Hypernym Detection

Shaurya Rawat

Ontology Engineering Group
Universidad Politécnica de Madrid, Spain

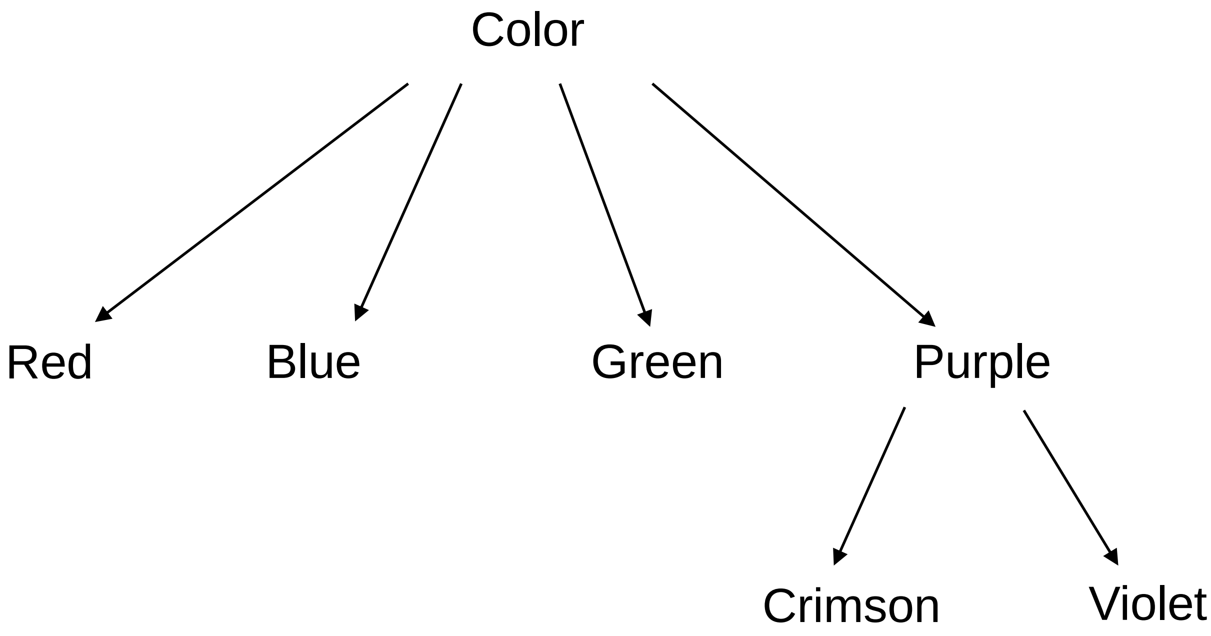
srawat@delicias.fi.upm.es

23/1/2020

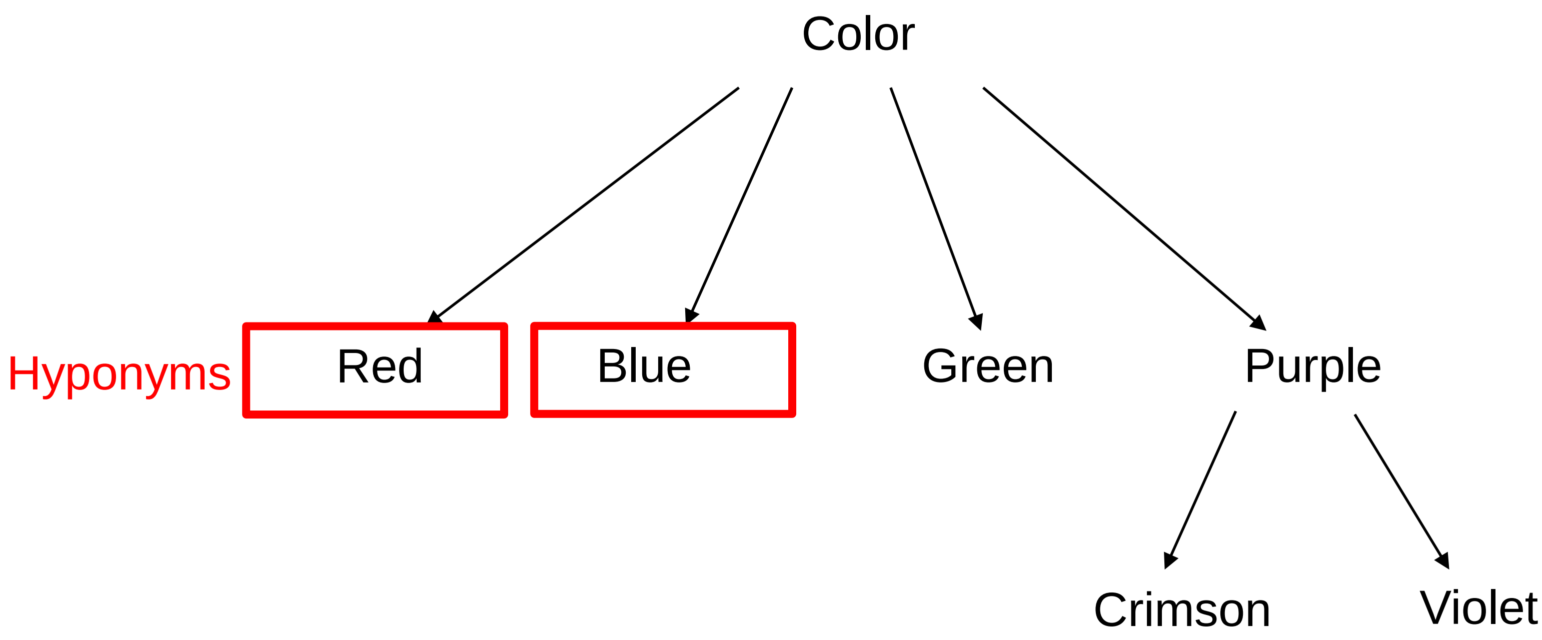
- What are Hypernyms?

