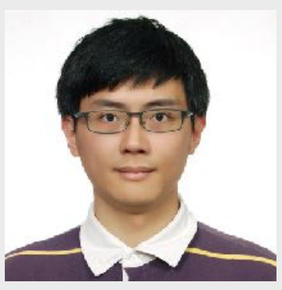


Haw-Shiuan Chang

NLP
Machine Learning

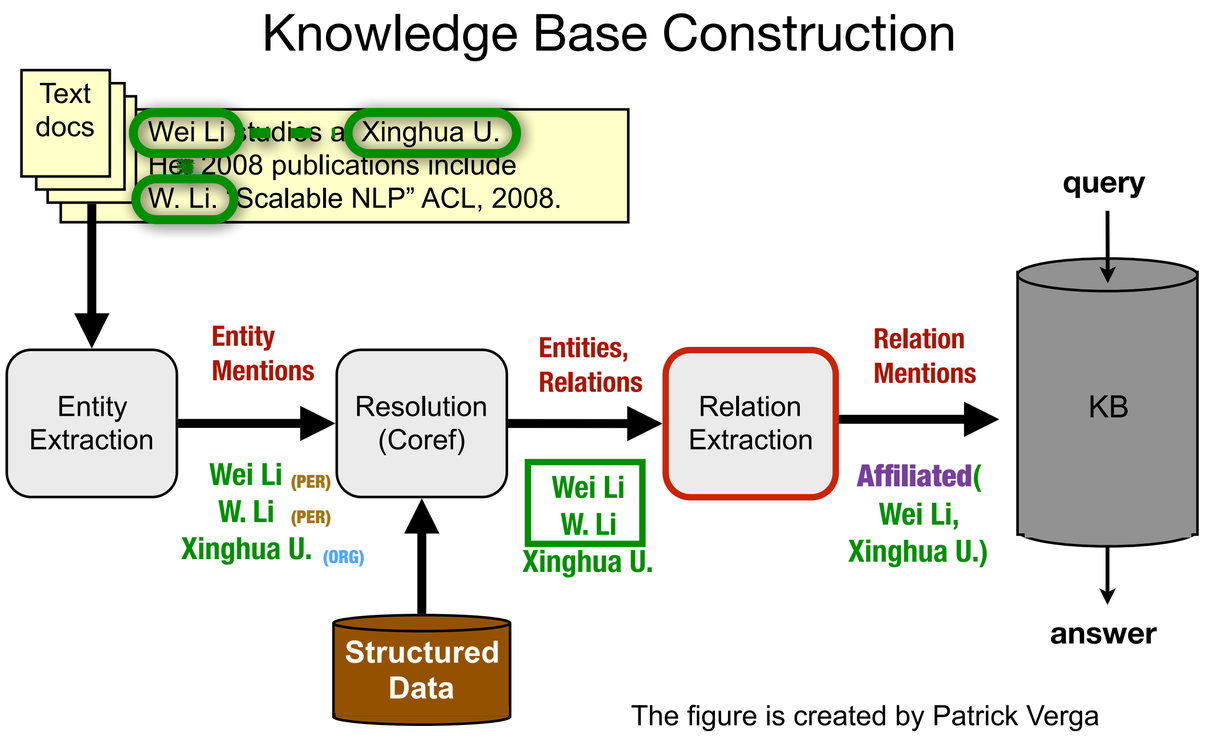
MS/PhD UMass CS, 2015-?

