

# Acquiring Comparative Commonsense Knowledge from the Web

Niket Tandon

Max Planck Institute for Informatics

Saarbrücken, Germany

Joint work with: Gerard de Melo, Gerhard Weikum



# Comparative Commonsense



- Siri shows nearby restaurants "In & Out Burger"
- "I would like something \*healthier than\* burgers."

# Related work

## Knowledge Harvesting

- Pattern Extraction & Open IE
- No comparative commonsense relations
- Disambiguation of triples
- Named entities but not nouns

## Commonsense Knowledge Bases

- Manual (Cyc),
- Semi-automated (ConceptNet),
- Automated (WebChild)
  - No comparative commonsense

## Comparative commonsense

This work: construction of a

- comparative commonsense KB,
- semantically refined,
- large-scale

# Semantically refined Comparative Commonsense

... “bullet trains” travel “quicker than” “a jaguar” ...

Pattern based extraction over ClueWeb



<bullet train, quick, jaguar>

1.  
Extraction



2.  
Disambiguation



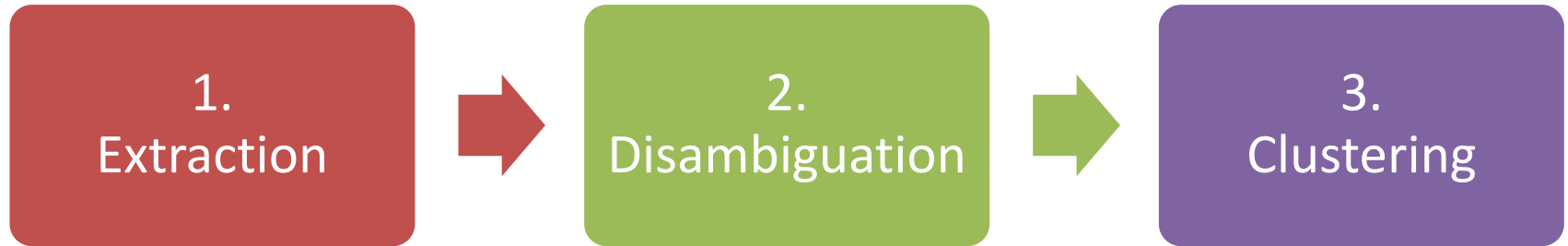
3.  
Clustering

# Semantically refined Comparative Commonsense



Open IE style extraction  
< bullet train , quick , jaguar >

# Semantically refined Comparative Commonsense

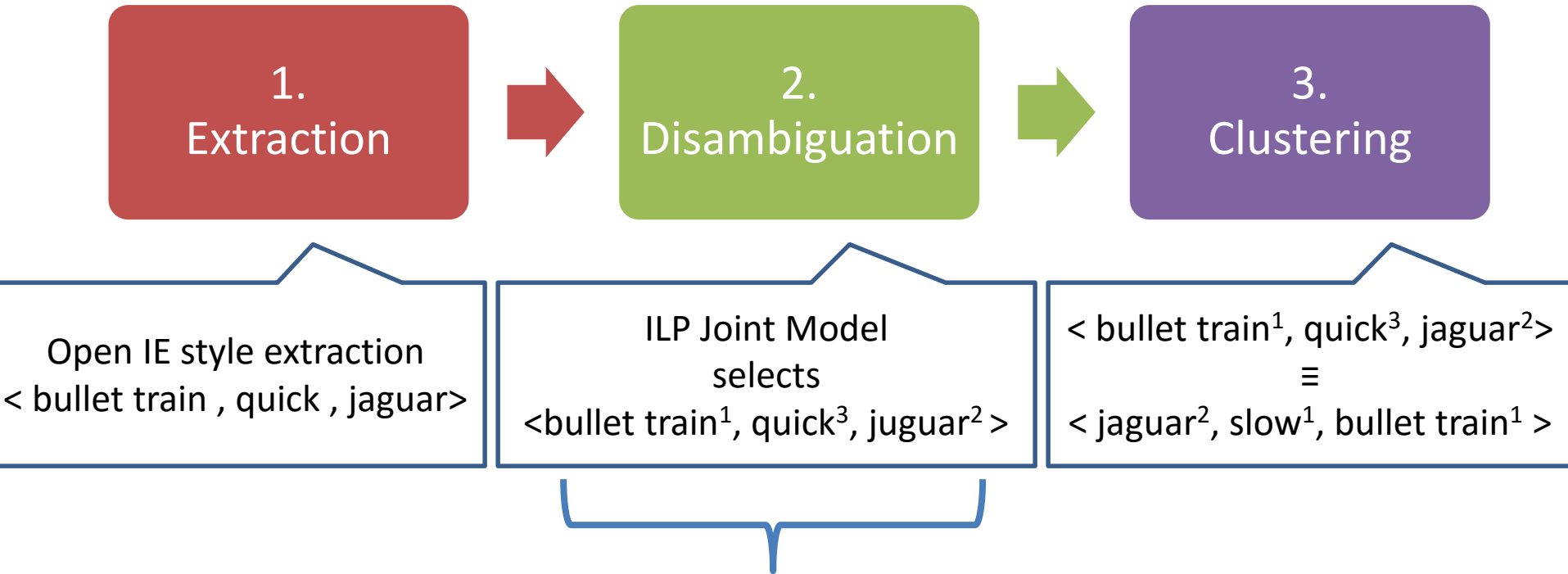


Open IE style extraction  
< bullet train , quick , jaguar >

ILP Joint Model  
selects  