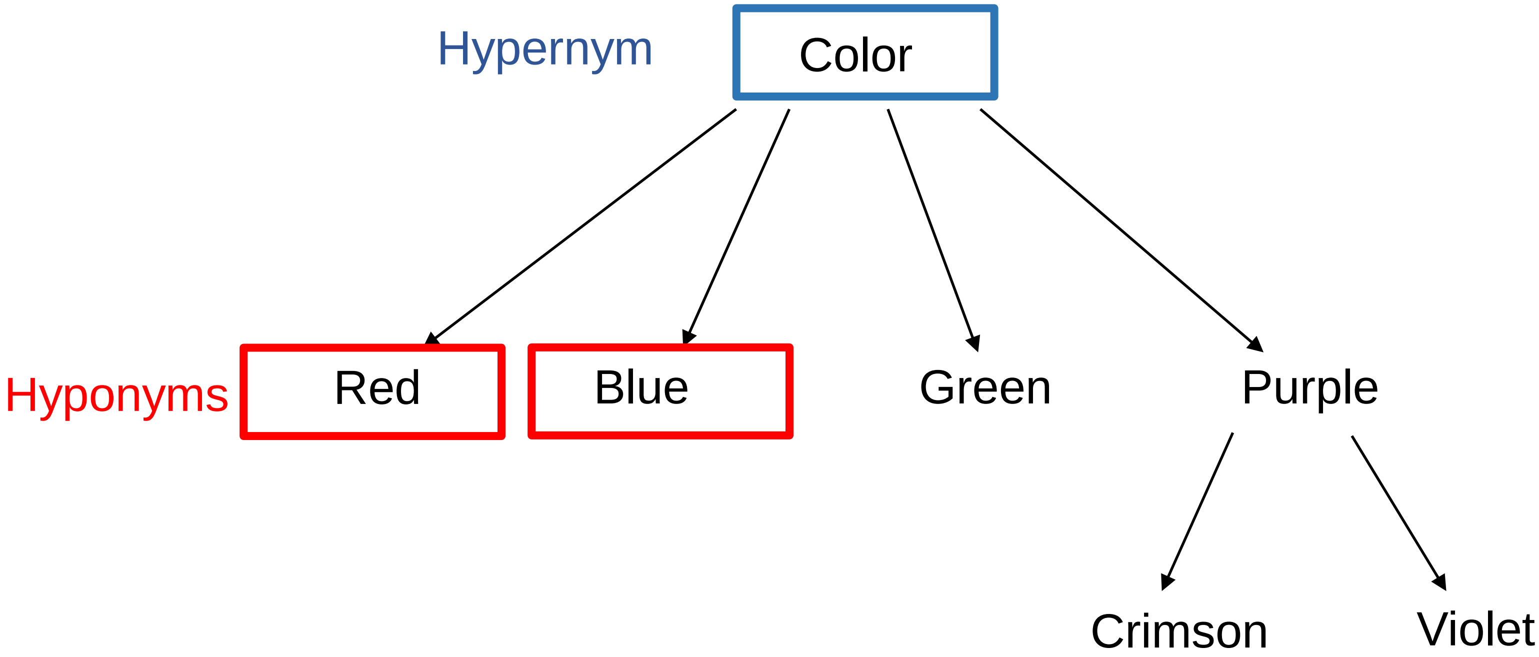
A *hyponym* is a word or phrase whose semantic field is included within that of another term, which is called it's ***hypernym***

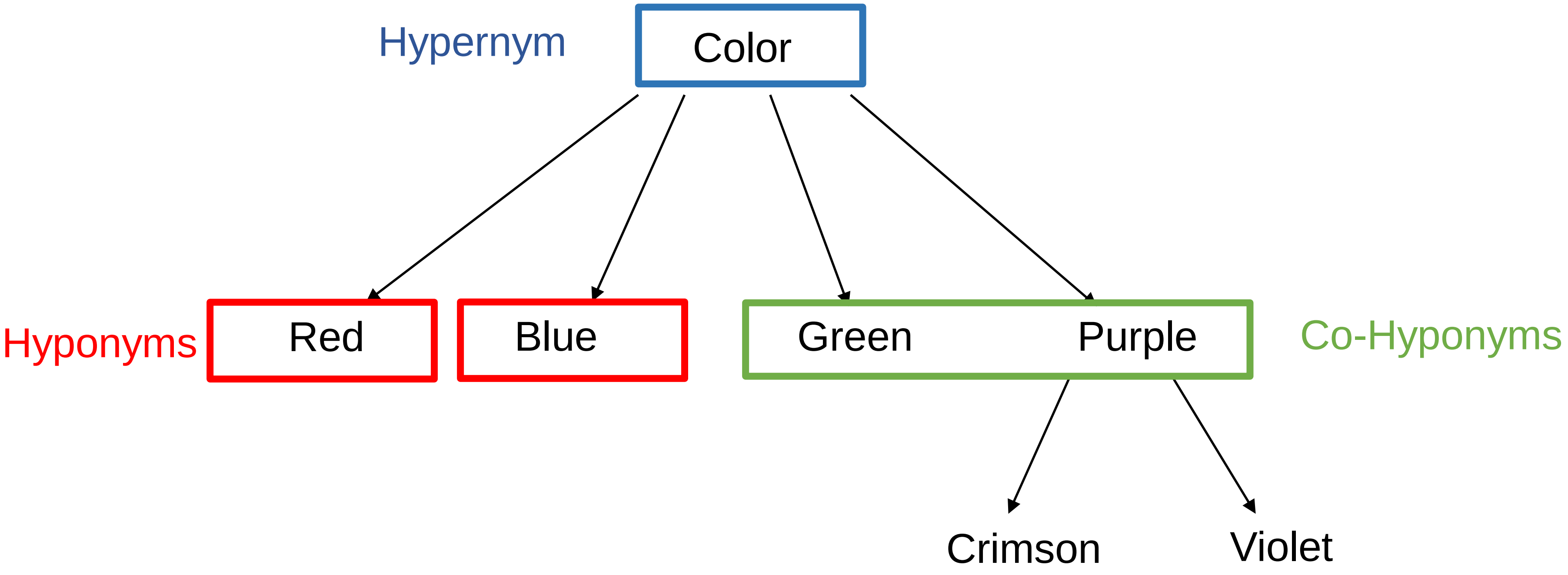
It generally covers the ***type-of*** or ***is-a*** relationship between 2 words.