Seeking: **Internship Summer 2017**

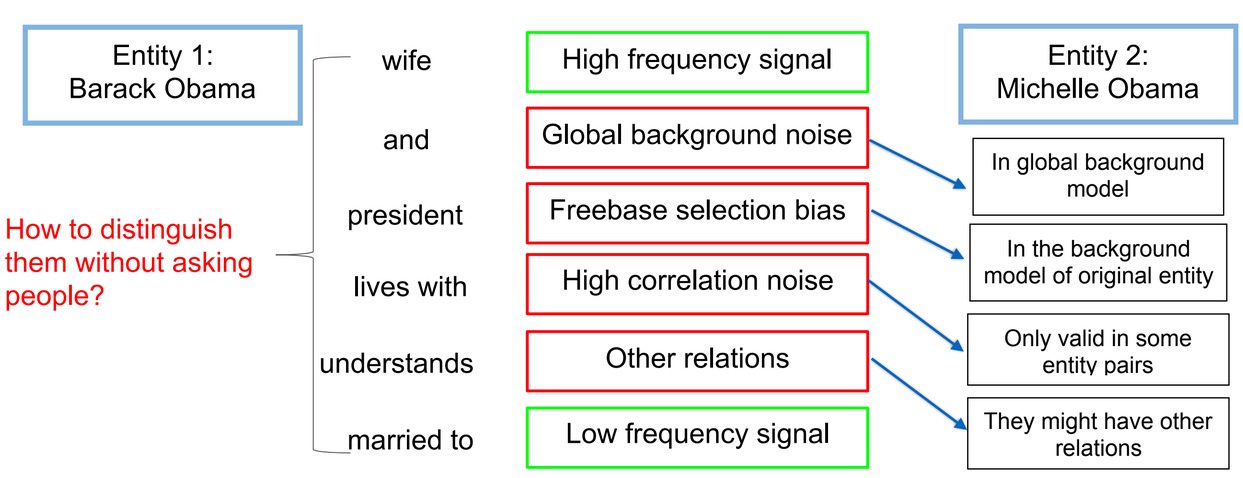
Currently work on NLP and Information extraction in IESL
Three years research experience on computer vision and active learning

