Data Science Talk II:

Clustering and *Visualization* (Demos)

Textual Pattern Mining for Attribute Discovery (Tutorial)

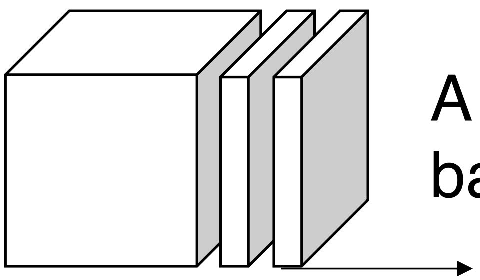
Meng Jiang
University of Notre Dame

http://www.meng-jiang.com

Clustering and Visualization

- Heterogeneous Information Network: http://www.meng-jiang.com/demos/hindblp/

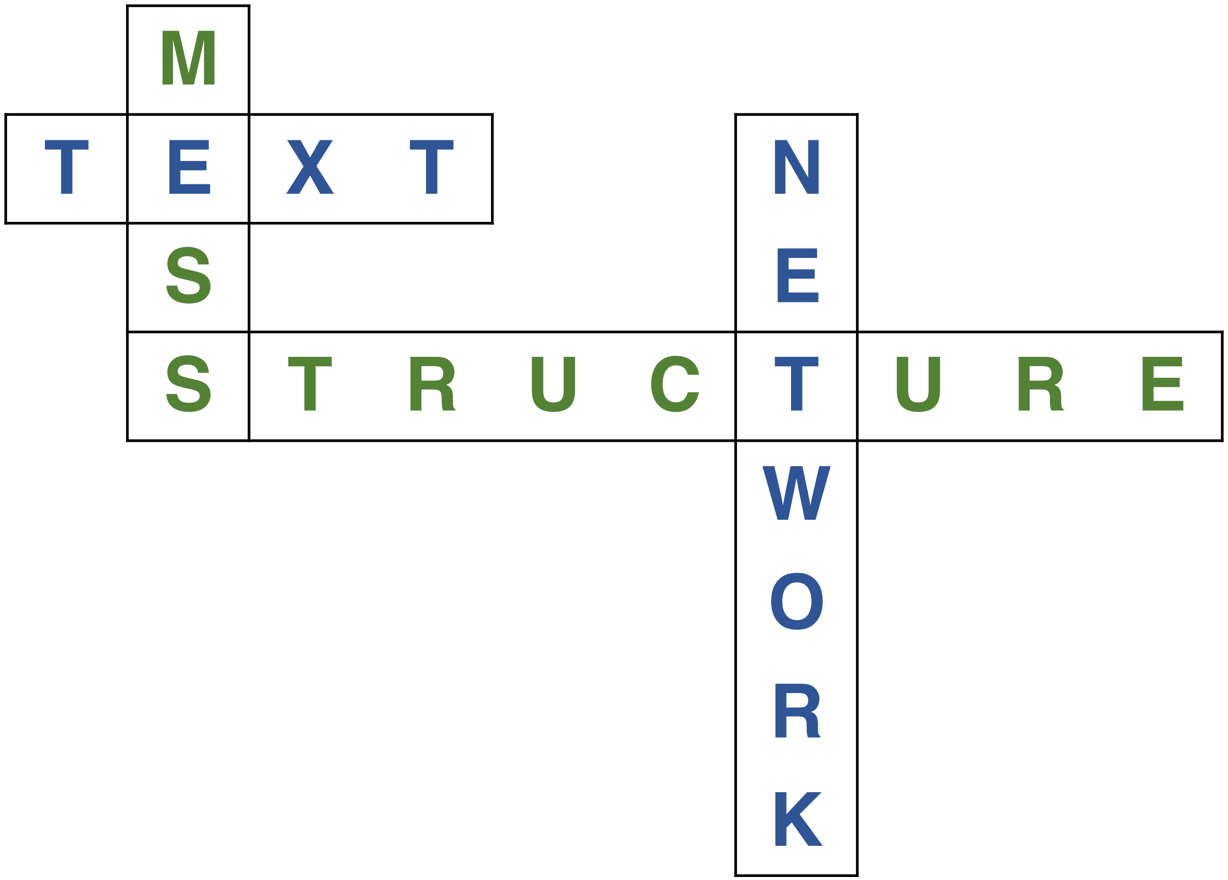
- Multi-dimensional data clustering: “who-where-what”?