Color
Red
Blue
Green
Purple
Crimson
Violet

Example




Hypernym
Color
Hyponyms
Red
Blue
Green
Purple
Crimson
Violet


Hypernym
Color
Hyponyms
Red
Blue
Green
Purple
Co-Hyponyms
Crimson
Violet

Approaches for Hypernym detection from text:

Approaches for Hypernym detection from text:

1) Pattern Based Approaches

Approaches for Hypernym detection from text:

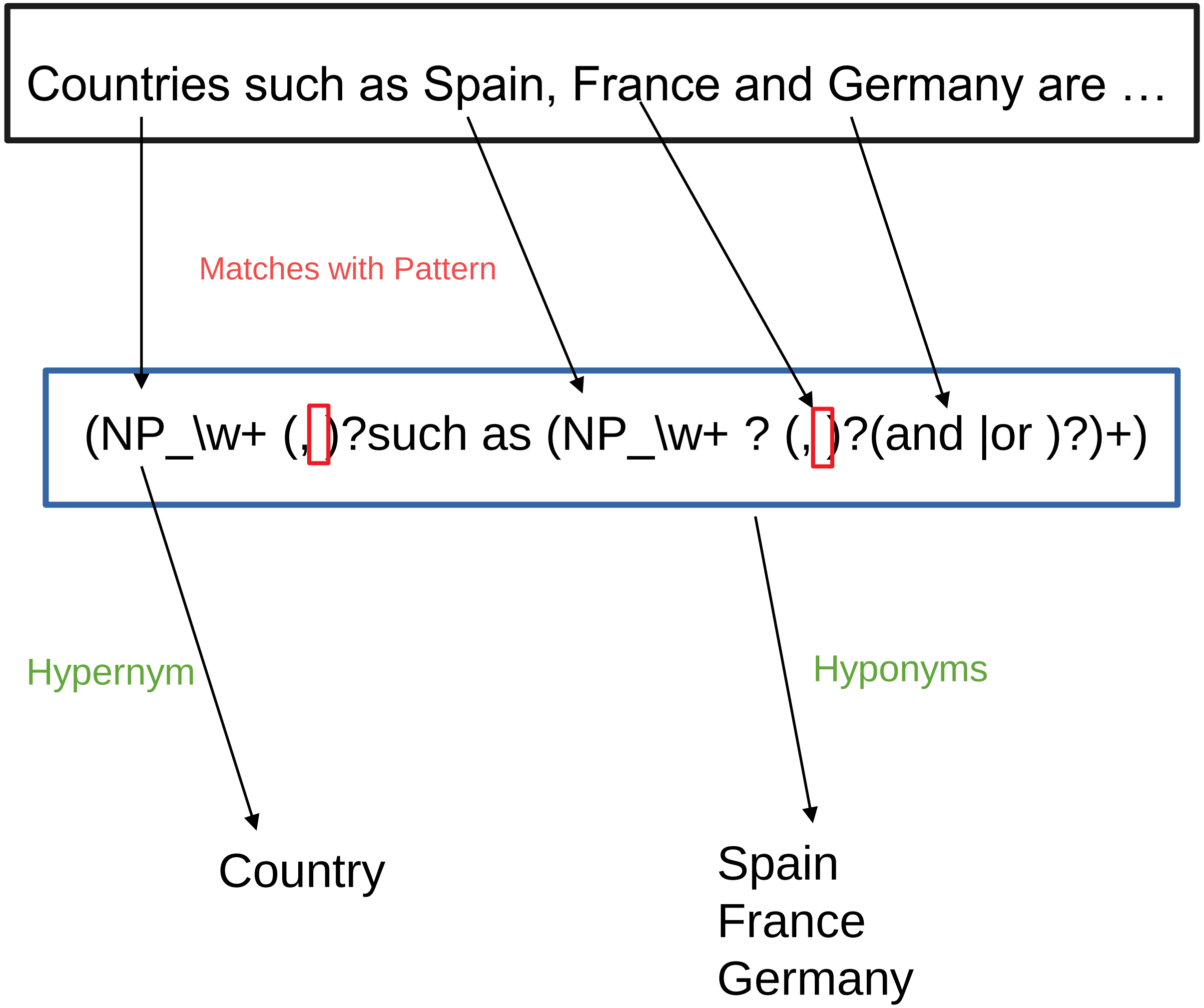
1) Pattern Based Approaches

Using lexico-synctactic patterns (LSPs) to extract Hypernyms from text

Most popular: Hearst Patterns by Marti Hearst

Example:

“NPx such as NPy” or “NPx and other NPy”


Countries such as Spain, France and Germany are …
Matches with Pattern
(NP_\w+ (,)?such as (NP_\w+ ? (,)?(and |or)?)+)
Hypernym
Hyponyms
Country
Spain
France
Germany

Approaches for Hypernym detection from text:

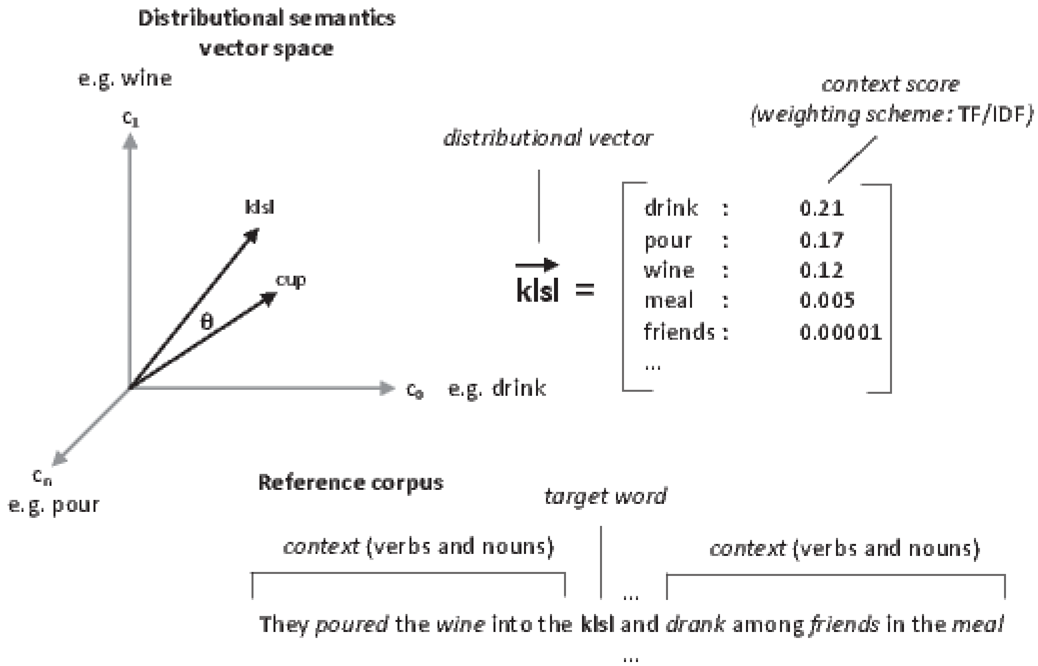
1) Pattern Based Approaches

1) Unsupervised Distributional Approaches

Approaches for Hypernym detection from text:

2) Unsupervised Distributional Approaches

Creating distributional semantic spaces based on target and context of different words



Approaches for Hypernym detection from text:

2) Unsupervised Distributional Approaches

Most approaches based on ***DIH(Distributional Inclusion Hypothesis)***

Measures:

- WeedsPrec (Weeds et al. 2004)
- invCL (Lenci and Benotto 2012)
- SLQS (Santus et al. 2014, Shwartz et al. 2017)