Asking Google: Construct Knowledge Base from Cleaner Relation Patterns

Problem

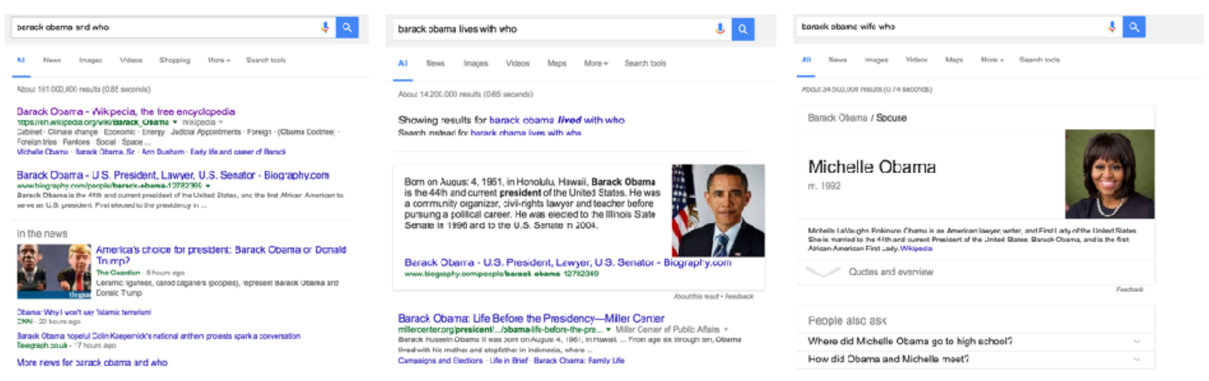


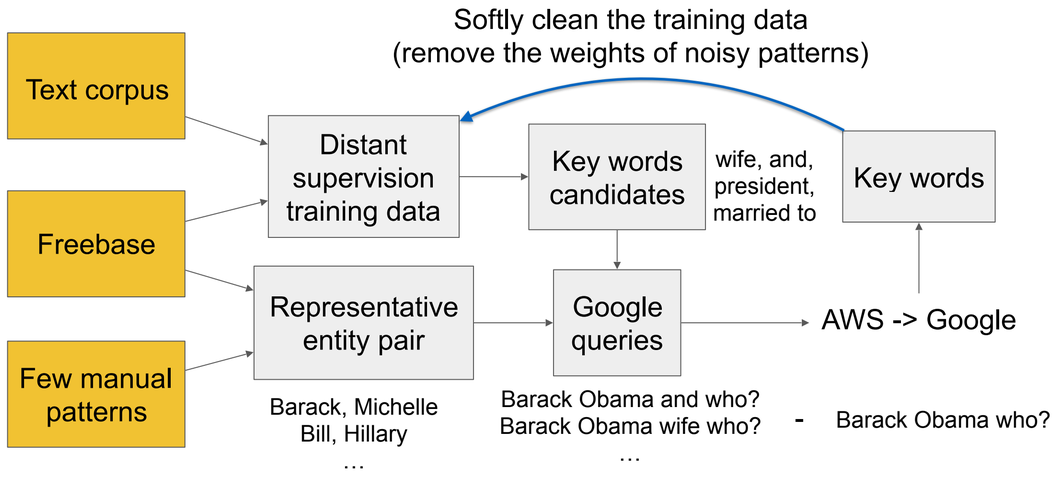
The figure is created by Patrick Verga



Solution

Search engines have been optimised to perform such type of queries





Results

Distant supervision (after applying several noises removal techniques)

wife
"his wife"
his
married
husband
"wife of"
"married to"
“her husband”
to
her

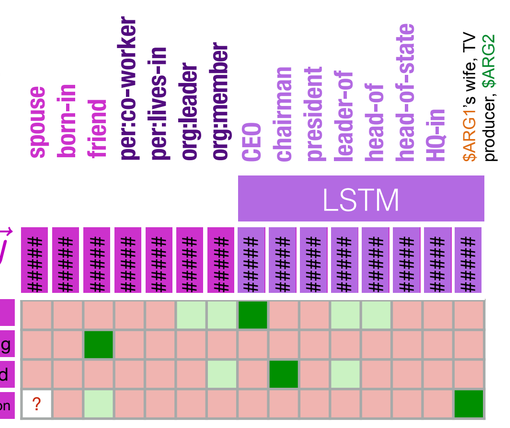
After applying google supervision

wife
"his wife"
married
husband
"married to"
"wife of"
"is married"
"is married to"
“her husband”
“(wife”

TAC slot filling 2012-2015

Query example:
[Christopher Noxon, spouse]
(i.e., Who is Christopher’s wife)

Answer Example:
Slot filler: Jenji Kohan
Provenance: Christopher Noxon’s wife, TV producer, Jenji Kohan

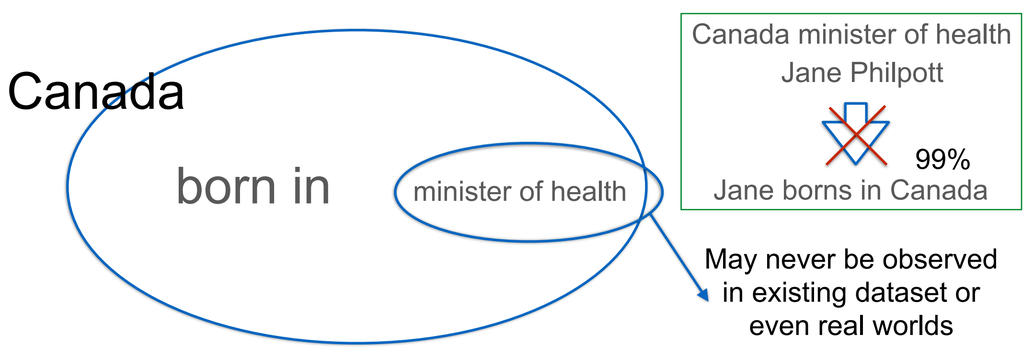


The figure is originally created by Patrick Verga

Model	Training data	High recall			High precision		
		F1	Precision	Recall	F1	Precision	Recall
USchema	Distant supervision (DS)	0.243	0.504	0.160	0.215	0.537	0.135
	DS+google	0.261	0.506	0.176	0.214	0.525	0.134
LSTM	Distant supervision (DS)	0.331	0.400	0.282	0.293	0.431	0.222
	DS+google	0.337	0.430	0.277	0.311	0.455	0.236

Future

Solving high correlation false positives



If Jane actually born in USA, but spend most of her time in USA. USA and Jane would be an ideal entity pair for google, because the relation between them is very simple.

Seeking

Internship related to knowledge graph, active learning, NLP, Information retrieval, machine learning, crowdsourcing, etc