<bullet train<sup>1</sup>, quick<sup>3</sup>, jaguar<sup>2</sup>>

Argument Type	Argument1	Relation/ Adjective	Argument2
<b>both WN</b>	snow-n-2	less dense-a-3	rain-n-2
<b>WN/ad hoc</b>	little child-n-1	happier (happy-a-1)	adult-n-1
<b>both ad hoc</b>	wet wood-n-1	softer (soft-a-1)	dry wood-n-1

# Semantically refined Comparative Commonsense



Argument Type	Argument1	Relation/ Adjective	Argument2
<b>both WN</b>	snow-n-2	less dense-a-3	rain-n-2
<b>WN/ad hoc</b>	little child-n-1	happier (happy-a-1)	adult-n-1
<b>both ad hoc</b>	wet wood-n-1	softer (soft-a-1)	dry wood-n-1

# Disambiguation of ambiguous comparative triples

bullet train<sup>1</sup>, quick<sup>3</sup>, jaguar<sup>2</sup>





# Disambiguation of ambiguous comparative triples

< bullet train, quick, jaguar >  
 < train, slow, plane >  
 < plane, fast, train >  
 < bus, slow, plane >  
 < jaguar, slow, cheetah >

bullet train<sup>1</sup>, quick<sup>3</sup>, jaguar<sup>2</sup>

jaguar<sup>2</sup>, fast<sup>1</sup>, bus<sup>1</sup>  
 has a neighbor  
 bus<sup>1</sup>, slow<sup>1</sup>, car<sup>1</sup>

slow:  
 penalize  
 >1 senses

bus:  
 penalize  
 >1 senses

maximize

$$\sum_i \sum_j (\text{coh}_{ij} + \tau_{ij}) x_{ij} + \sum_i \sum_j \sum_k \sum_l \text{sim}_{ij,kl} B_{ij,kl} - \sum_{m \in \text{adjectives}} \sum_s a_{ms} - \sum_{m \in \text{nouns}} \sum_s n_{ms}$$