Approaches for Hypernym detection from text:

1) Pattern Based Approaches

1) Unsupervised Distributional Approaches

1) Machine Learning Based Approaches

Approaches for Hypernym detection from text:

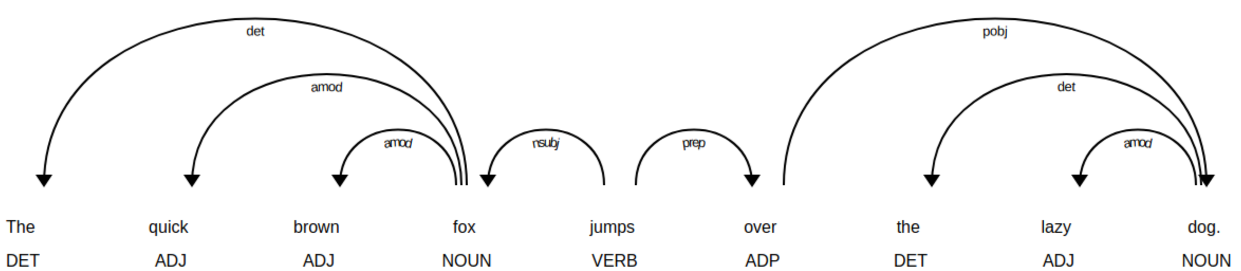
3) Machine Learning Based Approaches

Using Dependency Paths between known known hypernym pairs to train a classifier (Snow et al. 2005, Sheena et al. 2016)

Hyponymy shows the relationship between a generic term (hypernym)
hypernym, also known as a superordinate, is broader than that of a hyp
difficult with abstract words such as *imagine*, *understand* and *knowled*
broad category of actions. For example, verbs such as *stare*, *gaze*, *view*

Hypernyms and hyponyms are asymmetric. Hyponymy can be tested b
not 'A tool is a kind of screwdriver'.

Train



Hypernym Classifier

Takes dependency path between two words as input and classifies it as Hypernym or not

State of the Art

Paper: **Roller et al.2018 (Facebook AI)**

Hearst Patterns Revisited:
Automatic Hypernym Detection from Large Text Corpora

Stephen Roller, Douwe Kiela, and Maximilian Nickel
Facebook AI Research
{roller,dkiela,maxn}@fb.com

Pattern based and Distributional models were evaluated against each other on various datasets, using Average Precision.

→ Conclusions from the Paper:

Pattern based methods outperform Distributional methods.

State of the Art (Results from Paper):

	Detection (AP)				
	BLESS	EVAL	LEDS	SHWARTZ	WBLESS
Cosine	.12	.29	.71	.31	.53
WeedsPrec	.19	.39	.87	.43	.68
invCL	.18	.37	**.89**	.38	.66
SLQS	.15	.35	.60	.38	.69
p(x, y)	.49	.38	.71	.29	.74
ppmi(x, y)	.45	.36	.70	.28	.72
sp(x, y)	.66	.45	.81	.41	.91
spmi(x, y)	**.76**	**.48**	.84	**.44**	**.96**

Datasets Used

(For extraction of Hypernyms using Pattern based Approaches)

In the SOTA paper, Dataset used:

Gigaword (4 Billion words) + Wikipedia

We used:

UMBC Corpus (3 Billion Words) + Wikipedia

Link to UMBC Corpus: https://ebiquity.umbc.edu/resource/html/id/351

Datasets Used

(For extraction of Hypernyms using Pattern based Approaches)

In the SOTA paper, Dataset used:

Gigaword (4 Billion words) + Wikipedia

$6,000.00 Non-Member
$3,000.00 Reduced-License
$250.00 Extra Copy

Datasets Used

(For extraction of Hypernyms using Pattern based Approaches)

In the SOTA paper, Dataset used:

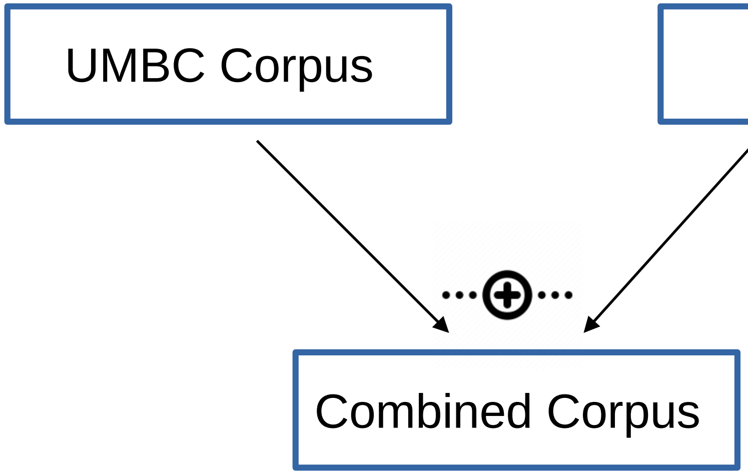
Gigaword (4 Billion words) + Wikipedia

$6,000.00 Non-Member
$3,000.00 Reduced-License
$250.00 Extra Copy

We used:

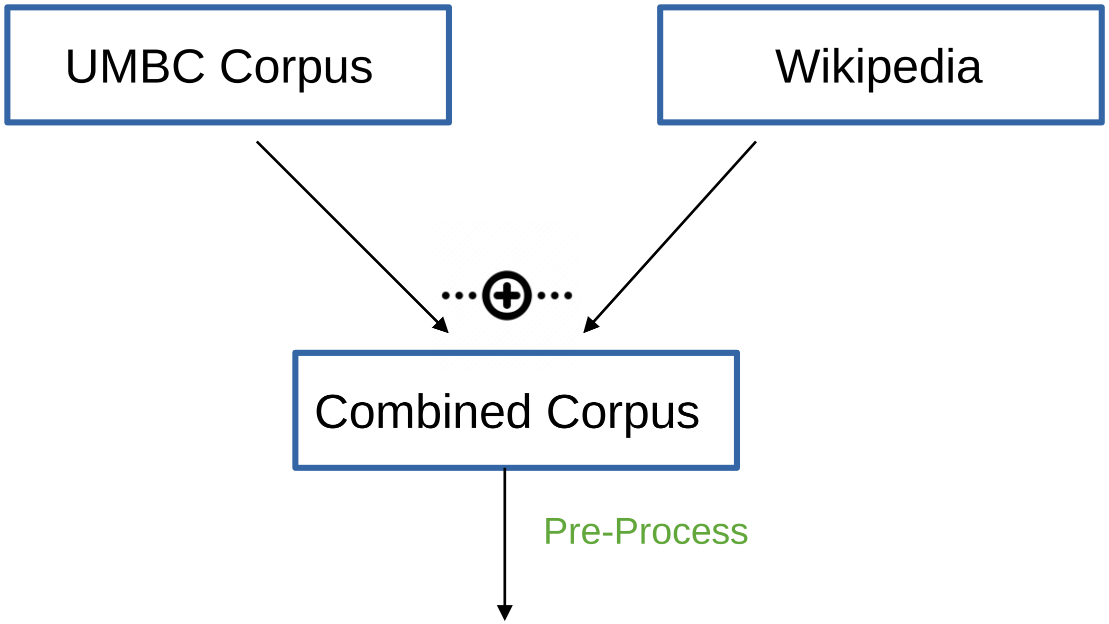
UMBC Corpus (3 Billion Words) + Wikipedia