A Multi-dimensional Clustering Framework based on Tensor Factorization

- Demo: http://www.meng-jiang.com/demos/fema/mas/
- Demo: http://www.meng-jiang.com/demos/fema/weibo/

Motivation:
Structuring Text into Networks



Attributes

Node attributes and link attributes

Given a sentence "*President Blaise Compaoré's government of **Burkina Faso** was founded…*", …

```
was,13897
…
president,2769
…
government,1886
…
blaise,42
…
compaore,15
…
```

vs

age: 65

Blaise Compaoré:
$PERSON.POLITICIAN

president

Burkina Faso: $COUNTRY

population: 17 million

Attribute Discovery

Given text corpora (news, tweets, paper text, etc.), **find**

1. ⟨entity, attribute name, attribute value⟩: *instance-level*

Ex. ⟨Burkina Faso, president, Blaise Compaoré⟩

⟨Burkina Faso, population, 17 million⟩

⟨Blaise Compaoré, age, 65⟩

2. ⟨entity type, attribute name⟩: *type-level*

Ex. ⟨$COUNTRY, president⟩

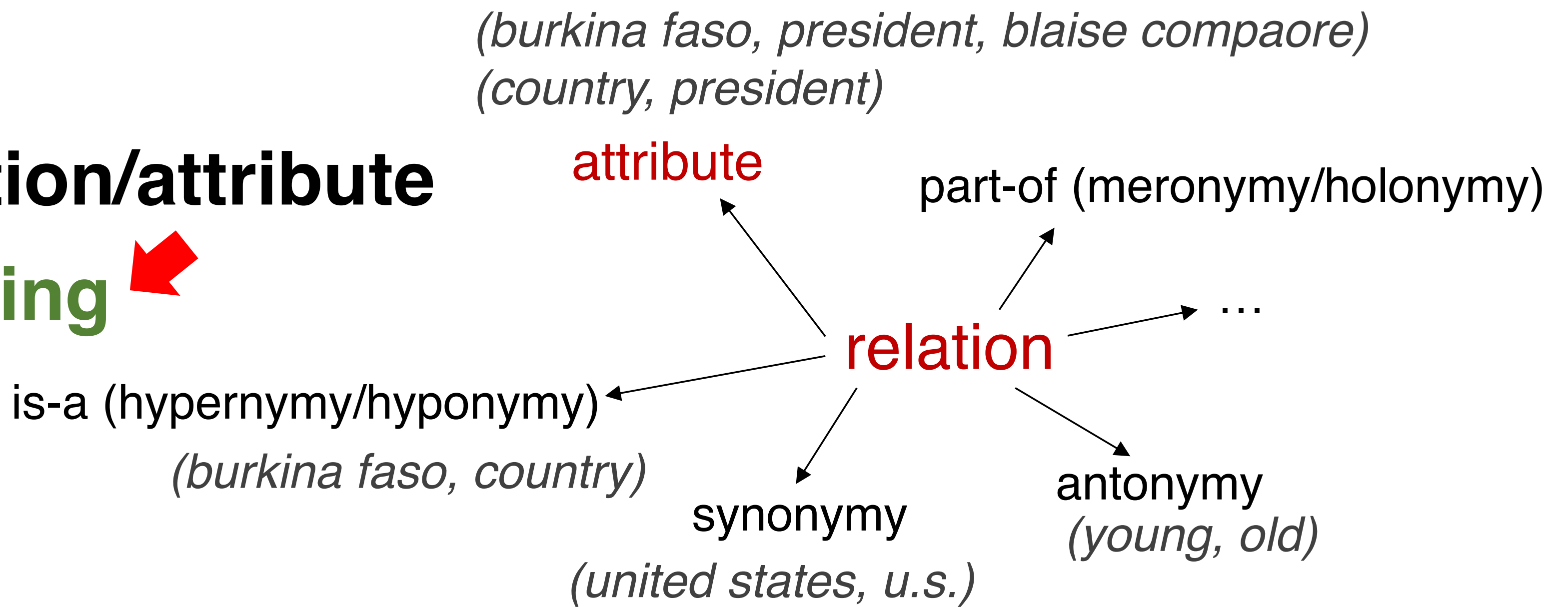
⟨$LOCATION, population⟩

⟨$PERSON, age⟩

Approaches

Relation/attribute

learning



Textual **pattern mining** and **bootstrapping**

- Hearst patterns
- Patterns in open-domain information extraction
- Textual patterns with semantic types

(Supervised) Relation/Attribute Learning

PRO: Pretty much any machine learning algorithm can work.

