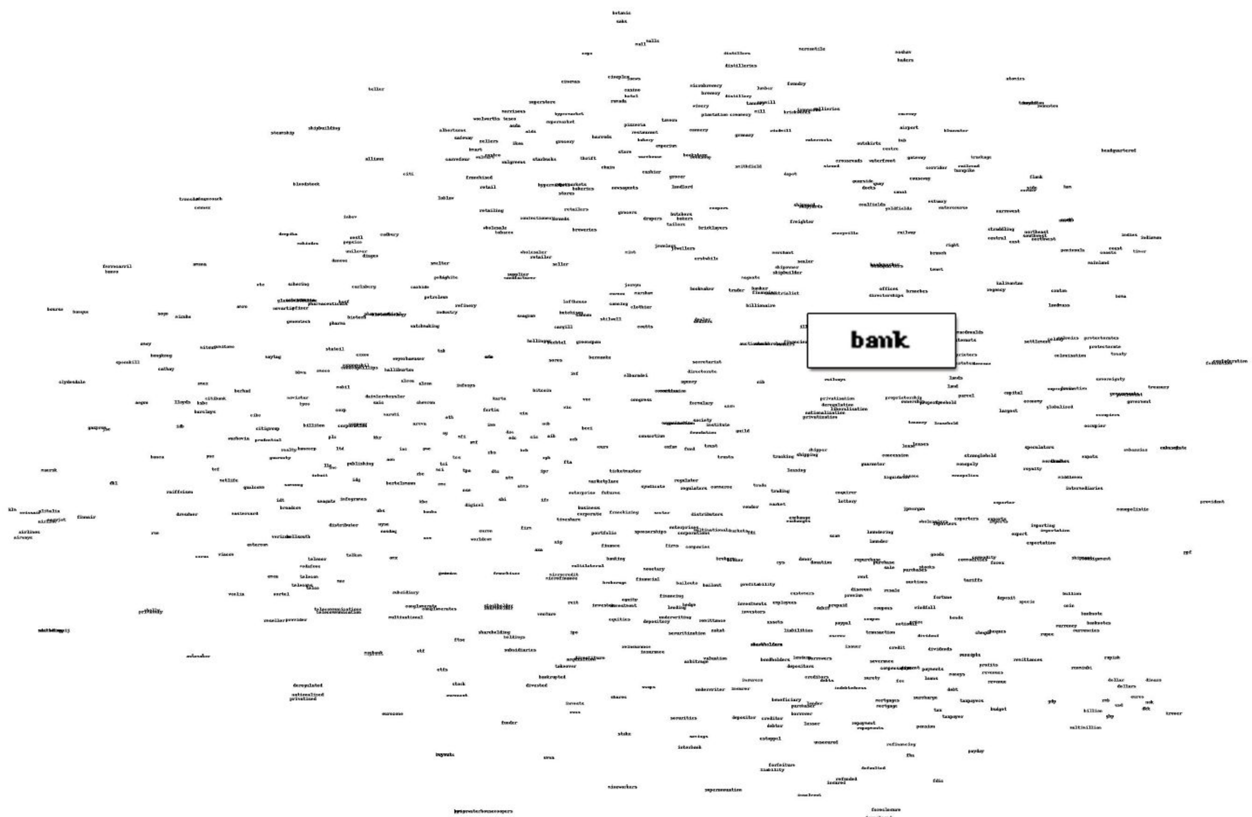
What are knowledge-based sense representations?

Semantic representations of lexical items (e.g. concepts, senses) which are linked to an external sense inventory or **lexical resource**.

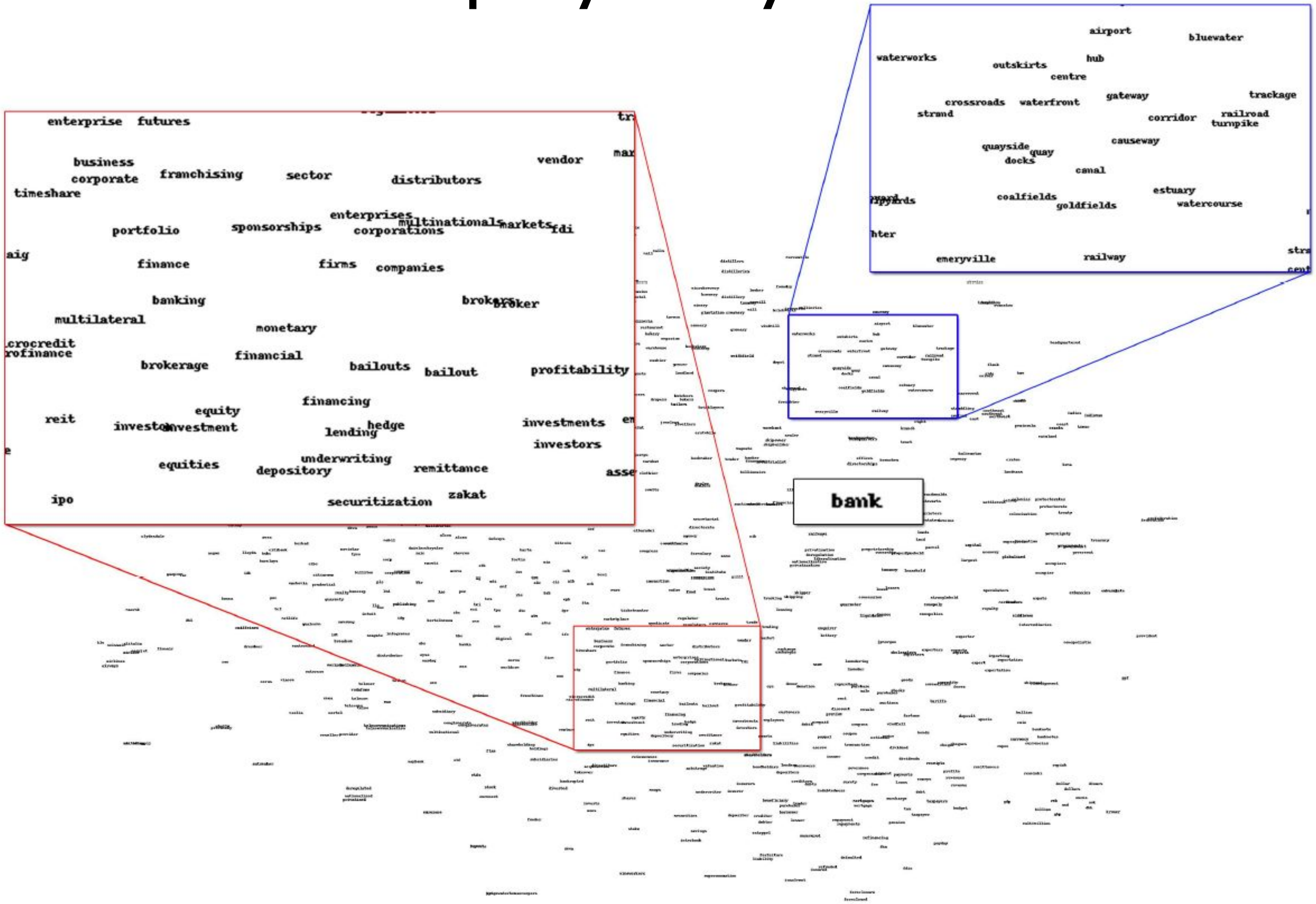
Why knowledge-based representations?

Word embeddings have shown powerful tools integrating useful semantic information, but they have some **limitations**

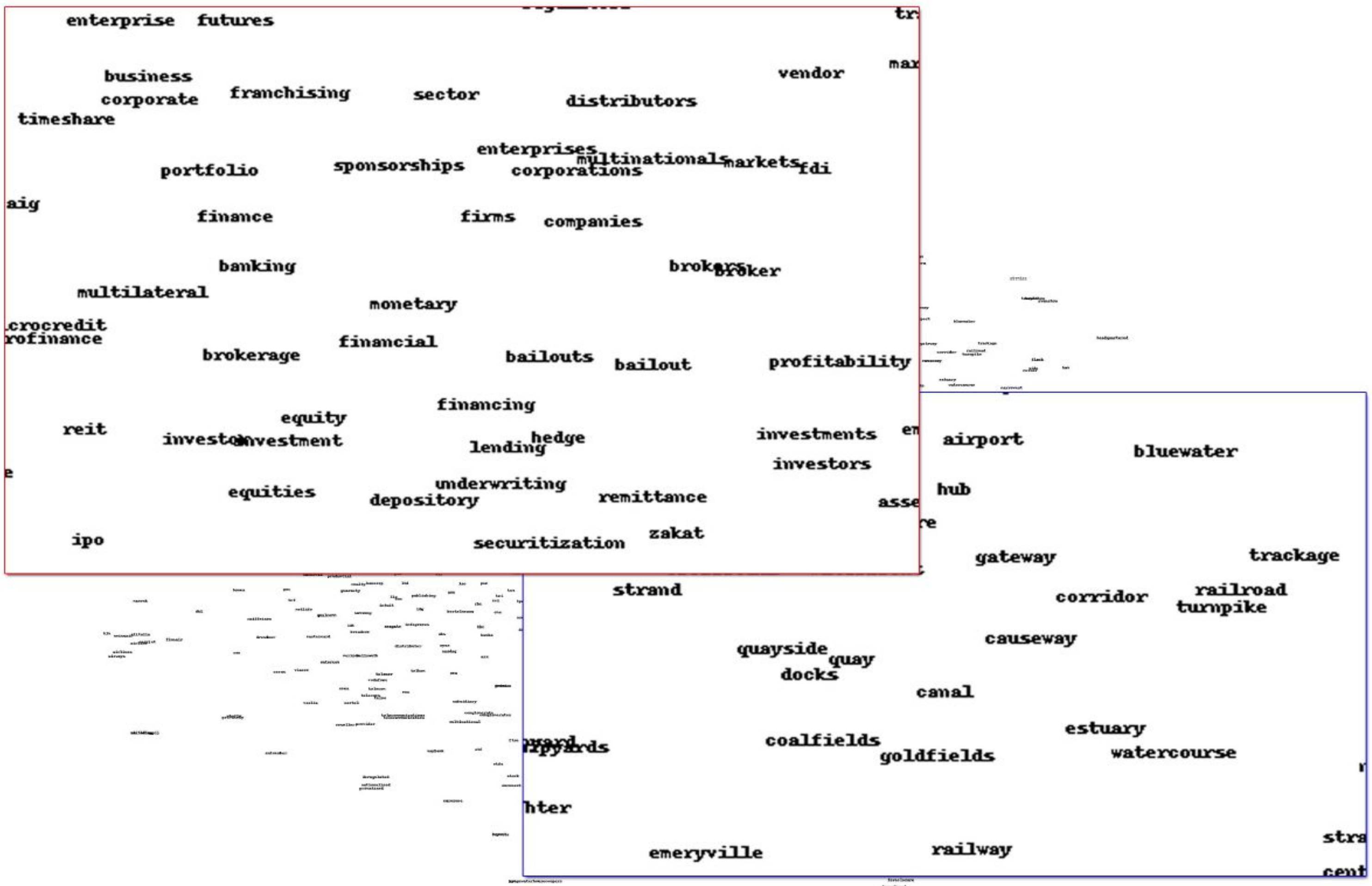
Problem 1: word representations cannot capture polysemy



Problem 1: word representations cannot capture polysemy



Problem 1: word representations cannot capture polysemy



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 $\longleftrightarrow$ refinery

plant


Word representations and the triangular inequality

Example from Neelakantan et al (2014)

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

pollen ⟷ refinery

plant^1 ⟷ plant^2

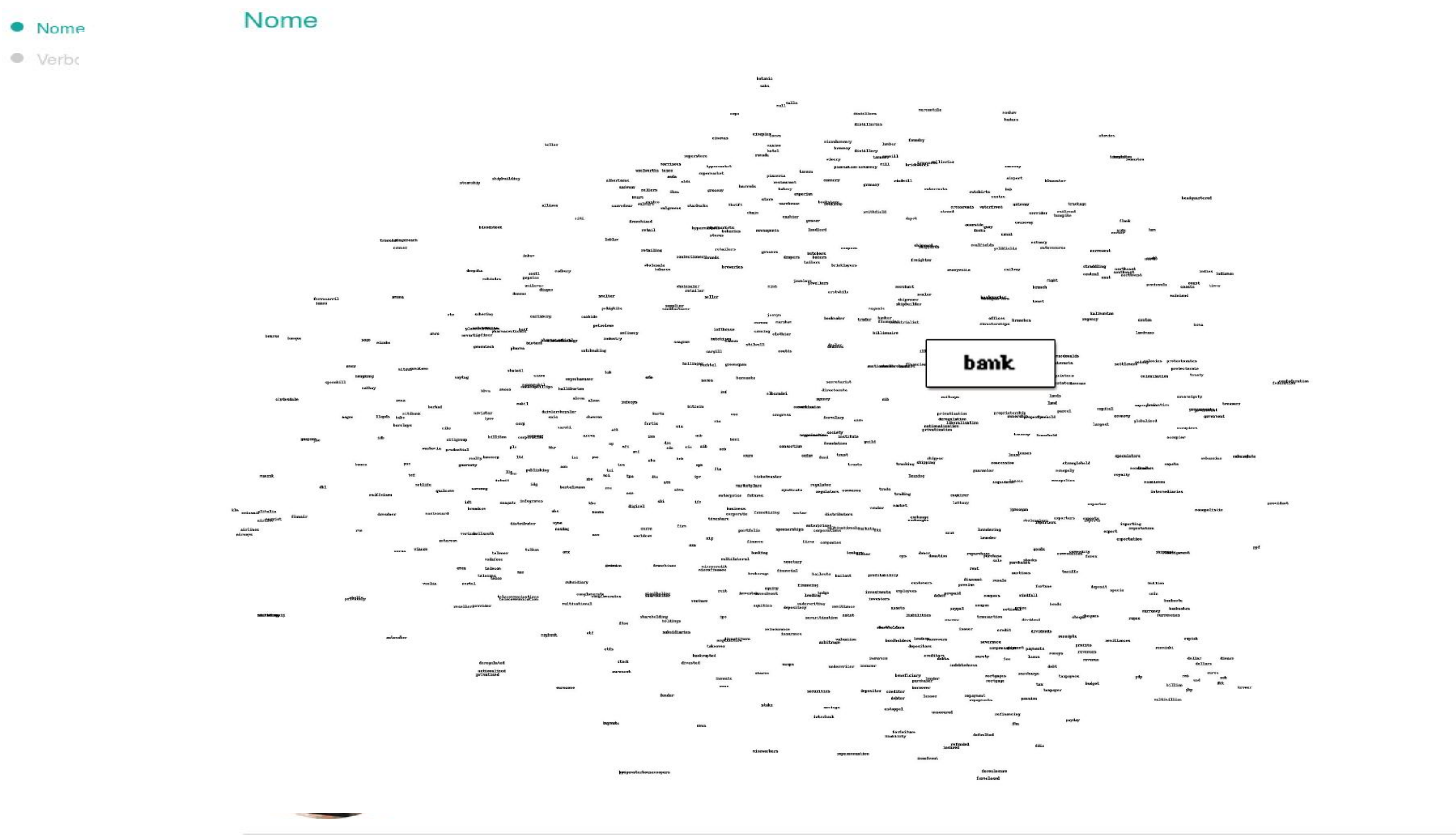
Problem 2: word representations do not take advantage of existing lexical resources