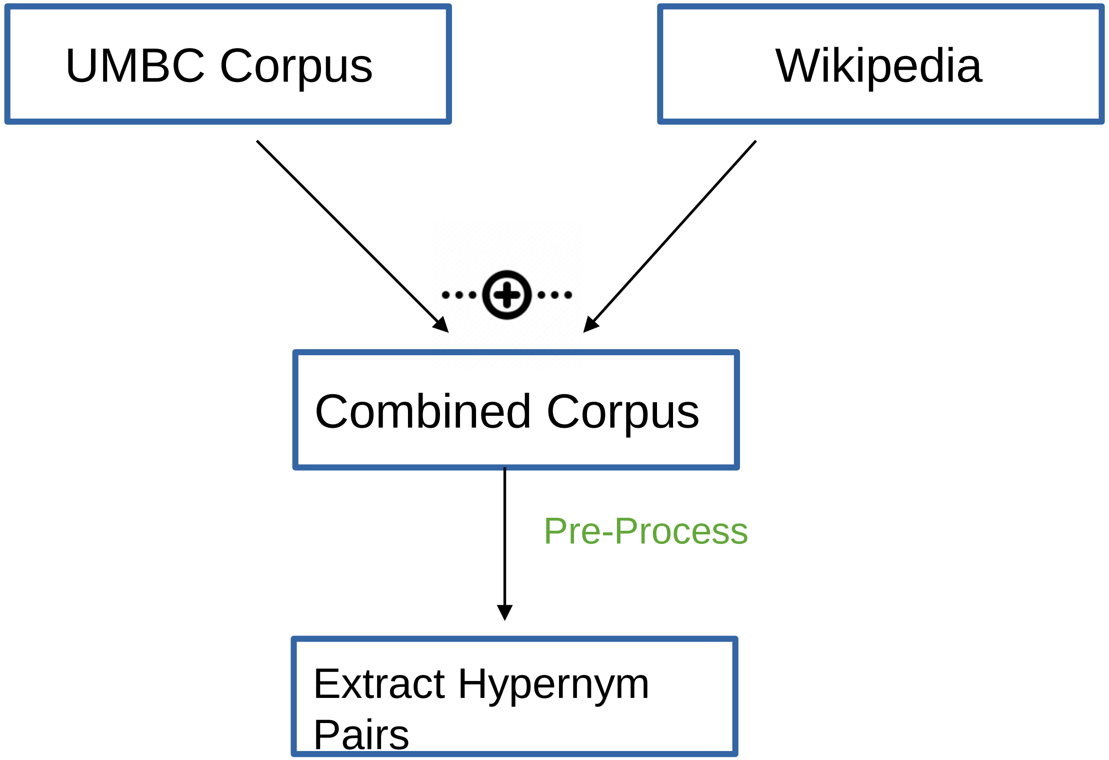
Link to UMBC Corpus(Freely available from University of Maryland Baltimore): https://ebiquity.umbc.edu/resource/html/id/351