CONs: (Needed)

An inventory of possible semantic relations/attributes;

Annotated positive/negative examples:

for training, tuning, and evaluation.

Algorithms:

1. Classification with kernels
2. Sequential labeling methods
3. Neural networks

Relation/Attribute Learning
- Classification with kernels -

Task: Classification (entity, value) → attribute name

Kernelizable classifiers: Support Vector Machines (SVM), logistic regression, kNN, Naïve Bayes, etc.

Kernels for linguistic structures

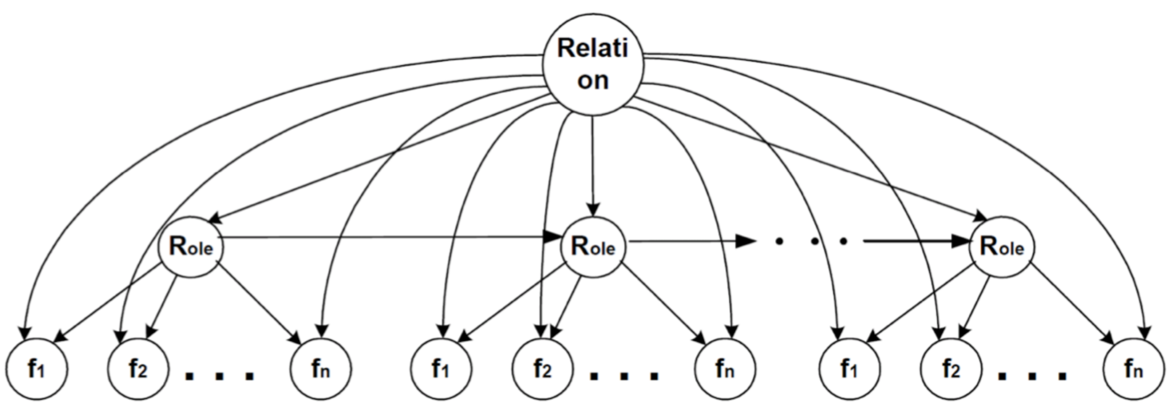
- string sequencies [Cancedda et al., 2003]
- dependency paths [Bunescu & Mooney, 2005]
- shallow parse trees [Zelenko et al., 2003]
- constituent parse trees [Collins & Duffy, 2001]
- dependency parse trees [Moschitti, 2006]
- directed acyclic graphs [Suzuki et al. 2003]

Relation/Attribute Learning
- Sequential labeling methods -

Task: Augment identification, when one augment is known (entity, attribute name) → value

Hidden Markov models (HMMs), Maximum-entropy Markov models (MEMMs), Conditional Random Fields (CRFs) [Bikel et al., 1999; Lafferty et al., 2001; McCallum & Li, 2003; Culotta et al., 2006; Bundschus et al., 2008]

Dynamic graphical models [Rosario & Hearst, 2004]



Relation/Attribute Learning
- Neural networks -

Task: Prediction by representing contexts

Recursive networks: Create a **bottom-up representation for a tree context** by recursively combining representations of siblings [Socher et al., 2012].

Convolutional networks: Create a representation by **sliding a window over the context** and **pooling the representations** at each step [Zeng et al., 2014; Liu et al., 2015; dos Santos et al., 2015].

Recurrent networks: Create a representation for a **sequence context** by processing each item in the sequence and updating the representation at each step [Li et al., 2015].