WIKIDATA

BabelNet

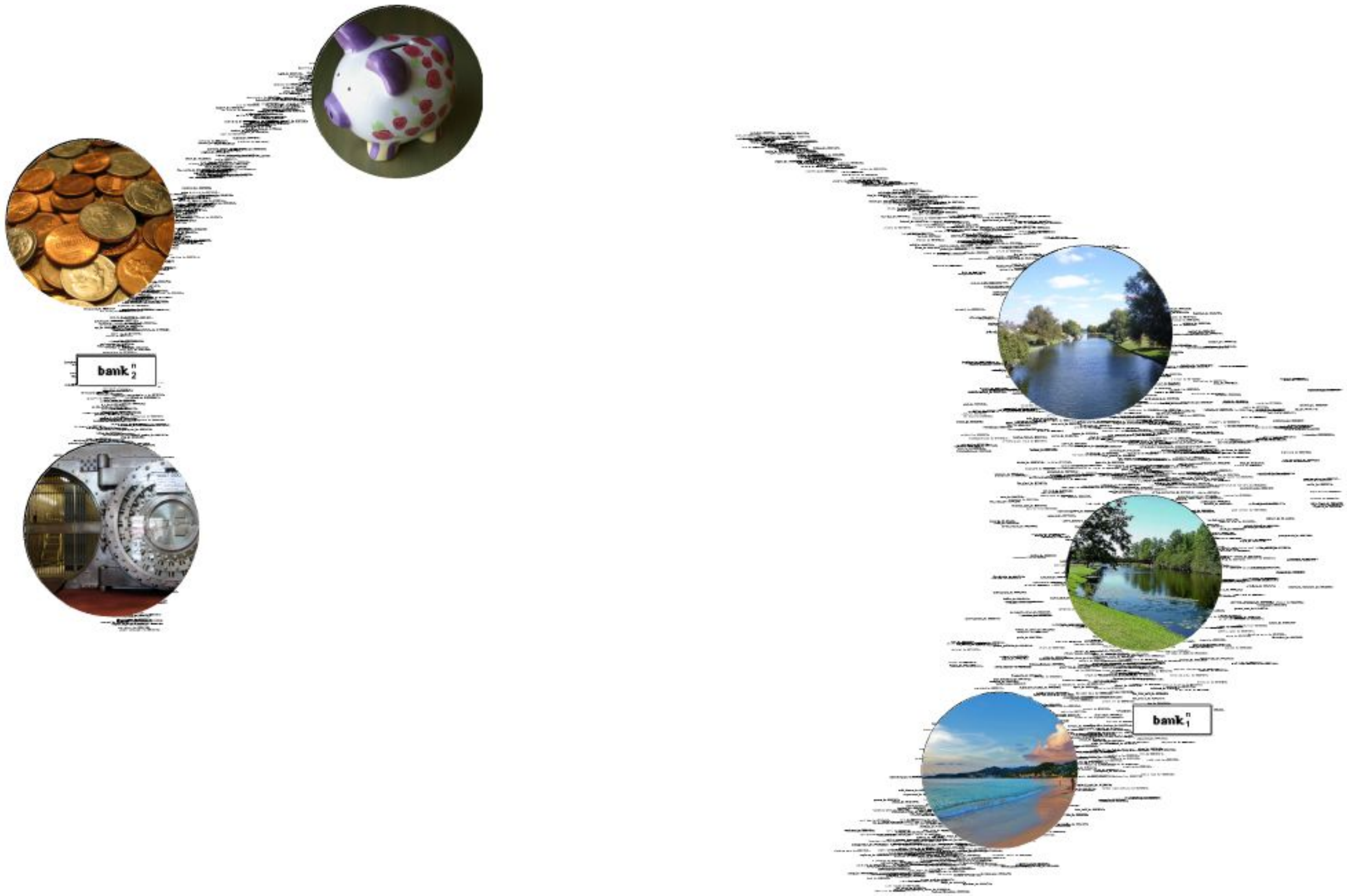
WIKIPEDIA

Key goal: obtain sense representations



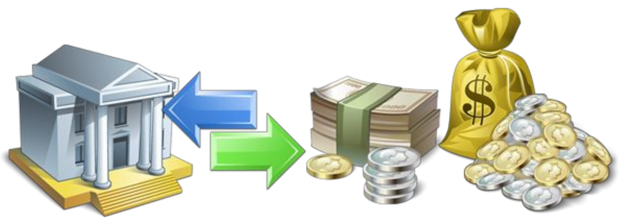
on

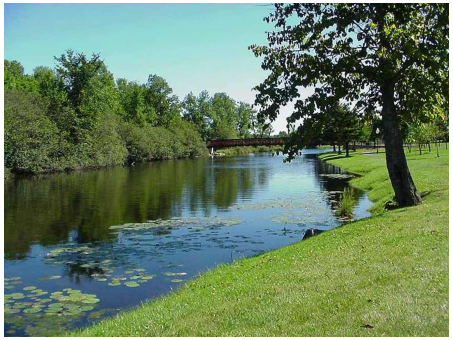
Key goal: obtain sense representations



Motivation: Model senses instead of only words

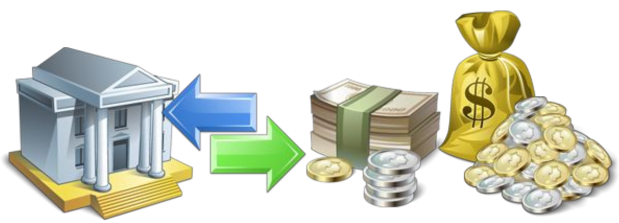
*He withdrew money from the **bank**.*





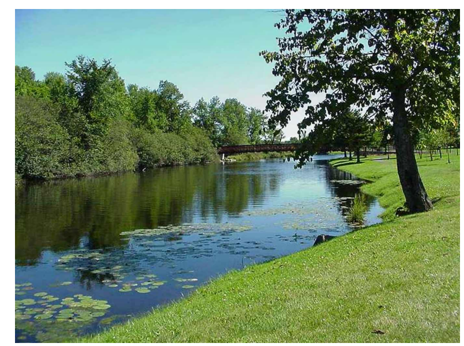
Motivation: Model senses instead of only words

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



bank#1

...

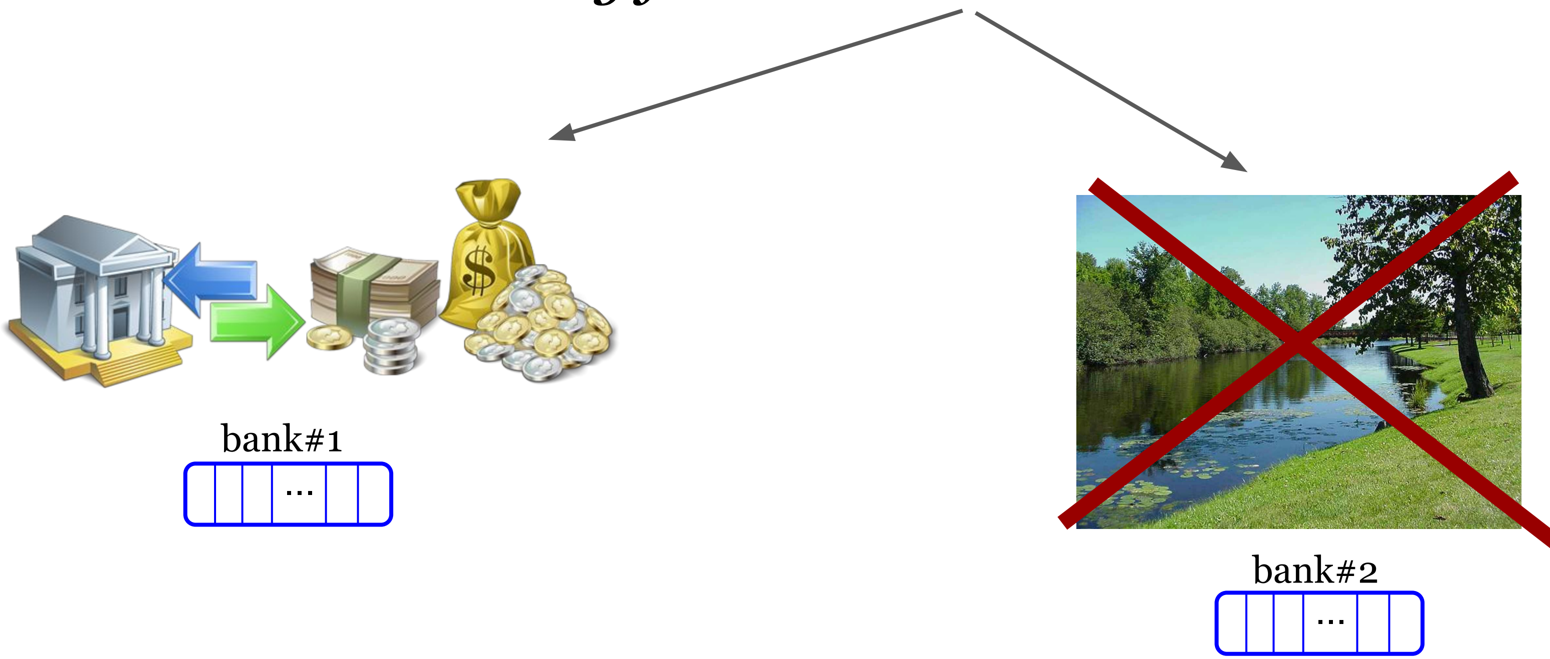


bank#2

...

Motivation: Model senses instead of only words

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



Sense representations: Two branches

- **Unsupervised sense embeddings**

- **Knowledge-based sense embeddings**

Sense representations: Two branches

- **Unsupervised sense embeddings**

Learn sense embeddings exploiting **text corpora only** *(Huang et al. ACL 2012; Neelakantan et al. EMNLP 2014; Tian et al. COLING 2014; Li and Jurafsky, EMNLP 2015...)*. **Easily adaptable to new domains.**

Drawbacks:

- Senses not interpretable (+change from model to model)
- Knowledge from resources cannot be easily exploited
- Senses (esp. not frequent ones) not easy to discriminate

Knowledge-based sense embeddings

Sense representations: Two branches

Unsupervised sense embeddings

- **Knowledge-based sense embeddings** (in this tutorial)

  Model senses as defined on a sense inventory or lexical resource.

  Usually leveraging corpus-based cues as well.

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)

Chen et al (EMNLP 2014)

A Unified Model for Word Sense Representation and Disambiguation

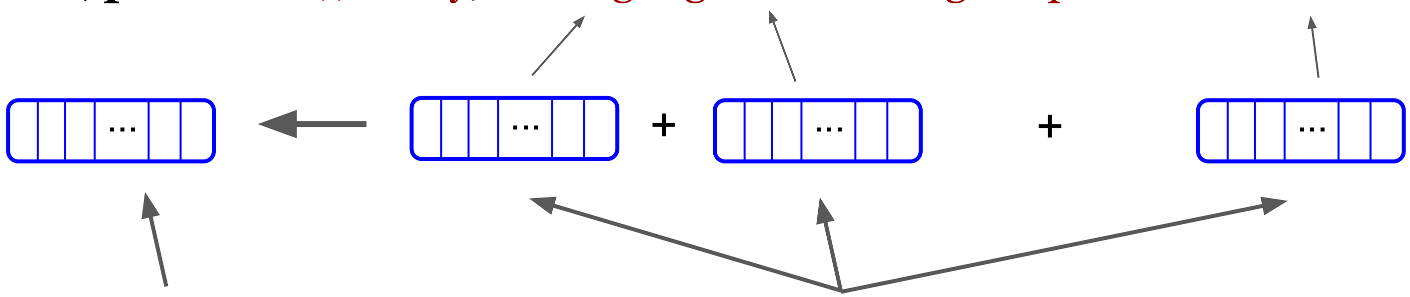
Basic idea: word sense representation and Word Sense Disambiguation can benefit from each other

⇨ Joint word sense representation and disambiguation

Chen et al (EMNLP 2014)

1- Use a sense definition to initialize its representation

plant, **flora**, **plant life** ((botany) a living organism lacking the power of locomotion)



Sense representation word embeddings

Chen et al (EMNLP 2014)