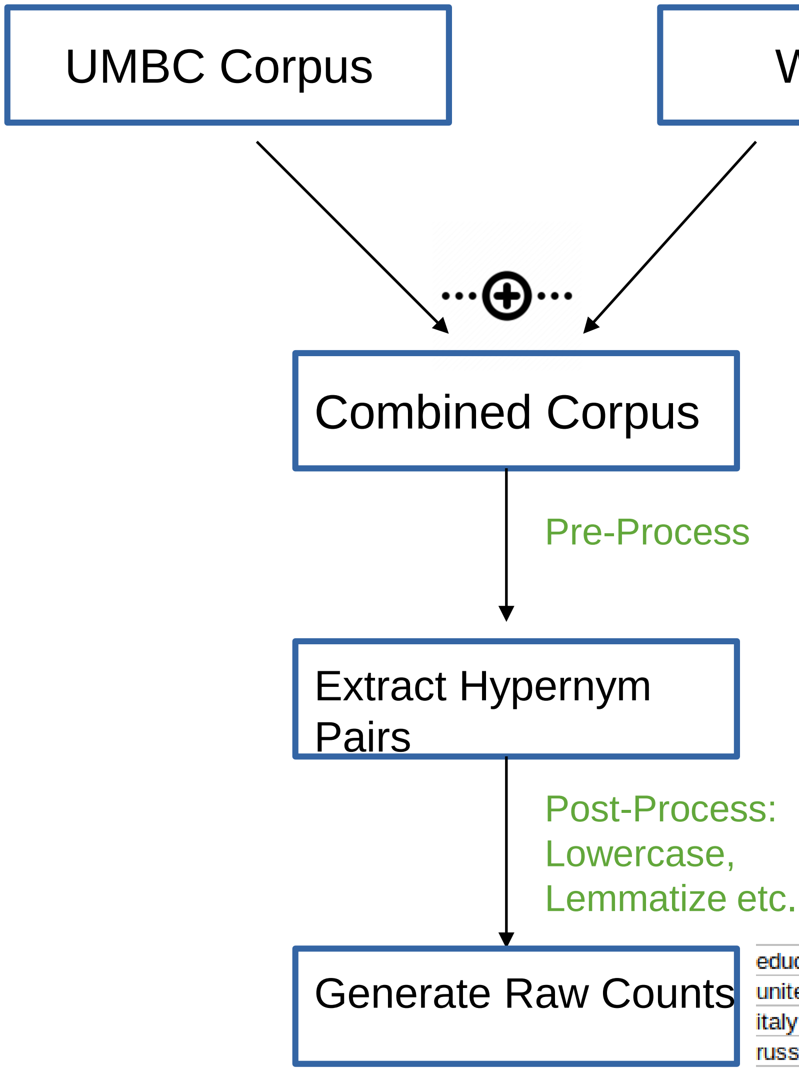
31 GB .txt file with raw text

Methodology



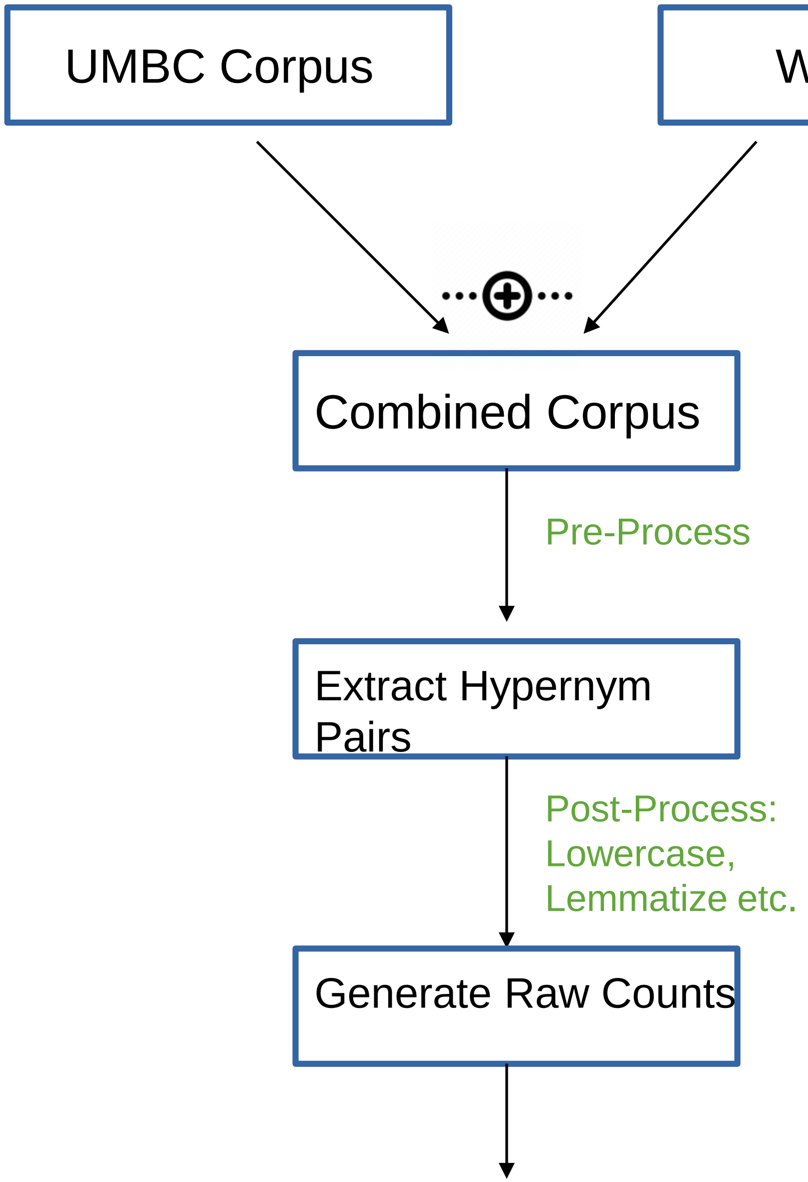



```
pigs    animal species
cattle  animal species
dogs    animal species
fish    animal species
cats    animal species
rabbits animal species
the World Health Organization   other international groups
```



education	area	1339
united_kingdom	country	1304
italy	country	1234
russia	country	1228
cancer	disease	1142