Approaches

Relation/attribute learning

(burkina faso, president, blaise compaore)
(country, president)

attribute

part-of (meronymy/holonymy)

relation

…

is-a (hypernymy/hyponymy)

(burkina faso, country)

synonymy

(united states, u.s.)

antonymy

(young, old)

Textual pattern mining and bootstrapping

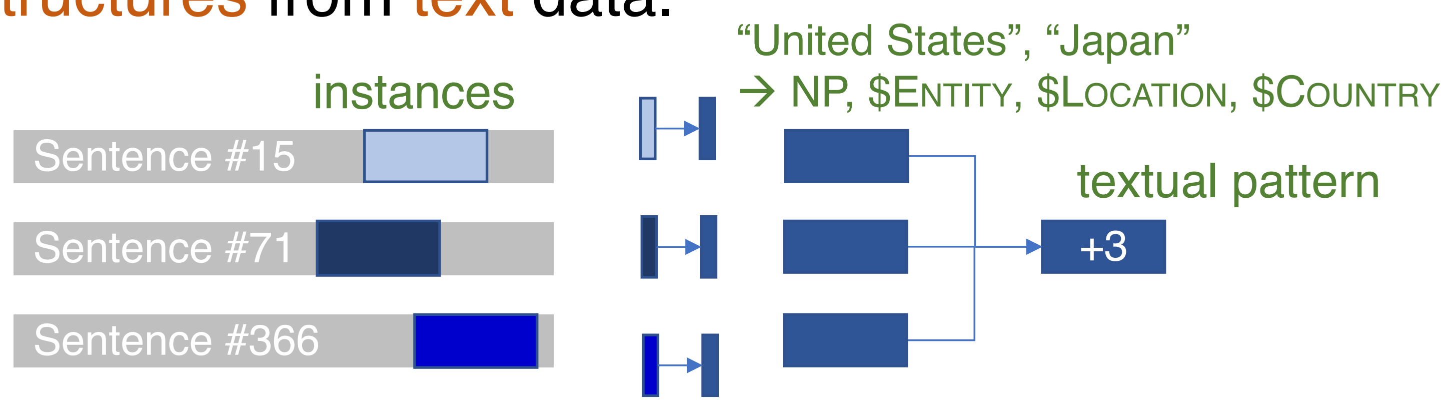
- Hearst patterns
- Patterns in open-domain information extraction
- Textual patterns with semantic types

Text Mining and Textual Pattern Mining

Text mining is the process of deriving high-quality information from text. High-quality information is typically derived through the devising of patterns and trends …

Frequent patterns are itemsets, subsequences, or substructures that appear in a data set with frequency no less than a user-specified threshold. [Wikipedia]

Definition (Textual pattern mining). Mining frequent substructures from text data.



Hearst's Lexico-Syntactic Patterns (1992)

NP such as {NP,}* {(or l and} NP

such NP as {NP,}* {(or l and)} NP

NP {, NP}* (,} (or l and) other NP

NP {,} (including l especially) {NP,}* (or l and) NP …

PRO: Designed for very high precision.

CONs: But low recall. Only cover "is-a", later, extended to "part-of" relation – more like "typing". Unclear if such patterns can be signed for all relations/attributes.

How to increase recall?

Bootstrapping

Initialization:

few seed examples, e.g., for "is-a",

for "is-a", cat-animal, banana-fruit …

for "organization-location_of_headquarters",

microsoft-redmond, boeing-seattle …

Expansion:

new patterns

new instances

Dual Iterative Pattern Expansion (DIPRE):

[Brin, 1998]

$STRING1's headquarters in $STRING2
$STRING2-based $STRING1
$STRING1, $STRING2

Several iterations

CON: Semantic drift. "pattern-based method", "turn-based strategy"

Tackling Semantic Drift using **Semantic Types**

The **Snowball** System [Agichtein & Gravano, 2000]

$STRING1's headquarters in $STRING2 $STRING2-based $STRING1 $STRING1, $STRING2	→	$ORGANIZATION's headquarters in $LOCATION $LOCATION-based $ORGANIZATION $ORGANIZATION, $LOCATION

Never-**E**nding **L**anguage **L**earner (**NELL**) [Mohamed et al. 2011]