1- Use a sense definition to initialize its representation

2- Automatically disambiguate large amounts of text

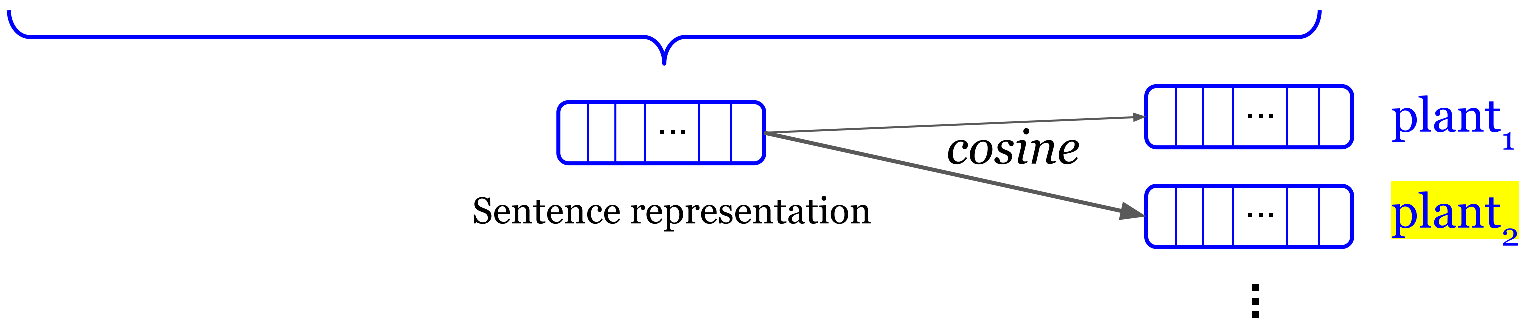
They **proposed simple disambiguation techniques** based on the obtained initial sense representations and used these disambiguation techniques to **disambiguate large amounts of texts**

Chen et al (EMNLP 2014)

Disambiguation Technique

To disambiguate a content word (plant):

water is absorbed by roots of a plant from the soil



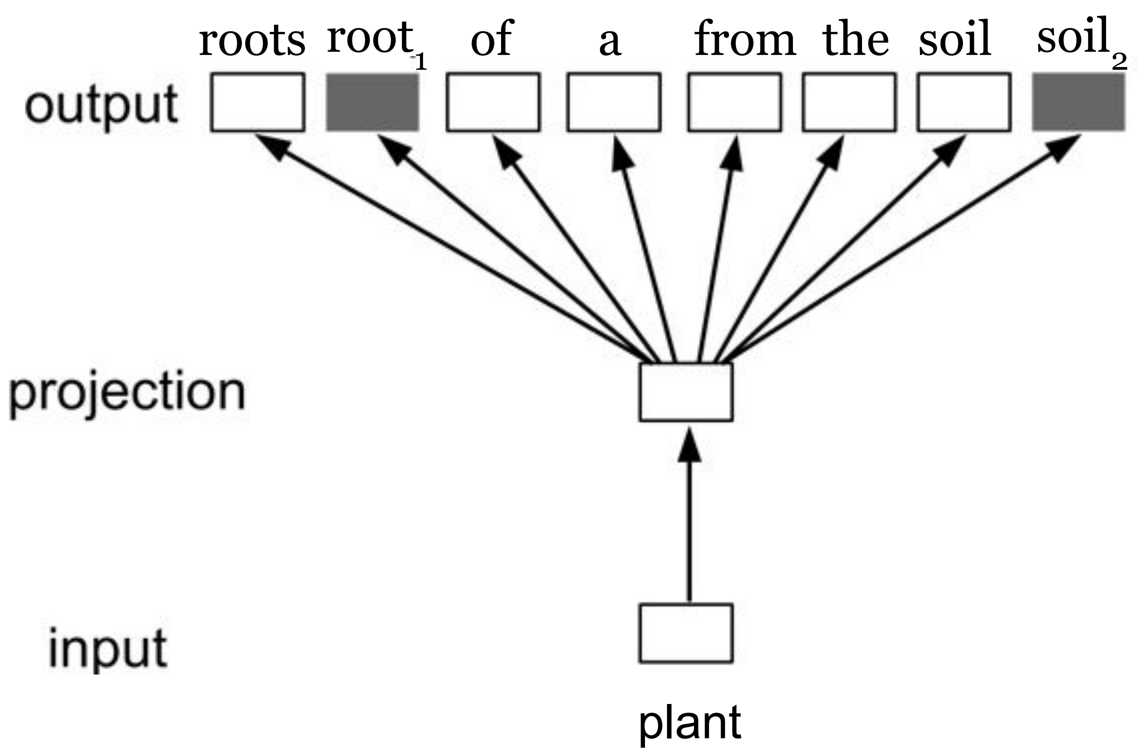
- Obtain the sentence representation (by averaging word embeddings)
- Pick the sense of *plant* which has the highest cosine similarity to the sentence vector

Chen et al (EMNLP 2014)

1- Use a sense definition to initialize its representation

2- Automatically disambiguate large amounts of text

3- Modify the objective of Skip-gram to learn sense representations



Chen et al (EMNLP 2014)

Results on the SCWS dataset:

word embeddings

sense embeddings

Model	$\rho \times 100$
Our Model-S	64.2
Our Model-M	**68.9**

Sense representations usually improve over word representations on word similarity benchmarks

Chen et al (EMNLP 2014)

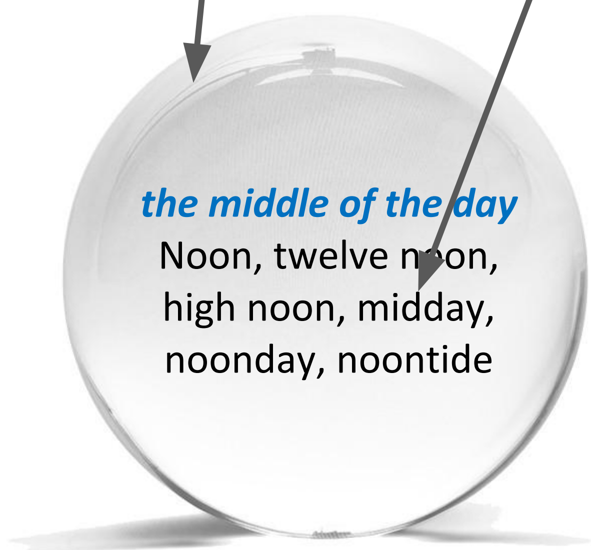
Limitations:

- Content words in definitions are not always enough for accurately pinpointing the semantics of a word sense

- The disambiguation technique is far from optimal, which introduces noise to the representation procedure

Rothe and Schütze (ACL 2015)

AutoExtend: Extending Word Embeddings to Embeddings for Synsets and Senses



Rothe and Schütze (2015)

AutoExtend: Extending Word Embeddings to Embeddings for Synsets and Senses

Leverages WordNet properties (**constraints**) for learning sense representations

Rothe and Schütze (2015)

AutoExtend: Extending Word Embeddings to Embeddings for Synsets and Senses

Leverages WordNet properties (**constraints**) for learning sense representations

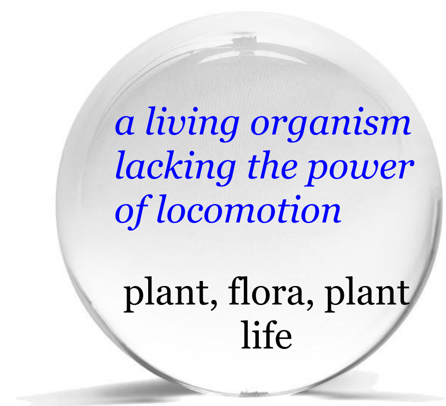
polysemy and synonymy

Rothe and Schütze (2015)

Two basic premises for an **autoencoder**:

1- **A word is the sum of its senses**

e.g., embedding of plant is the sum of embeddings of plant(organism), plant(industry), etc.



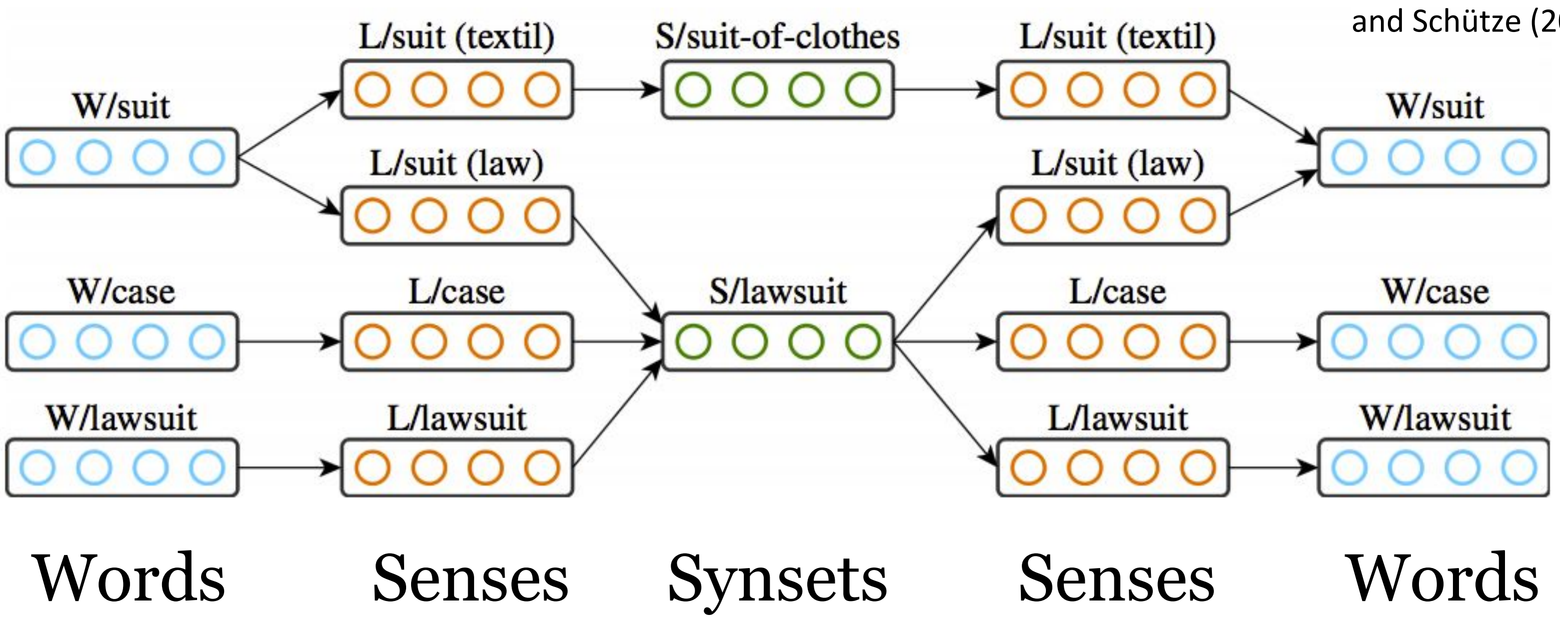
2- **A synset is the sum of its senses**

e.g., embedding of this synset is:
plant (organism) + flora (organism) + plant_life (organism)

Rothe and Schütze (2015)

An autoencoder framework

Illustration from Rothe and Schütze (2015)



Johansson and Nieto Piña (2015)

Embedding a Semantic Network in a Word Space
(NAACL 2015)

Learns sense embeddings in the same semantic space as (pre-trained) word embeddings

Applied to Swedish data:

SALDO semantic network

Johansson and Nieto Piña (2015)

target and neighbour sense representations

$$\underset{E,p}{\text{minimize}} \quad \sum_{i,j,k} w_{ijk} \Delta(E(s_{ij}), E(n_{ijk}))$$

$$\text{subject to} \quad \sum_{j} p_{ij} E(s_{ij}) = F(l_i) \quad \forall i$$

word representation

The distances between neighbours to be minimized, while satisfying the <u>mix constraint</u> for each lemma

a word vector is a convex combination of its senses vectors

Johansson and Nieto Piña (2015)

Evaluation on

classifying frames in FrameNet

Frame	P	R	F
ANIMALS	0.741	0.643	0.689
FOOD	0.684	0.679	0.682
PEOPLE_BY_VOCATION	0.595	0.651	0.622
ORIGIN	0.789	0.691	0.737
PEOPLE_BY_ORIGIN	0.693	0.481	0.568
Overall	0.569	0.292	0.386

(a) Using lemma embeddings.

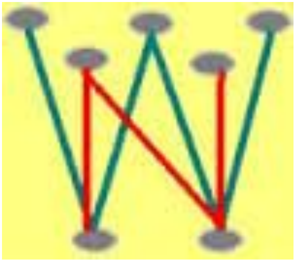
Frame	P	R	F
ANIMALS	0.826	0.663	0.736
FOOD	0.726	0.743	0.735
PEOPLE_BY_VOCATION	0.605	0.637	0.621
ORIGIN	0.813	0.684	0.742
PEOPLE_BY_ORIGIN	0.756	0.508	0.608
Overall	0.667	0.332	0.443

(b) Using sense embeddings.