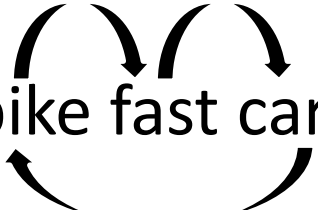
subject to

$$x_{ij} = x_{kl} \quad \forall i, j, k, l : \mu_{ij,kl} = 1$$

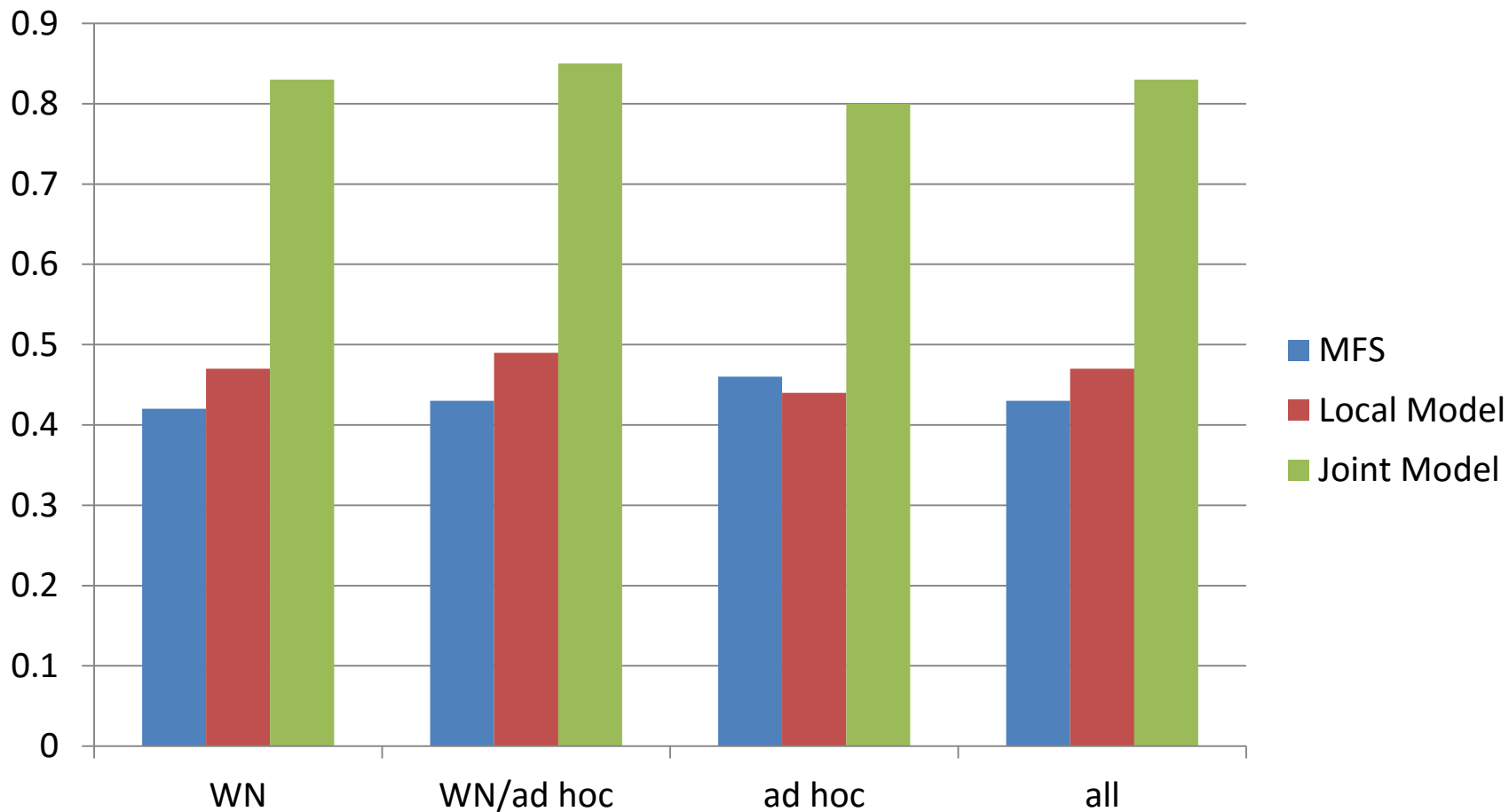
jaguar<sup>2</sup>, slow<sup>1</sup>, bullet train<sup>1</sup>  
 ≡  
 bullet train<sup>1</sup>, quick<sup>3</sup>, jaguar<sup>2</sup>

# Experiments

- Dataset for extraction:
  - ClueWeb09: 500 Million pages.
  - ClueWeb12: 733 Million pages.
- Extraction output (not disambiguated, noisy):
  - More than 1 million comparative facts extracted (e.g. bike, fast, car)
- Baselines (task: clean and disambiguate triples)
  - MFS: Most frequent sense: bike-n-1, fast-a-1, car-n-1

– Local Model:  bike fast car

# Evaluation Results (precision)



# Resultant Comparative commonsense KB more than 1 million semantically refined triples.

Argument Type	Argument1	Relation/ Adjective	Argument2
<b>both WN</b>	snow-n-2	less dense-a-3	rain-n-2
	marijuana-n-2	more dangerous-a-1	alcohol-n-1
<b>WN/ad hoc</b>	little child-n-1	happier (happy-a-1)	adult-n-1
	private school-n-1	more expensive-a-1	public institute-n-1
<b>both ad hoc</b>	peaceful resistance-n-1	more effective-a-1	violent resistance-n-1
	wet wood-n-1	softer (soft-a-1)	dry wood-n-1

# Conclusion

- First large-scale, semantically-refined Comparative Commonsense KB.
- Publicly available at:  
[mpii.de/yago-naga/webchild](http://mpii.de/yago-naga/webchild)