- Specify relations/attributes, e.g.,

  country:president → $COUNTRY × $POLITICIAN

  organization:headquarters → $ORGANIZATION × $LOCATION

- Start with seed examples
- Learn: new entities, new instances, novel relations

Approach: bootstrapping + coupled learning

called *continuous open-domain information extraction*

Machine Reading at University of Washington

- **KnowItAll** [Etzioni et al., 2005] – bootstrapping using Hearst patterns
- **TextRunner** [Banko et al., 2007] – self-supervised, specific relation models from a small corpus, applied to a large corpus
- **Kylin** [Wu & Weld, 2007] and **WPE** [Hoffmann et al., 2010] – bootstrapping starting with Wikipedia infoboxes and associated articles
- **WOE** [Wu & Weld, 2010] – extends **Kylin** to open information extraction, using part-of-speech or dependency patterns
- **ReVerb** [Fader et al., 2011] – lexical and syntactic constraints on potential relation expressions

ReVerb

Problems in TextRunner:

Incoherent *verb* relation

The guide *contains* dead links and *omits* sites.
The Mark 14 *was central* to the *torpedo* scandal of the fleet.
They *recalled* that Nungesser *began* his career as a precinct leader.

Uninformative *verb* relation

Faust made a deal with the devil. (Faust, made, a deal) vs (Faust, made a deal with, the devil)
Burkina Faso has a population of 17 million. (Burkina Faso, has, a population) vs (Burkina Faso, has a population of, 17 million)
Vince gave a talk at the 2016 MSA Meeting. (Vince, gave, a talk) vs (Vince, gave a talk at, the 2016 MSA Meeting)

Scope: Open IE of relations centered around verbs

ReVerb

Preprocessing: POS tagging, NP chunking

Model: Fixed syntactic pattern + classifier

Pattern: $NP1 … $VP … $NP2

$VP satisfies syntactic constraint : V | VP | VW*P

V = verb particle? adv?

W = (noun | adj. | adv. | pron. | det)

P = (prep. | particle | inf. marker)

Learning Algorithm:

- Use *logistic regression* to assign a confidence to each triple:
  - Classifier trained manually labeled triples from 1,000 sentences
- Trade precision for recall using a confidence threshold

Machine Reading at University of Washington

- **KnowItAll** [Etzioni et al., 2005] – bootstrapping using Hearst patterns
- **TextRunner** [Banko et al., 2007] – self-supervised, specific relation models from a small corpus, applied to a large corpus
- **Kylin** [Wu & Weld, 2007] and **WPE** [Hoffmann et al., 2010] – bootstrapping starting with Wikipedia infoboxes and associated articles
- **WOE** [Wu & Weld, 2010] – extends **Kylin** to open information extraction, using part-of-speech or dependency patterns
- **ReVerb** [Fader et al., 2011] – lexical and syntactic constraints on potential relation expressions
- **OLLIE** [Mausam et al., 2012] – extends **WOE** with better patterns and dependencies (e.g., some relations are true for some period of time, or are contingent upon external conditions)