Retrofitting (Faruqui et al., NAACL 2015)

Retrofitting Word Vectors to Semantic Lexicons. Manaal Faruqui, Jesse Dodge, Sujay K. Jauhar, Chris Dyer, Eduard Hovy, and Noah A. Smith (NAACL 2015)

+



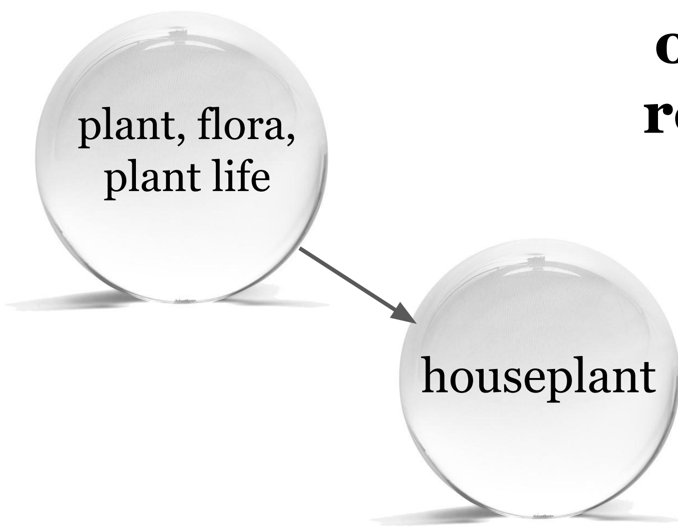
Distributional approaches usually rely **only** on the **statistics** derived from text corpora They usually **ignore** all the valuable information encoded in **knowledge resources**

Retrofitting (Faruqui et al., NAACL 2015)

WordNet

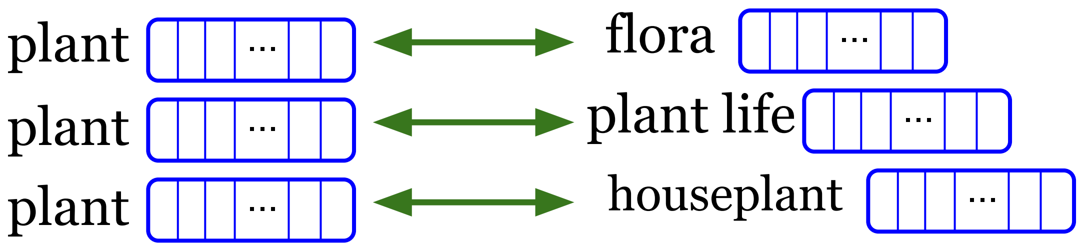
The Paraphrase Database

Benefit from synonymy and other semantic relationships in resources

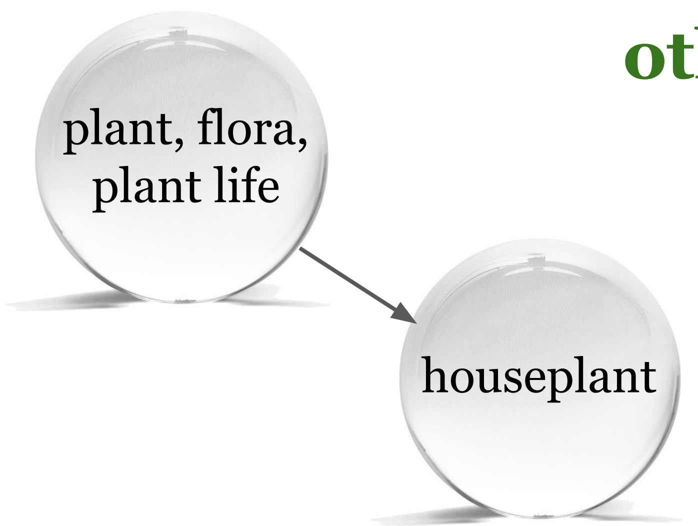


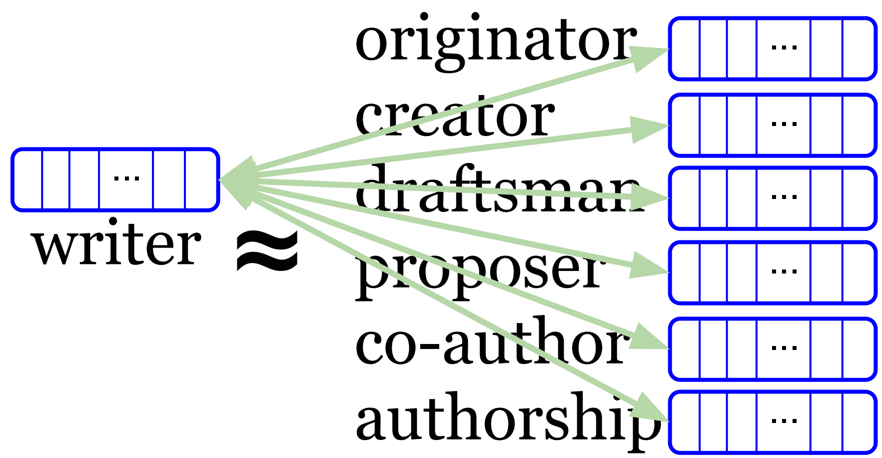
writer ≈ originator
creator
draftsman
proposer
co-author
authorship

Retrofitting (Faruqui et al., NAACL 2015)



Make these vectors more similar to each other





Jauhar et al. (NAACL 2015)

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

Two techniques for learning sense-specific embeddings that are linked to WordNet: **Retro** and **EM**

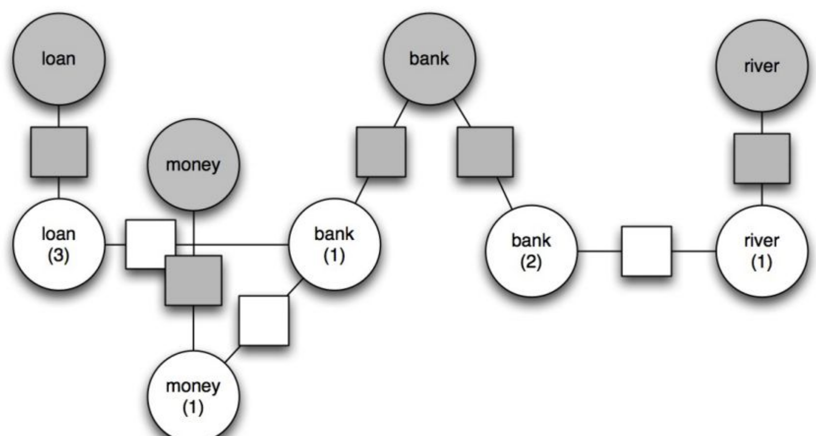
Jauhar et al. (NAACL 2015)

RETRO

$$C(V) = \arg\min_V \sum_{i-ij} \alpha \|\hat{u}_i - v_{ij}\|^2 + \sum_{ij-i'j'} \beta_r \|v_{ij} - v_{i'j'}\|^2$$

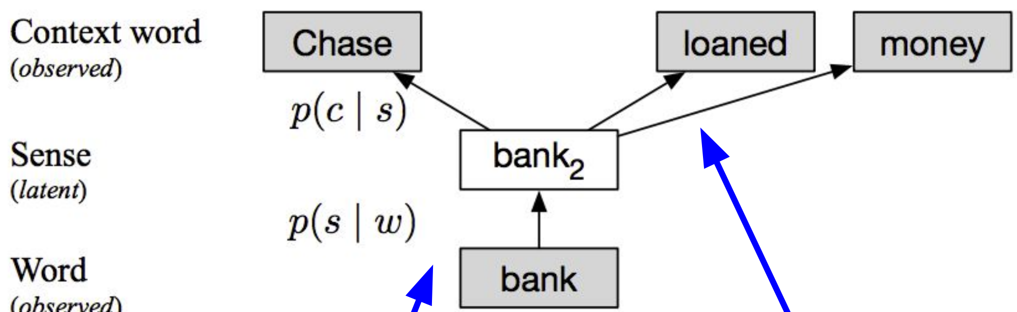
Initial word vectors

Sense vectors



Jauhar et al. (NAACL 2015)

EM: Extends the skip-gram model to learn ontologically-grounded sense vectors



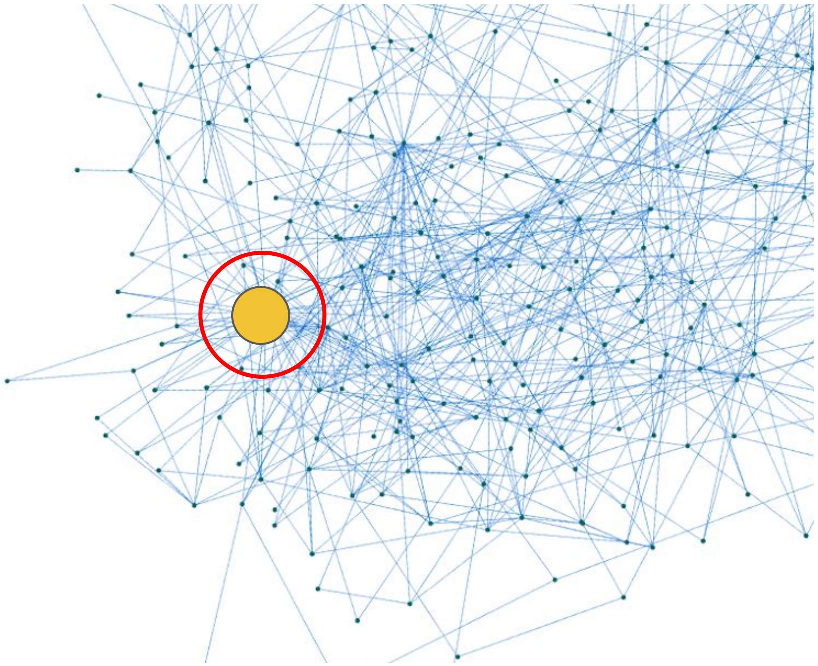
$$C(\theta) = \arg\max_{\theta} \sum_{(w_i, c_i) \in D} \log \Big(\sum_{s_{ij}} p(c_i \mid s_{ij}; \theta) \times p(s_{ij} \mid w_i; \theta) \Big) - \gamma \sum_{ij - i'j'} \beta_r \| v_{ij} - v_{i'j'} \|^2$$

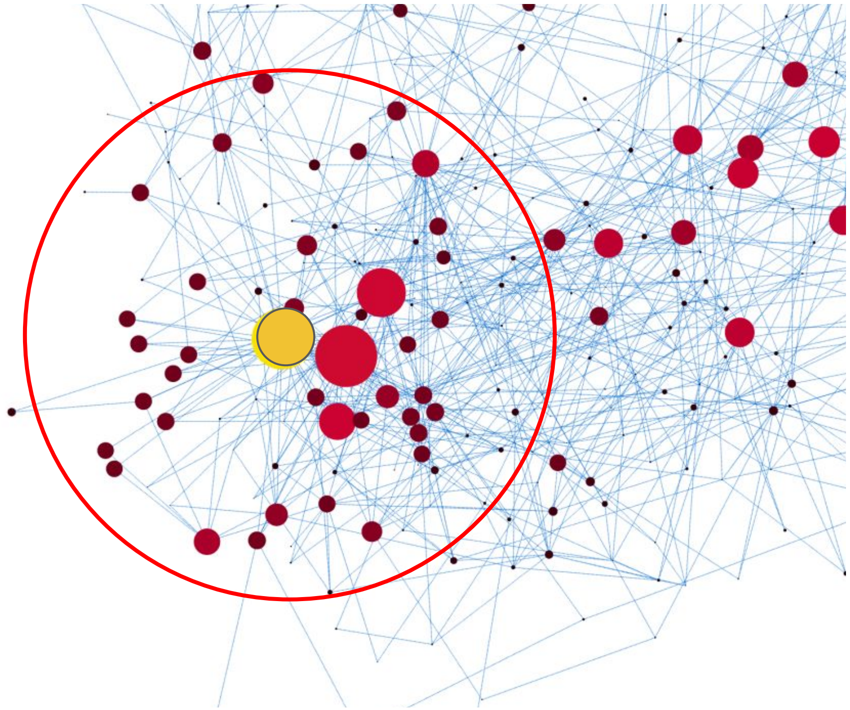
Ontological prior

De-Conflated Semantic Representations

Approaches so far

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

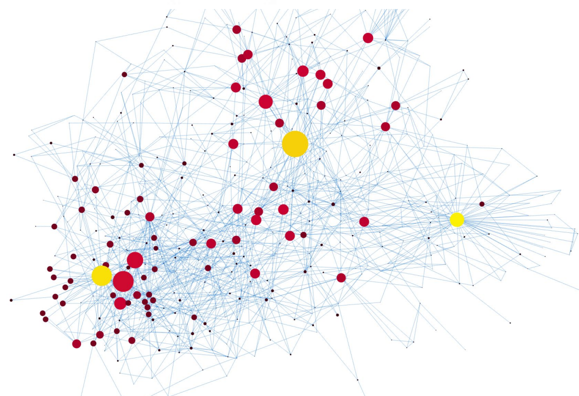




De-Conflated Semantic Representations (EMNLP 2016)

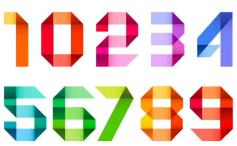
Uses Personalized PageRank algorithm to exploit WordNet for sense specific information

$$\vec{v}^{(t)} = (1-\alpha)\, M\vec{v}^{(t-1)} + \alpha\, \vec{v}^{(0)}$$



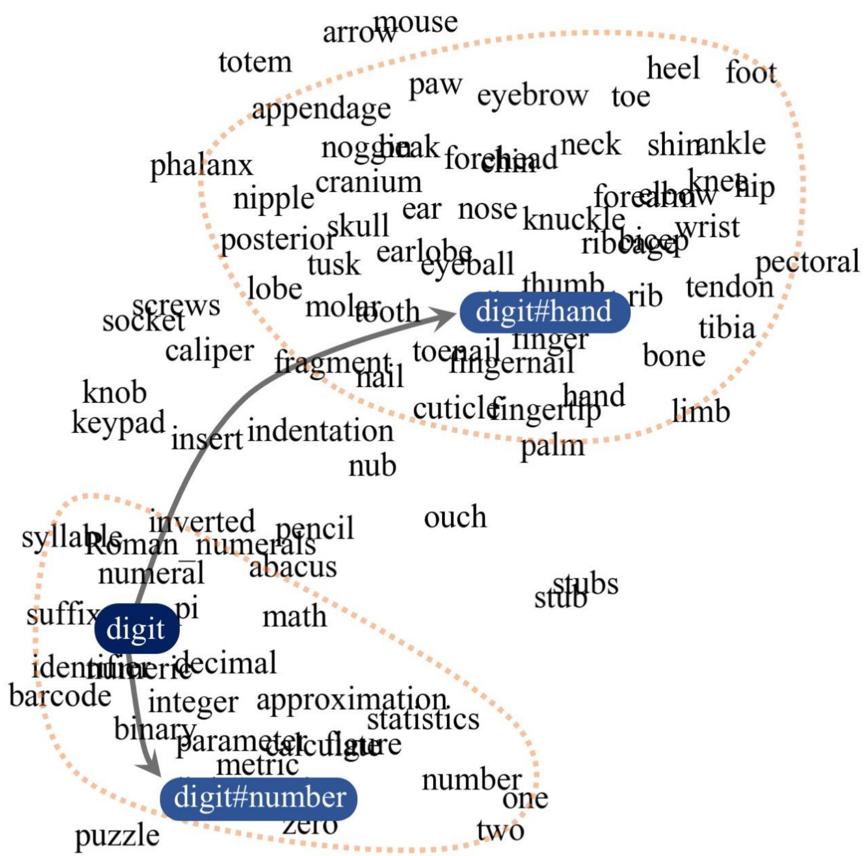
Digit

#	Sense biasing words
1	dactyl, finger, toe, thumb, pollex, body_part, nail, minimus, tarsier, webbed, extremity, appendage
2	figure, cardinal_number, cardinal, integer, whole_number, numeration_system, number_system, system_of_numeration, large_integer, constituent, element, digital