Methodology



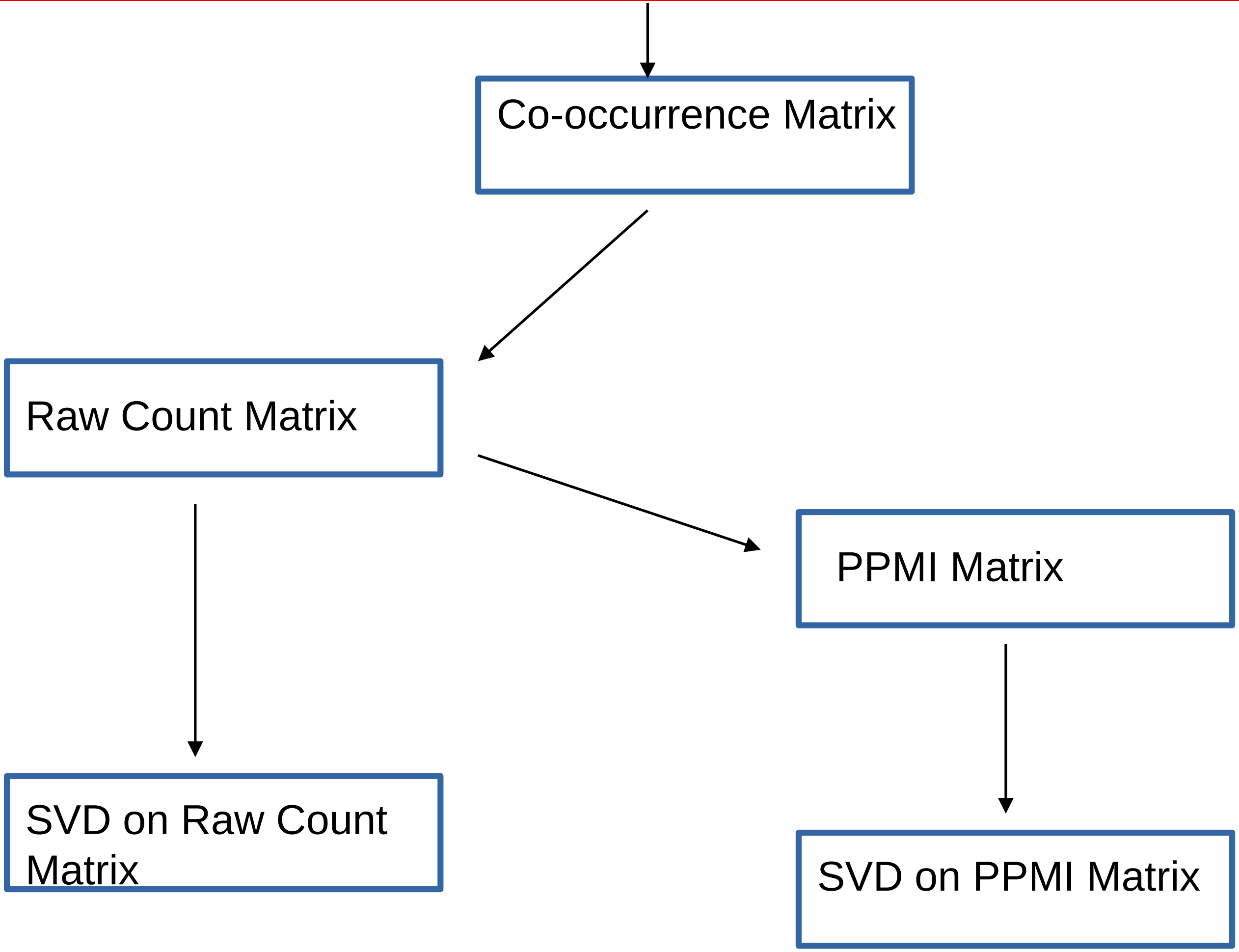
Co-occurrence Matrix

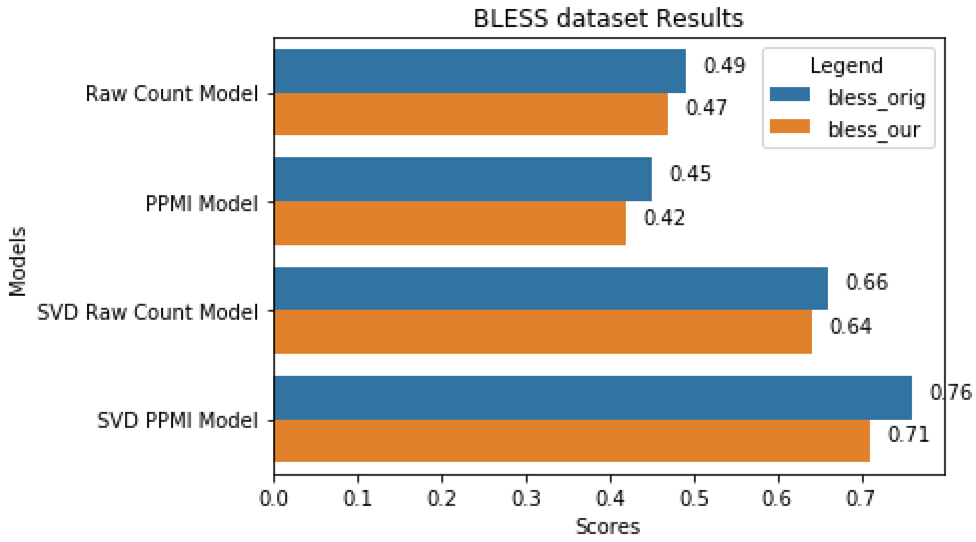
w = words

c = contexts

f_{ij} = frequency of cooccurrence

	C1	C2	C3	C4	C5
W1	1	0	0	2	0
W2	0	4	1	0	0
W3	2	0	0	1	0

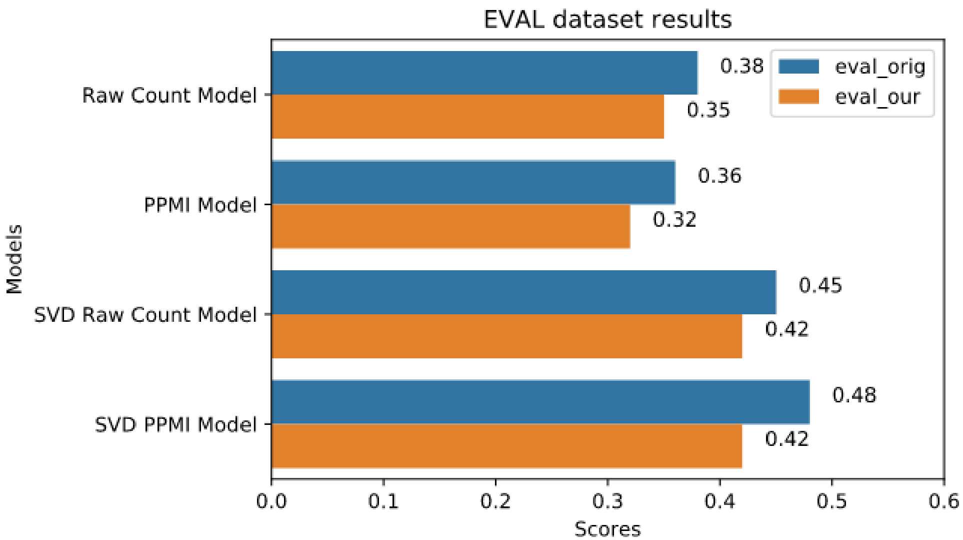

Co-occurrence Matrix
Raw Count Matrix
PPMI Matrix
SVD on Raw Count Matrix
SVD on PPMI Matrix



- **BLESS (Baroni and Lenci 2011)**

- 14 572 pairs
- 1337 positive examples

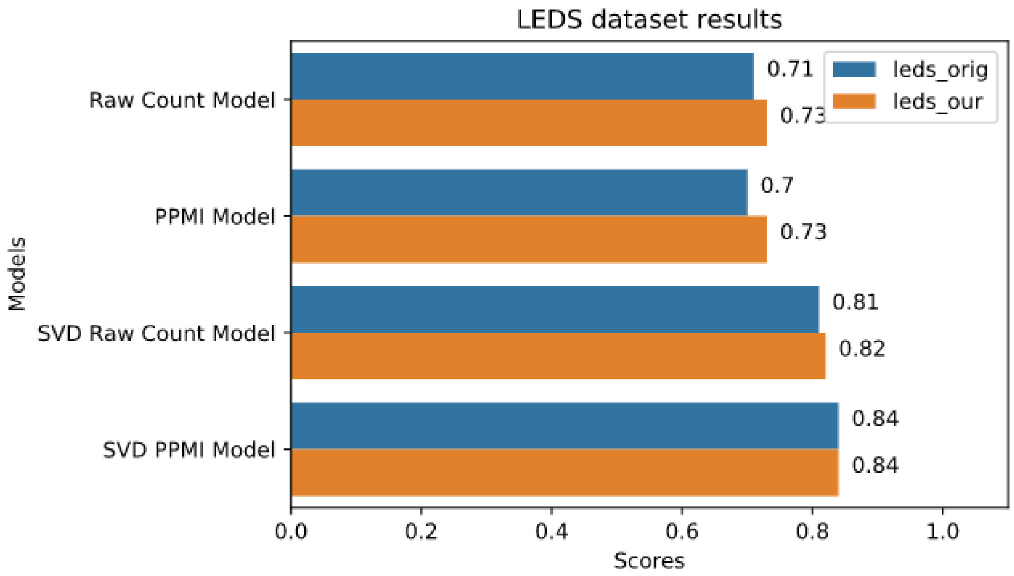
alligator	crocodile	False
alligator	frog	False
alligator	lizard	False
alligator	snake	False
alligator	toad	False
alligator	turtle	False
alligator	animal	True
alligator	beast	True
alligator	carnivore	True
alligator	chordate	True
alligator	creature	True
alligator	predator	True
alligator	reptile	True
alligator	vertebrate	True
alligator	eye	False
alligator	foot	False