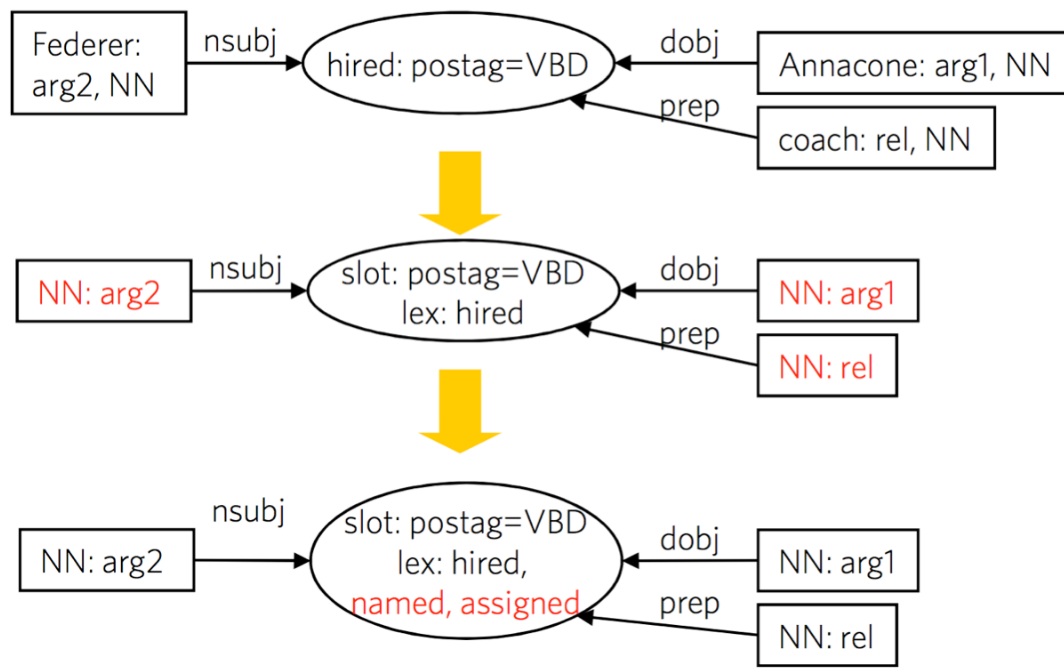
Open Language Learning IE

Input: Seed triples like (Annacone, is the coach of, Federer)

"Paraphrase" of seed triple: Sentence contains content words linked by a linear dependency path

Model: Generalize "paraphrases" + classifier

Logistic regression classifier: identify other likely triples: Trained on manually labeled triples from 1,000 sentences



Clause-based Open IE

ClausIE [Del Corro and Gemulla, 2013] Task: Separate the identification of information from its representation

Identifies essential and optional arguments in a clause

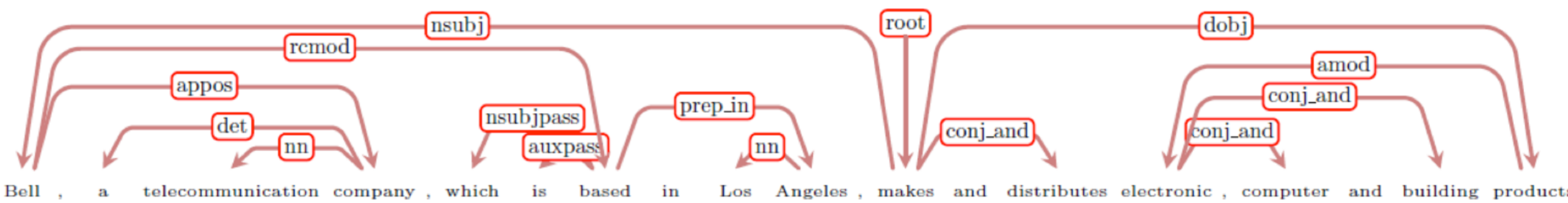
7 essential clauses: **SV, SVA, SVO, SVC, SVOO, SVOA, SVOC**

A *minimal clause* is a clause without the optional adv. (A)

Algorithm

- Clause identification: Walk the dependency tree and identify clauses using a deterministic flow chart of decision questions
- Proposition generation: For each clause, generate one or more propositions

ClausIE Example

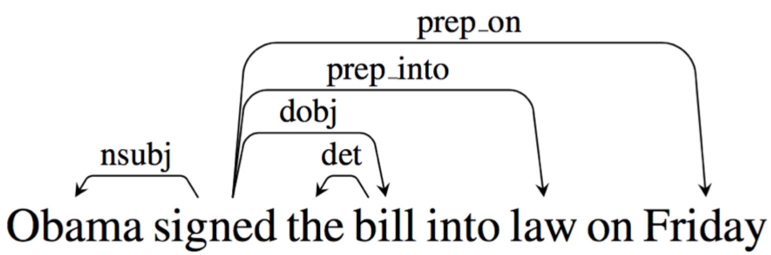


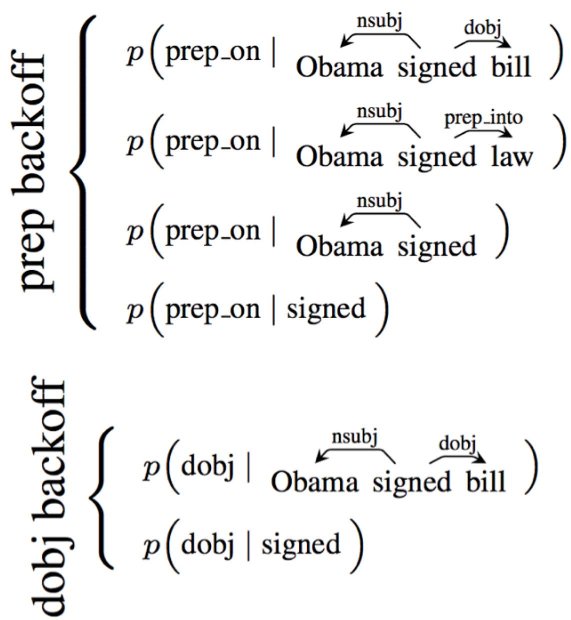
Bell, **a telecommunication company**, which is based in **Los Angeles**, makes and distributes **electronic and building products**.