De-Conflated Semantic Representations

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

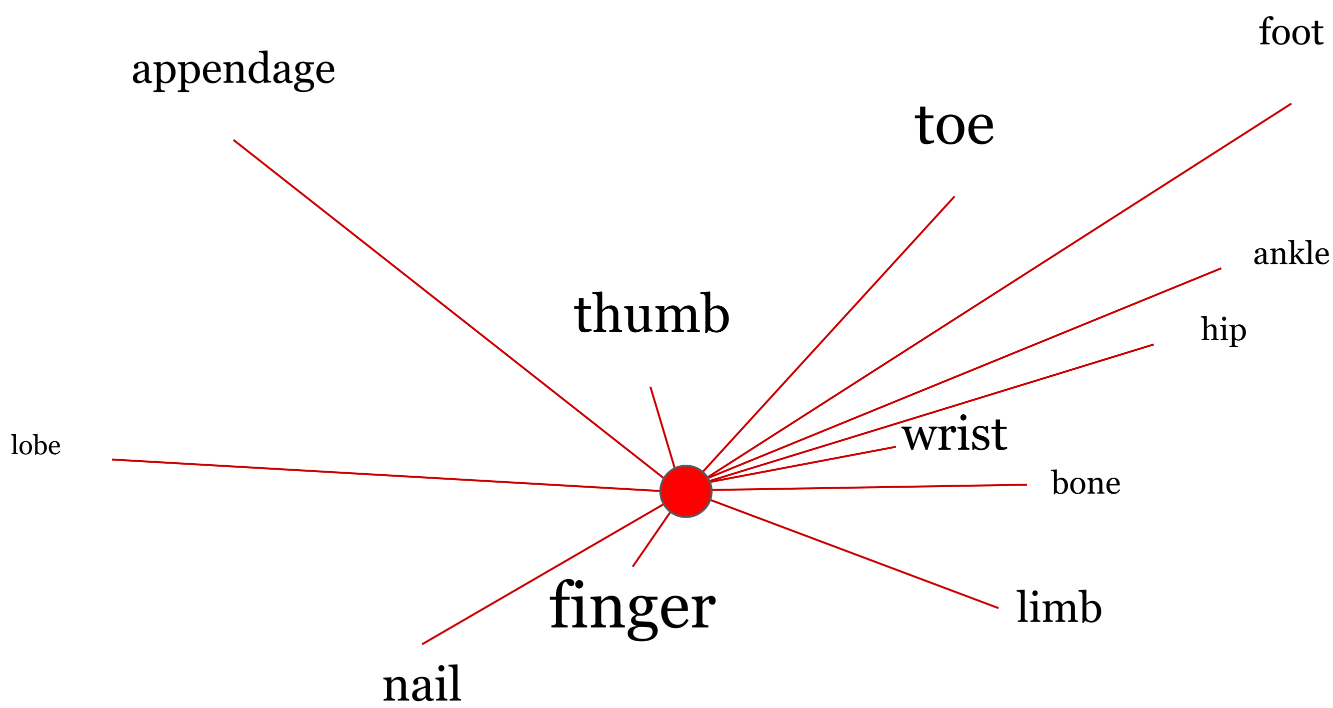


De-Conflated Semantic Representations

- Learns a representation $v^*_{s_i}$ for a sense s_i that is:
  - Close to its **lemma embedding**
  - Close to a weighted average of **embeddings of its sense biasing words**

$$\arg\min_{v^*} \alpha\, d(v^*_{s_i}, v_{s_i}) + \sum_{b_{ij} \in \mathcal{B}_i} \delta_{ij}\, d(v^*_{s_i}, v_{b_{ij}})$$

De-Conflated Semantic Representations



Knowledge Representations using WordNet

Advantages and limitations

+ Manually curated
+ Rich and highly accurate representations:
  state-of-the-art performance on multiple NLP tasks and datasets
- Limited coverage (that of WordNet)
  > Solution: use large-scale lexical resources

Large knowledge resources

Large knowledge resources

Wikipedia

BabelNet

FreeBase/Wikidata

Knowledge-based sense representations exploiting Wikipedia and BabelNet

- **NASARI** (Camacho-Collados et al., AIJ 2016)

- **SensEmbed** (Iacobacci et al., ACL 2015)

- **SW2V** (Mancini et al., CoNLL 2017)

NASARI: Integrating explicit knowledge and corpus statistics for a multilingual representation of concepts and entities (Camacho-Collados et al., AIJ 2016)

Goal

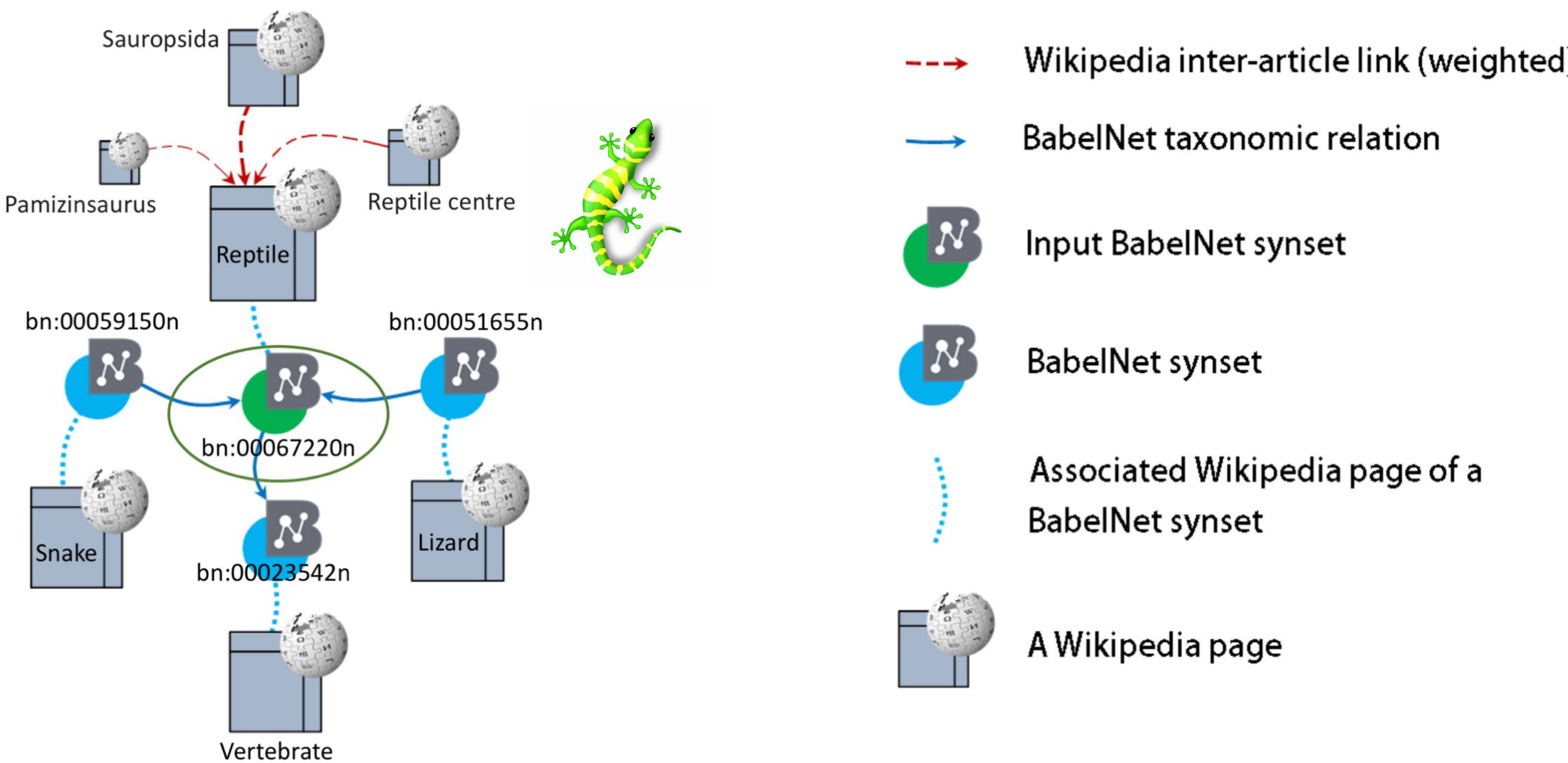
Build vector representations for multilingual BabelNet synsets.

How?

It exploits **Wikipedia semantic network** and the **WordNet taxonomy** to construct a subcorpus contextual information for any given BabelNet synset.

NASARI (AIJ 2016)

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



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

NASARI (AIJ 2016)

Three types of vector representations:

- **Lexical** (dimensions are words)

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

- **Embedded** (latent dimensions)

NASARI (AIJ 2016)

Three types of vector representations:

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

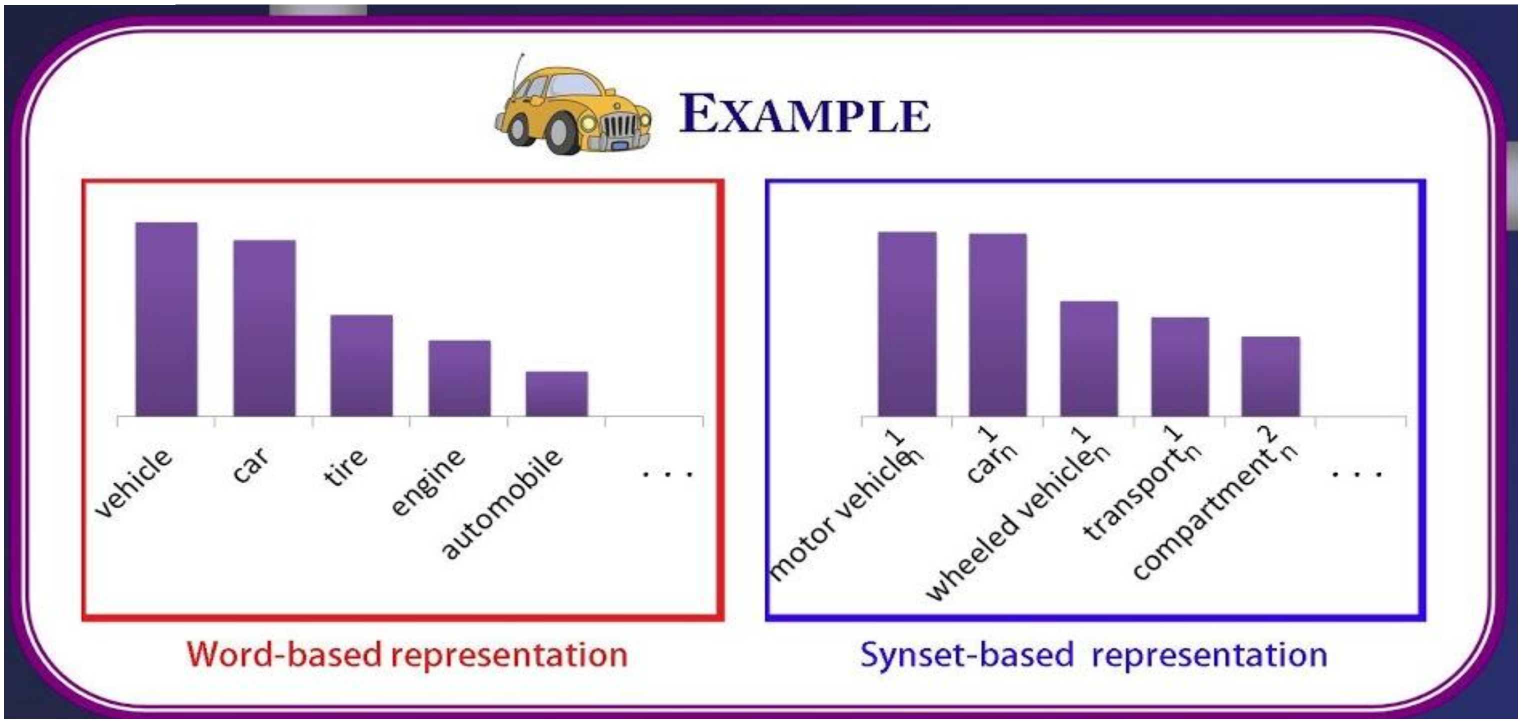
- **Unified** (dimensions are multilingual BabelNet synsets): This representation uses a **hypernym-based clustering technique** and can be used in **cross-lingual** applications
-
- **Embedded** (latent dimensions)

NASARI (AIJ 2016)

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)

NASARI (AIJ 2016)



From a lexical vector to a unified vector

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

$\textbf{motor\_vehicle}^{1}_{n}$

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

NASARI (AIJ 2016)

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.

NASARI (AIJ 2016)

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!

NASARI (AIJ 2016)

High coverage of concepts and named entities in several languages (covers all Wikipedia pages).