- **EVAL (Santus et al. 2015)**

- 7 378 pairs

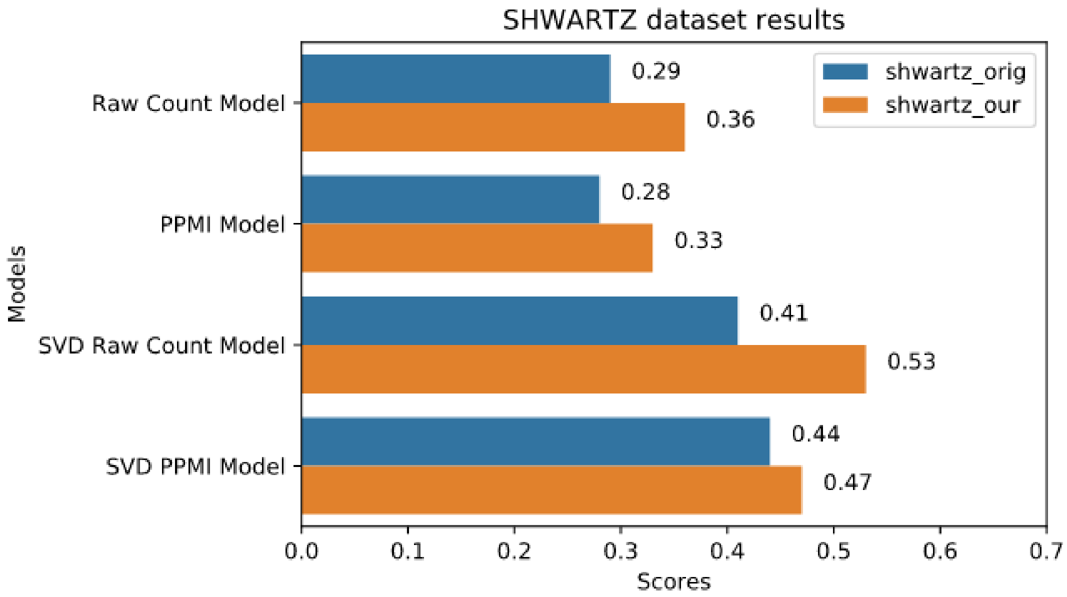
hate	dislike	True	hyper	val
hate	dislike	True	hyper	val
hate	dislike	True	hyper	val
heart	courage	False	hyper	val
black	color	False	hyper	val
black	color	False	hyper	val
black	color	False	hyper	val
want	demand	True	hyper	val
want	demand	True	hyper	val
queen	insect	True	hyper	val



- **LEDS (Baroni et al. 2012)**

- 2 770 pairs
- balanced randomly shuffled pairs

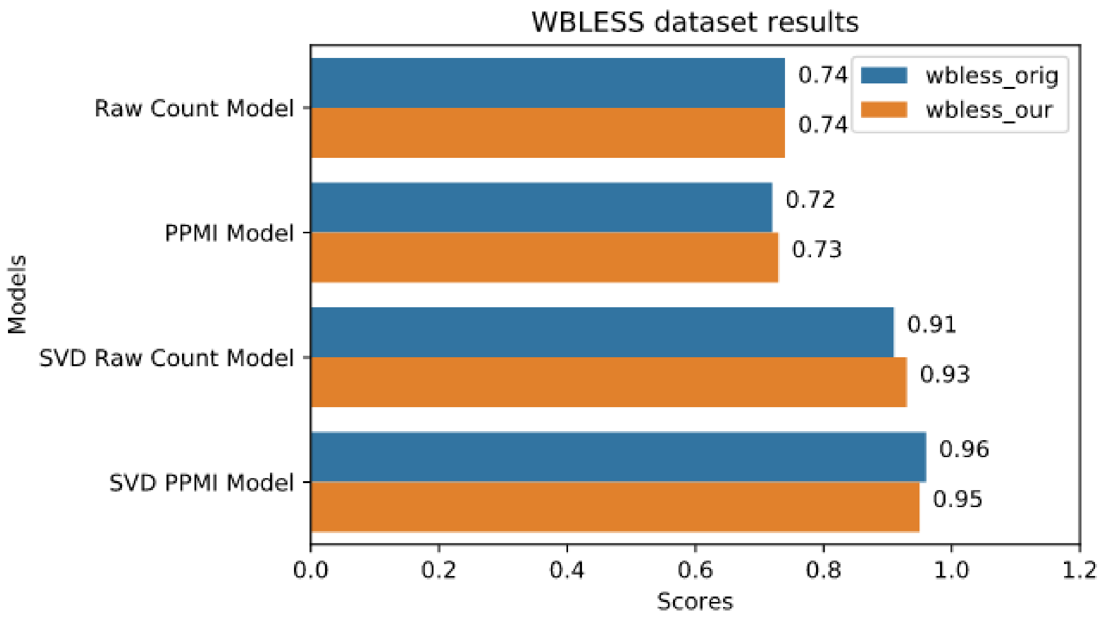
meal	sport	False	val
poet	writer	True	val
garment	jean	False	val
worker	traveler	False	val
hamlet	community	True	val
velocity	rate	True	val
berry	diplomat	False	val
integer	eleven	False	val



- **SHWARTZ (Shwartz et al. 2016)**

- 52 578 pairs

icomb	gloucestershire	False	val
cumhuriyet	turkey	False	val
lomonosov	russia	False	val
personality	soul	False	val
beyer	garratt	False	val
kimberley	australia	False	val
nesiotica	genus	False	val
oppland	norway	False	val
anglet	bayonne	False	val
nicholas	russia	False	val



- **WBLESS (Weeds et al. 2014)**

- 1 668 pairs

cloth	table	False	other	val
butt	gun	False	other	val
cow	cattle	True	hyper	val
whale	mammal	True	hyper	val
radio	device	True	hyper	val
cod	food	True	hyper	val
blouse	clothes	True	hyper	val
peel	banana	False	other	val
good	stove	False	other	val
cucumber	garlic	False	other	val
giraffe	violin	False	other	val
cloak	covering	True	hyper	val

Reasons for slight deviation in Results from SOTA:

On some datasets, we perform better and on some we are slightly below the SOTA.

1) Slight variations in datasets used
2) Difference in Pre and post processing methodologies

Results difference is negligible and our pattern based methods still outperform

Future work from here:

1) Extract dependency paths between known hypernym pairs from large text corpora and form new patterns

2) Explore Distributional methods that form space through linguistic features. Trying to encode patterns in DSMs

3) ML/DL approaches

Thank You!