Useful for **multilingual applications.**

SensEmbed (Iacobacci et al., ACL 2015)

It leverages **BabelNet** and **Word2Vec** to build sense embeddings. Two steps:

- First, it uses **Babelfy** (Moro et al., TACL 2014), a multilingual joint disambiguation and entity linking system, to disambiguate a corpus.

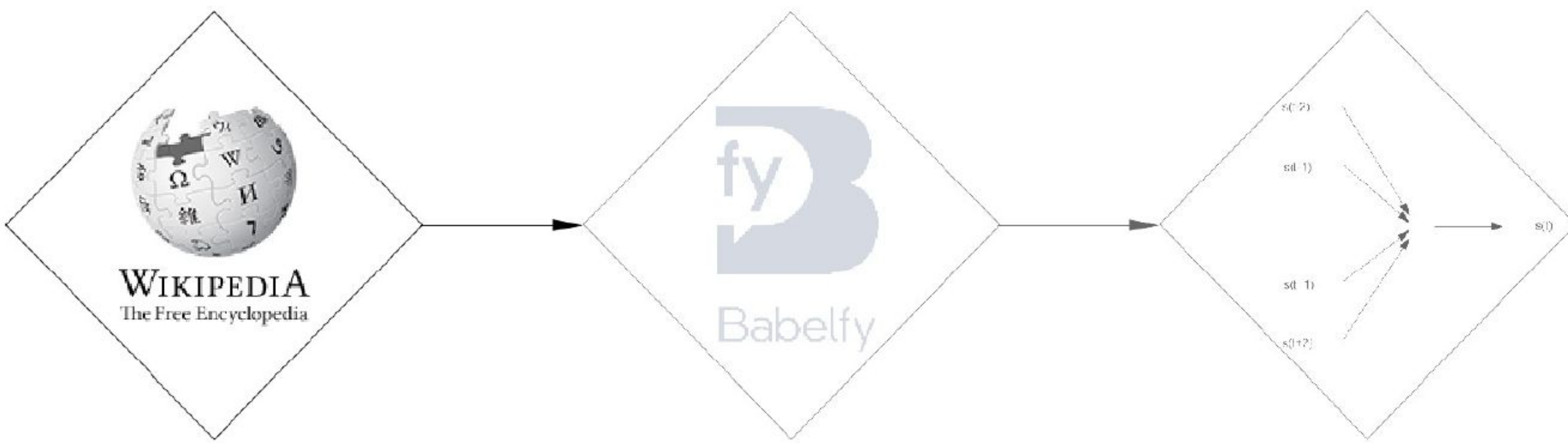
SensEmbed

It leverages **BabelNet** and **Word2Vec** to build sense embeddings. Two steps:

- First, it uses **Babelfy** (Moro et al., TACL 2014), a multilingual joint disambiguation and entity linking system, to disambiguate a corpus.
- Then, it uses **Word2Vec** to learn sense embeddings from the sense-annotated corpus.

SensEmbed

SensEmbed construction



...survey on the relationship between the **banks** and our industry , in preparation for a forthcoming forum.

...and it stands on the right **bank** of the Drava River , bounded by the river to the north...

... If you have dividend or receive **bank** or building society interest on which tax has been paid ,

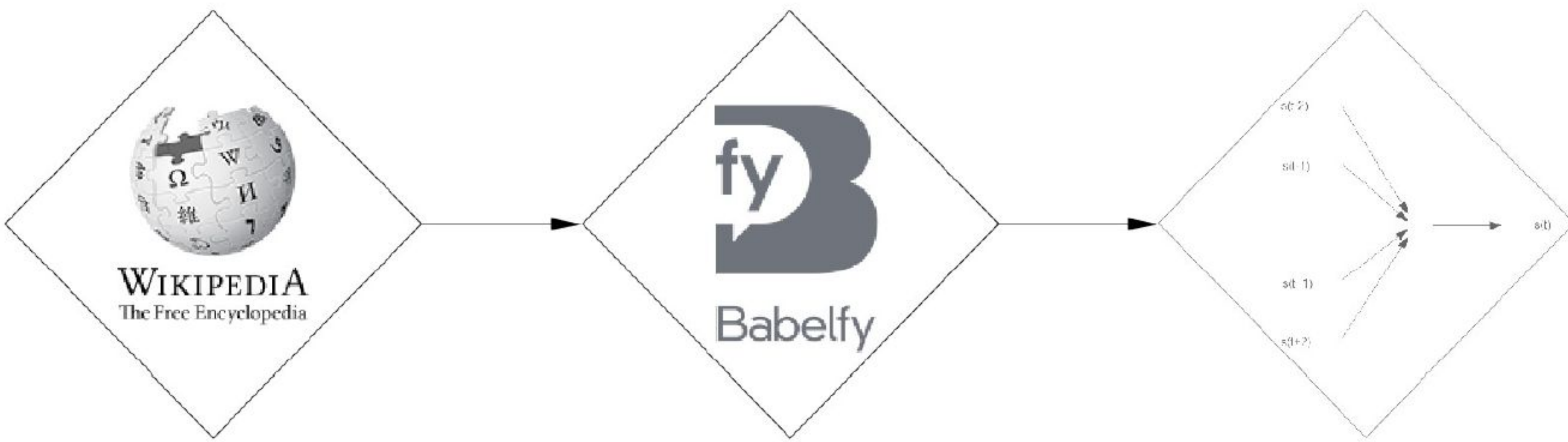
...workplaces and unions. Corporations, **banks** and trusts controlled a great deal and , although machines...

...The critical decision for the **banks** will come if their own adviser sticks to his view of the costs.

countryside of high hedges and tall earth **banks** with trees on top. The heavily wooded area was criss-crossed...

SensEmbed

SENSEMBED construction



...survey on the relationship between the **banks** and our industry , in preparation for a forthcoming forum.
...and it stands on the right **bank** of the Drava River , bounded by the river to the north...
... If you have dividend or receive **bank** or building society interest on which tax has been paid ,
...workplaces and unions. Corporations, **banks** and trusts controlled a great deal and , although machines...
...The critical decision for the **banks** will come if their own adviser sticks to his view of the costs.
countryside of high hedges and tall earth **banks** with trees on top. The heavily wooded area was criss-crossed...

SensEmbed

SENSEMBED construction

...survey on the relationship between the **banks** and our industry , in preparation for a forthcoming forum.
...and it stands on the right **bank** of the Drava River , bounded by the river to the north...
... If you have dividend or receive **bank** or building society interest on which tax has been paid ,
...workplaces and unions. Corporations, **banks** and trusts controlled a great deal and , although machines...
...The critical decision for the **banks** will come if their own adviser sticks to his view of the costs.
countryside of high hedges and tall earth **banks** with trees on top. The heavily wooded area was criss-crossed...

-2.19067 1.16642 -1.91385 -0.269672 0.712771 -0.623024 -3.20115 0.560895 0.891554 0.145258 1.26956 -0.221078 -0.0733777 2.08072 -3.30558 -0.727272 -0.902202 -1.84578 -1.38985 -0.0791954 0.989769 -1.34631 1.10242 -1.59836 -1.37341 -1.42038 0.238941 -2.98729 -0.730938 0.267584 0.0560677 -0.722721 2.23752 -2.99094 -1.45598 -0.645446 0.278277 2.28877 -0.926191 2.89934 -1.17254 1.38449 2.38617 -0.0838845 -1.80698 0.622097 0.223875 0.870654 -0.33808 -0.41957

1.16672 0.811884 -0.115492 -2.59049 -1.50286 1.2536 1.44281 0.0136615 0.131499 2.04445 -0.425782 1.29676 0.0996086 1.52687 -0.0951281 -0.715488 -0.71172 0.453871 1.08481 1.55074 0.385158 -0.116754 -0.582987 -1.56923 -0.488404 -1.07999 0.0447149 -0.733387 0.765212 2.67995 2.51105 0.192151 1.49743 2.91849 1.86901 0.23101 0.381663 1.20355 0.126758 1.57204 -0.372069 -2.45076 0.514557 -1.4028 -1.20396 0.726036 2.41265 -0.104843 2.26862 1.21729

SW2V (Mancini et al., CoNLL 2017)

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

SW2V (Mancini 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 shallowly link to each word its *associated senses in context*.

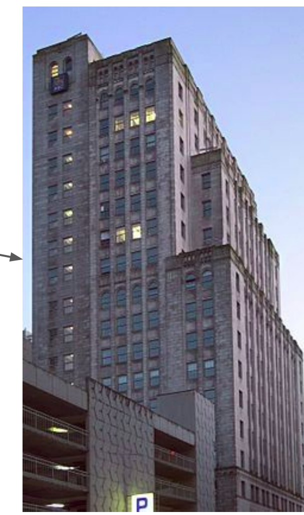
He withdrew money from the ***bank***.

SW2V: Idea

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

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

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



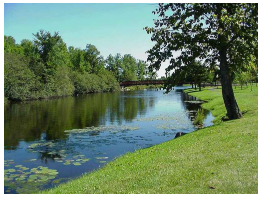
SW2V: Idea

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

retire

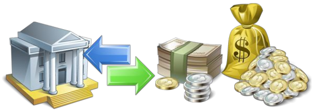
cash

geography



take out

financial institution



building

SW2V: Idea

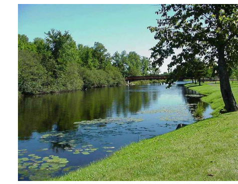
He **withdrew** **money** from the **bank**

retire

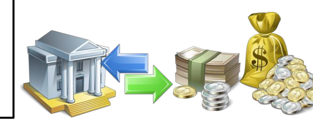
take out

cash

geography



financial institution

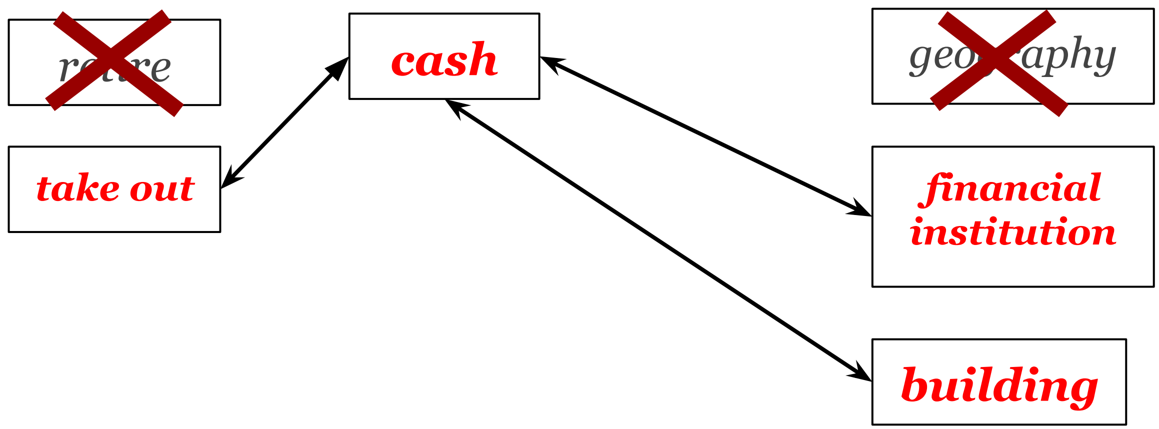


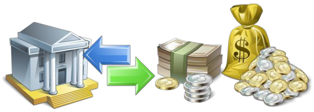
building

Graph-based representation of the sentence using semantic networks (e.g. WordNet, BabelNet)

SW2V: Idea

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





Graph-based representation of the sentence using semantic networks (e.g. WordNet, BabelNet)

SW2V: Idea

Given as input a corpus and a semantic network:

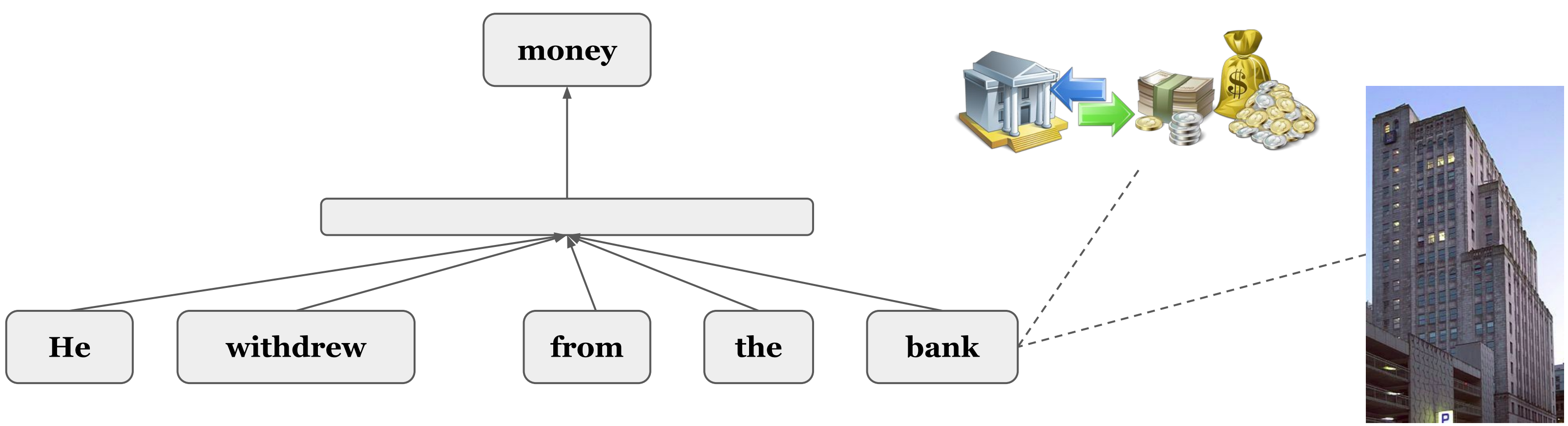
1. Use a semantic network to link to each word its *associated senses in context*.
2. Use a neural network where the update of word and sense embeddings is linked, exploiting *virtual* connections.

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.

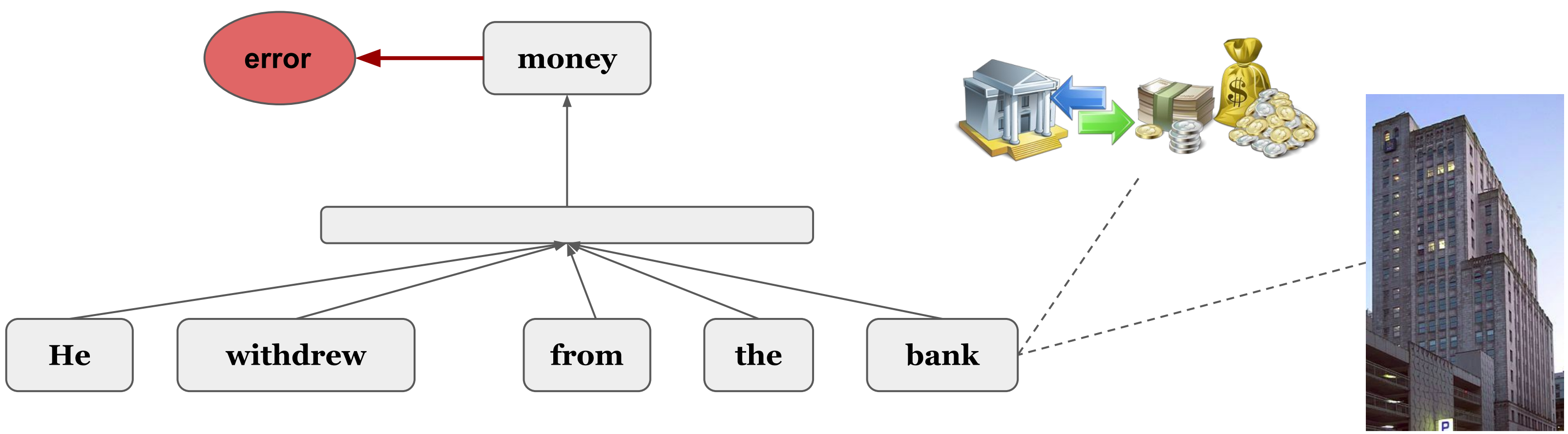


SW2V: Idea

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.

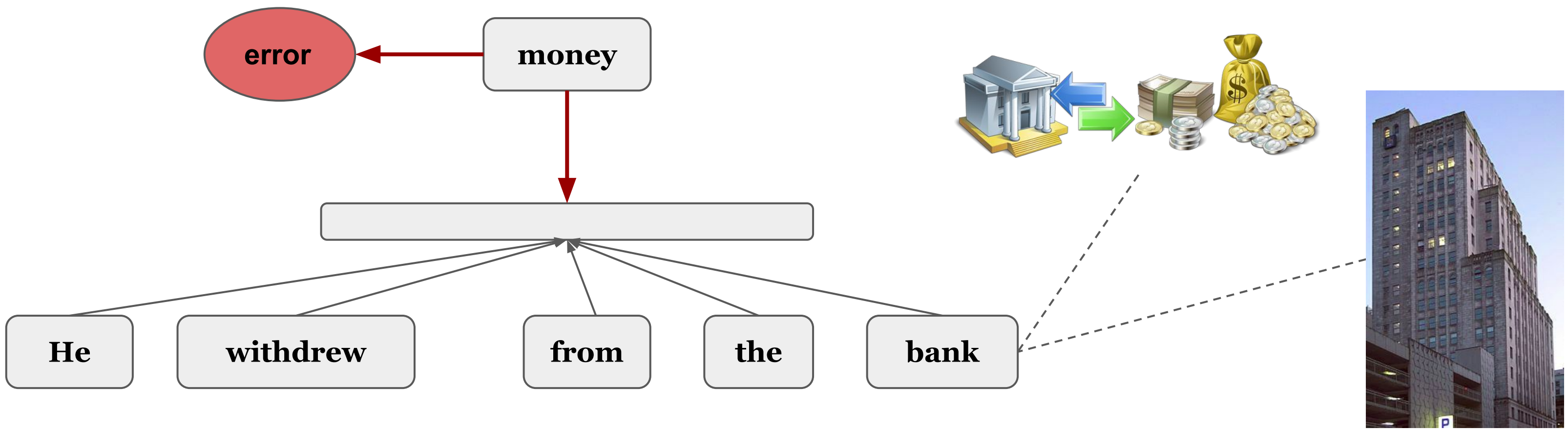


SW2V: Idea

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.

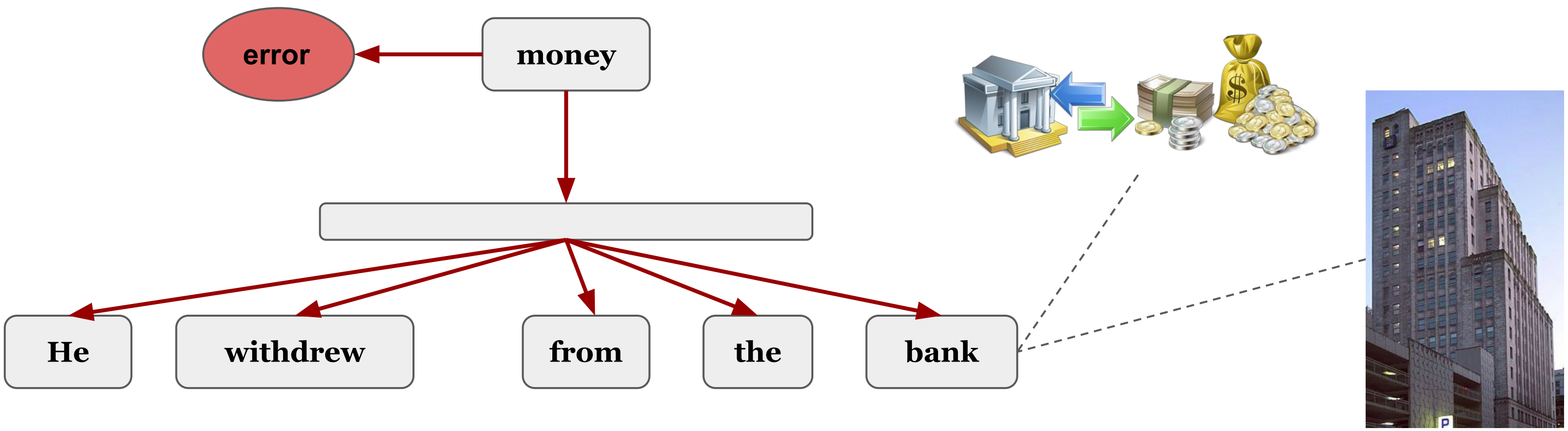


SW2V: Idea

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.

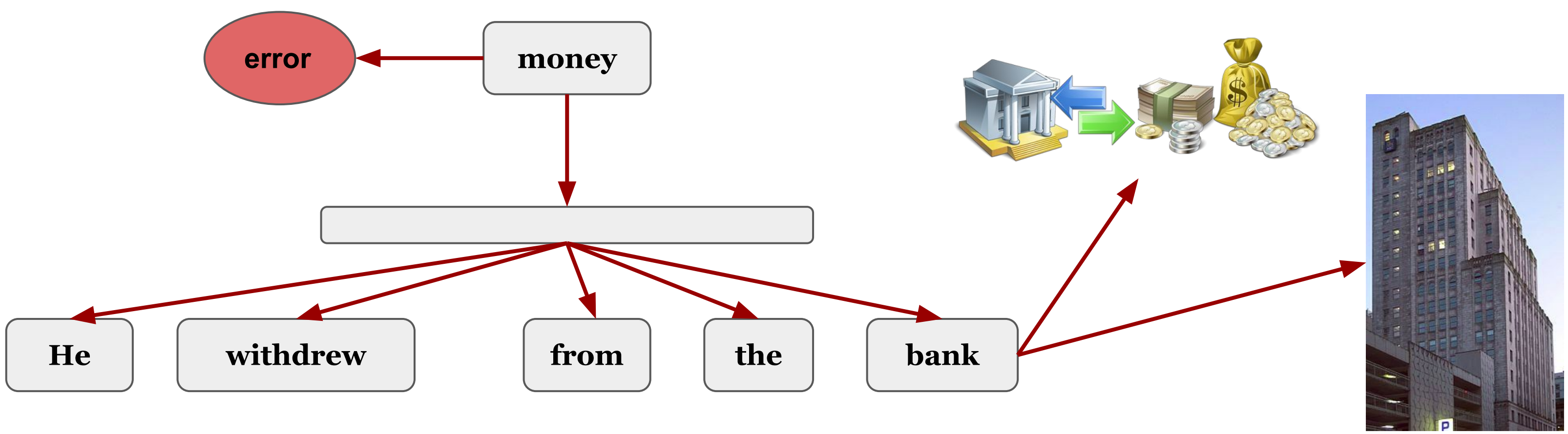


SW2V: Idea

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:

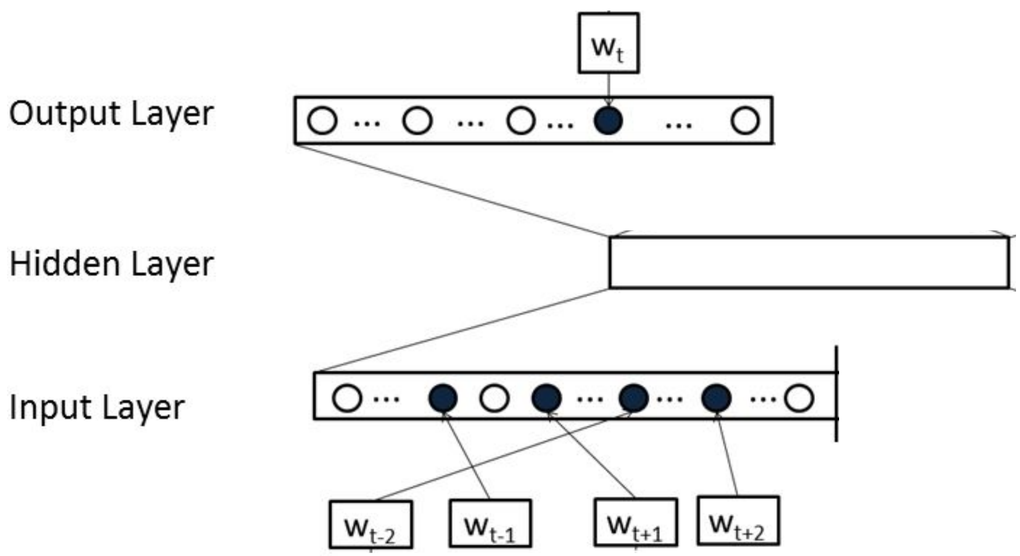
1. Use a semantic network to link to each word its *associated senses in context*.

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)

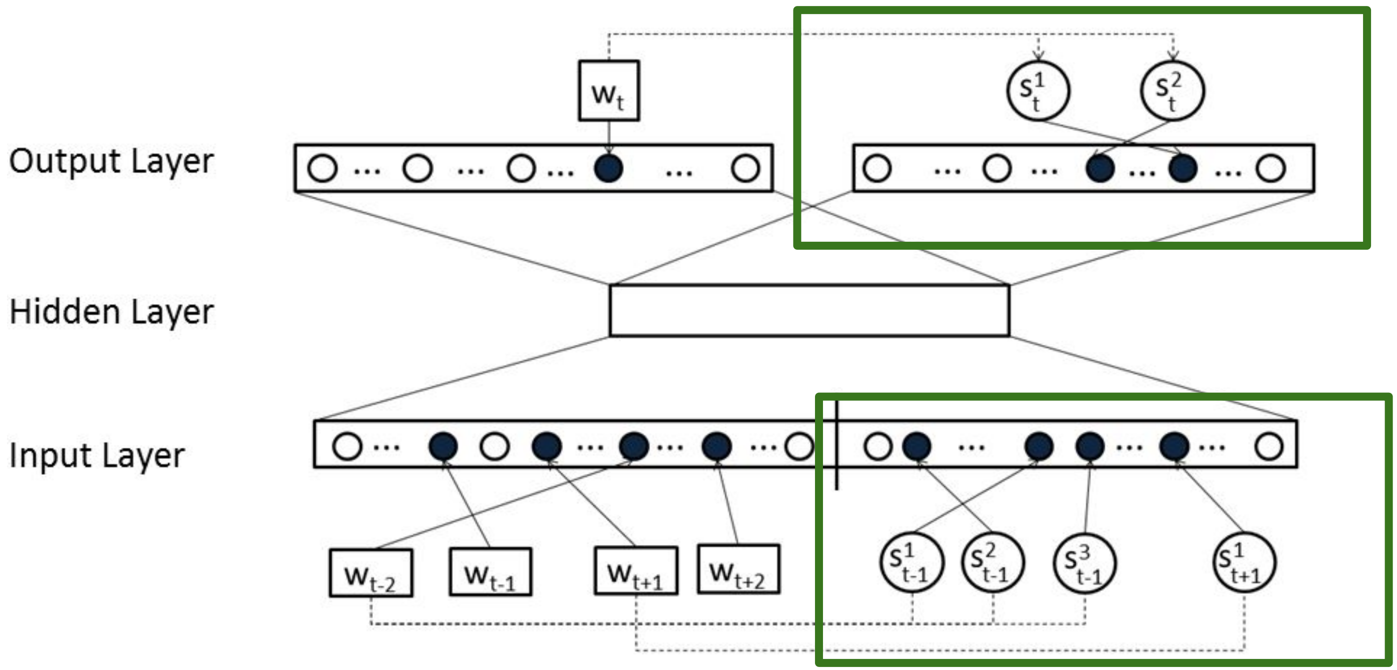
$$E=-\log(p(w_t|W^t))$$



Words and associated senses used both as input and output.

Full architecture of SW2V (Mancini et al. 2017)

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



Words and associated senses used both as input and output.

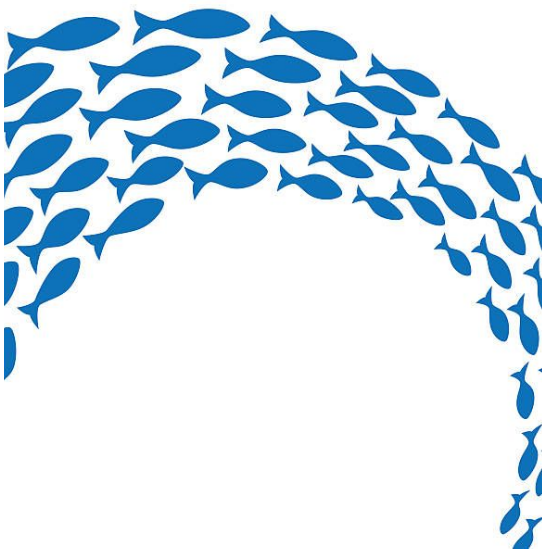
Word and senses connectivity: example 1

company^2_n (military unit)

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

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

Word and senses connectivity: example 2



$school_n^7$ (group of fish)

AutoExtend	SW2V
school	schools_n^7
school_n^4	sharks_n^1
school_n^6	sharks
school_v^1	shoals_n^3
school_n^3	fish_n^1
elementary	dolphins_n^1
schools	pods_n^3
elementary_a^3	eels
school_n^5	dolphins
elementary_a^1	whales_n^2

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

More information on knowledge-based embeddings

From Word to Sense Embeddings: A Survey on Vector Representations of Meaning (2018)

https://arxiv.org/abs/1805.04032

QUESTION 7
PIN: 7024700
www.kahoot.it

Integration of knowledge-based sense representations into NLP tasks

- **Taxonomy Learning** (Espinosa-Anke et al. AAAI, 2016)
- **Open Information Extraction** (Delli Bovi et al. EMNLP 2015).
- **Lexical entailment** (Nickel & Kiela, NIPS 2017)
- **Word/Entity Disambiguation** (Rothe & Schütze, ACL 2015)
- **Sentiment analysis** (Flekova & Gurevych, ACL 2016)
- **Lexical substitution** (Cocos et al., SENSE 2017)
- **Computer vision** (Young et al. ICRA 2017)

Text Classification (Pilehvar et al., ACL 2017)

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

Text Classification

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

Problems:

- WSD is not perfect

Text Classification

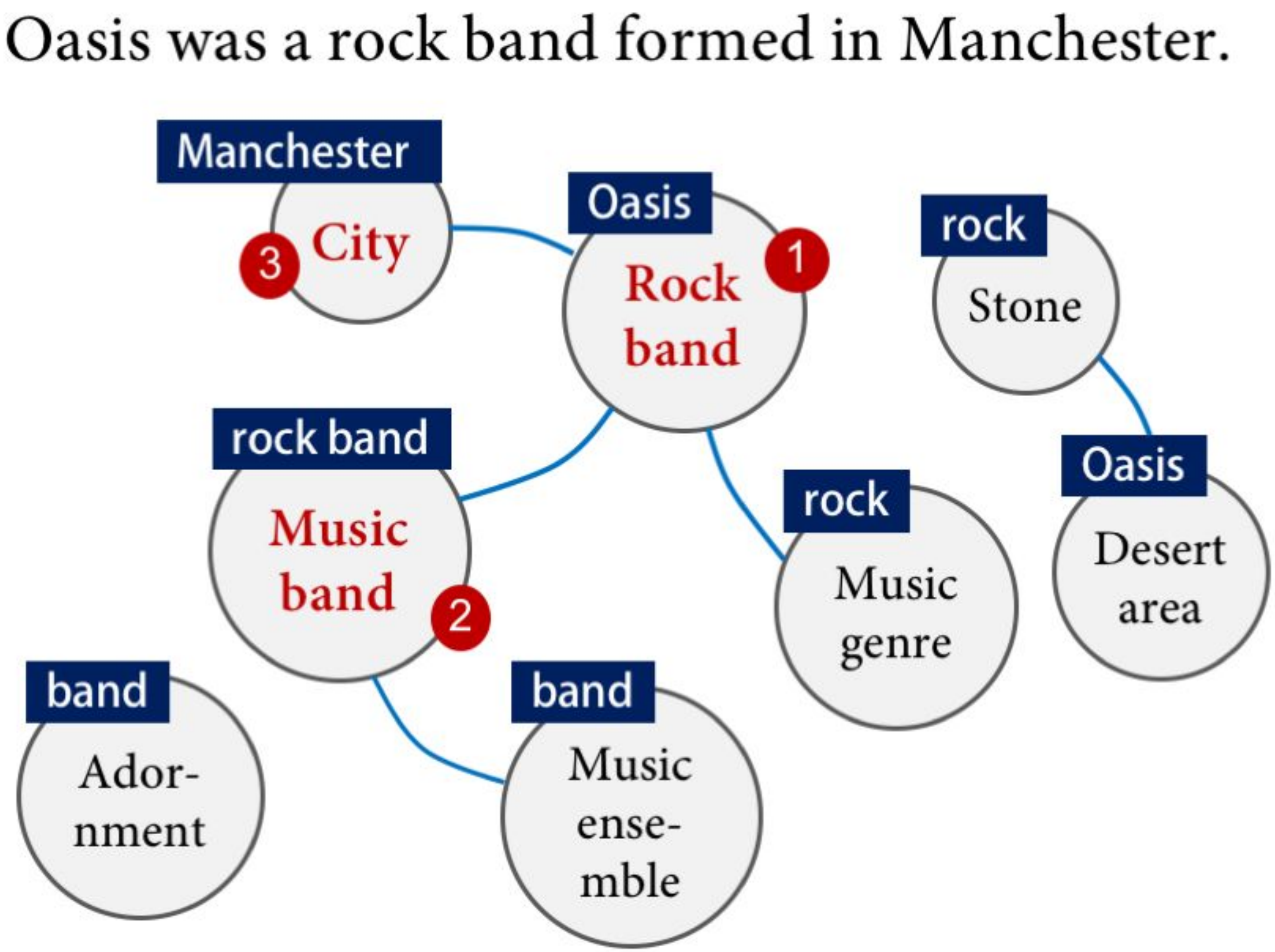
Question: What if we apply WSD and inject sense embeddings to a standard neural classifier?

Problems:

- WSD is not perfect

-> Solution: High-confidence disambiguation

High confidence graph-based disambiguation



Text Classification

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

Text Classification

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

11	noun.event	nouns denoting natural events
12	noun.feeling	nouns denoting feelings and emotions
13	noun.food	nouns denoting foods and drinks
14	noun.group	nouns denoting groupings of people or objects
15	noun.location	nouns denoting spatial position
16	noun.motive	nouns denoting goals
17	noun.object	nouns denoting natural objects (not man-made)
18	noun.person	nouns denoting people
19	noun.phenomenon	nouns denoting natural phenomena

-> **Solution:** Supersenses

Text Classification

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

Text Classification

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

WIKIPEDIA
The Free Encyclopedia

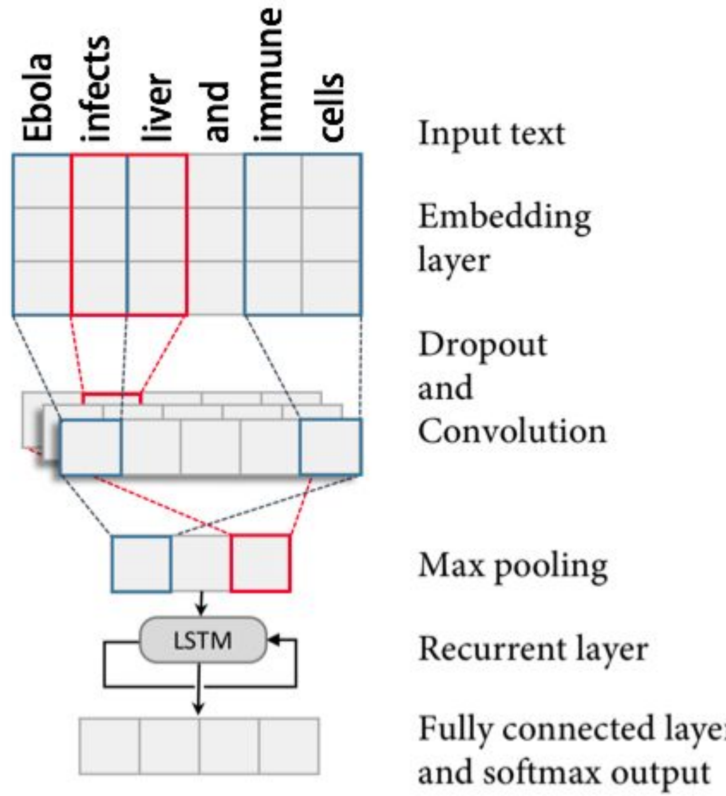
Tasks: Topic categorization and sentiment analysis

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

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

Word-based classification model

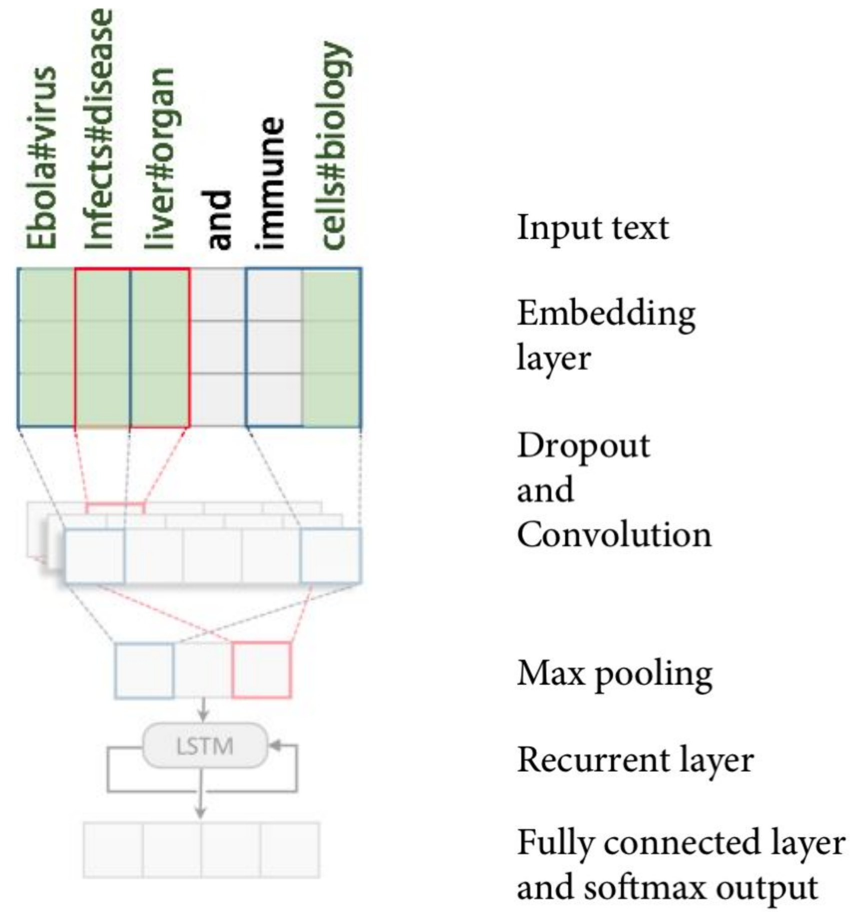
Standard CNN classifier inspired by Kim (2014) and Xiao and Cho (2016)



Conventional word-based system

Sense-based classification model

Standard CNN classifier inspired by Kim (2014) and Xiao and Cho (2016)



Sense-integrated system

Sense-based vs. word-based: Conclusions

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

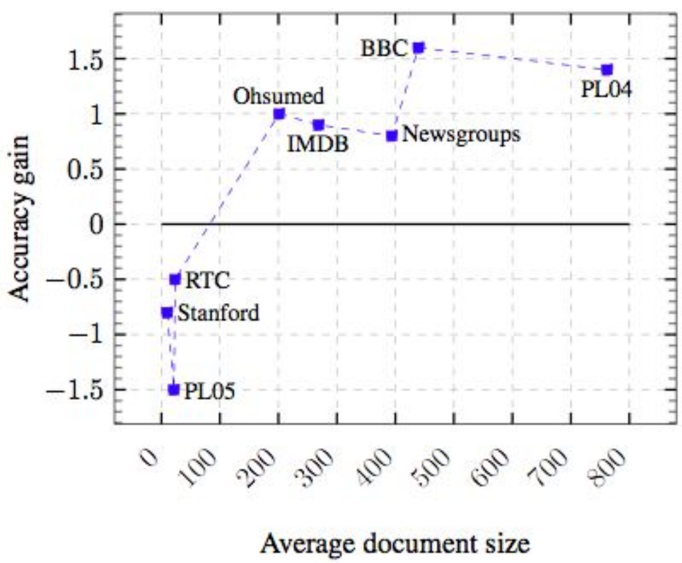
Sense-based vs. word-based: Conclusions

- 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

- Disambiguation increases sparsity

CONCLUSION

Conclusion

- NLP for Lexical Resources

- Lexical Resources for NLP

NLP for Lexical Resources

Challenge: Evaluation

- Comparison is difficult.
- Solving a problem from industrial and academic point of view is different.
  - If automatic extension, ideally extremely high quality at the expense of recall.
  - But high recall applications must also consider the posterior editing and validation.

Lexical Resources for NLP

Encouraging results at the lexical level.

Challenge: Scaling it to sentences and documents:

- Sensitivity to word order
- Combine vectors into syntactic-semantic structures
- Requires disambiguation, semantic parsing, etc.
- Compositionality

Challenges

Challenge: Addressing multilinguality

- Most work/resources so far for English
- Potential in multilingual and cross-lingual applications (e.g. BabelNet, ConcepNet)

Google Group: goo.gl/JEazYH