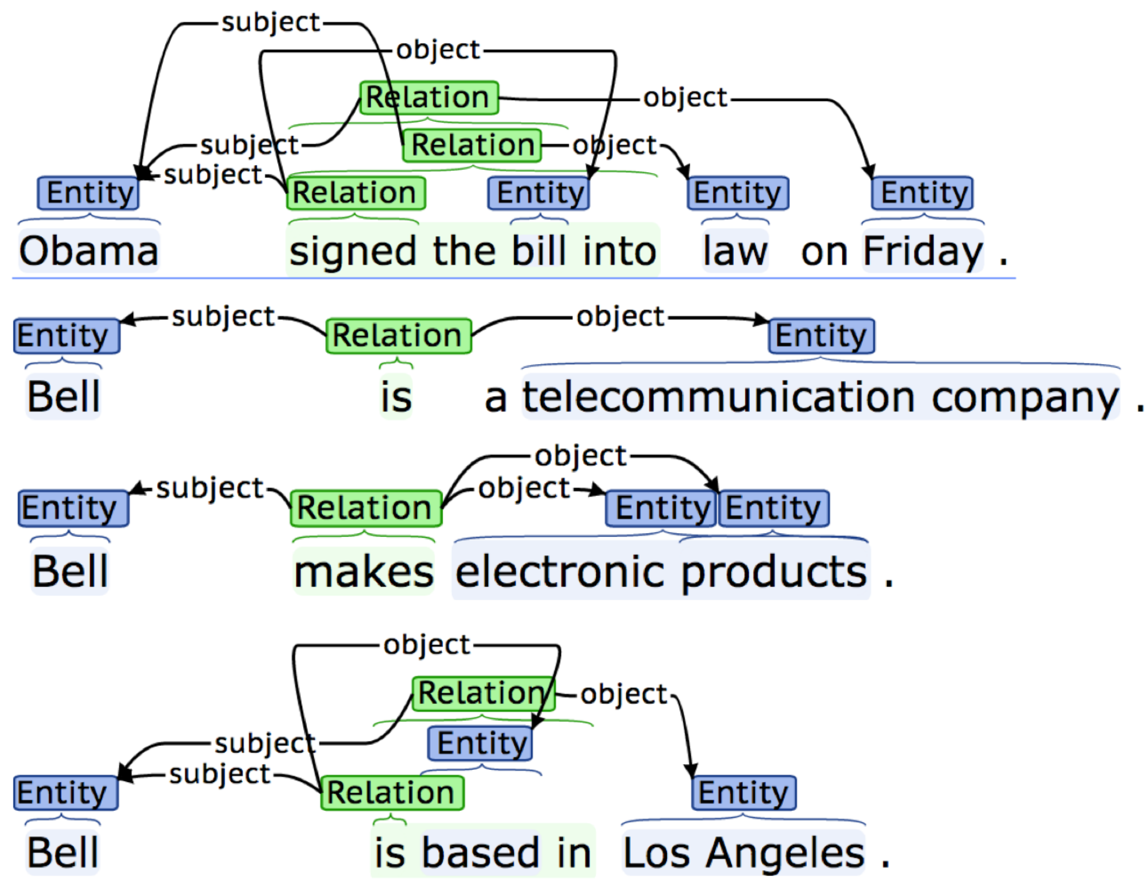
S: Bell	V: “is”	C: a telecommunication company
S: Bell	V: is based	A: in Los Angeles
S: Bell	V: makes	O: electronic products
S: Bell	V: makes	O: computer products
S: Bell	V: makes	O: building products
S: Bell	V: distributes	O: electronic products
S: Bell	V: distributes	O: computer products
S: Bell	V: distributes	O: building products

Clause-based Open IE
(Inter-Clause and Intra-Clause)

Stanford's Open IE [Angeli et al., 2015] http://corenlp.run/

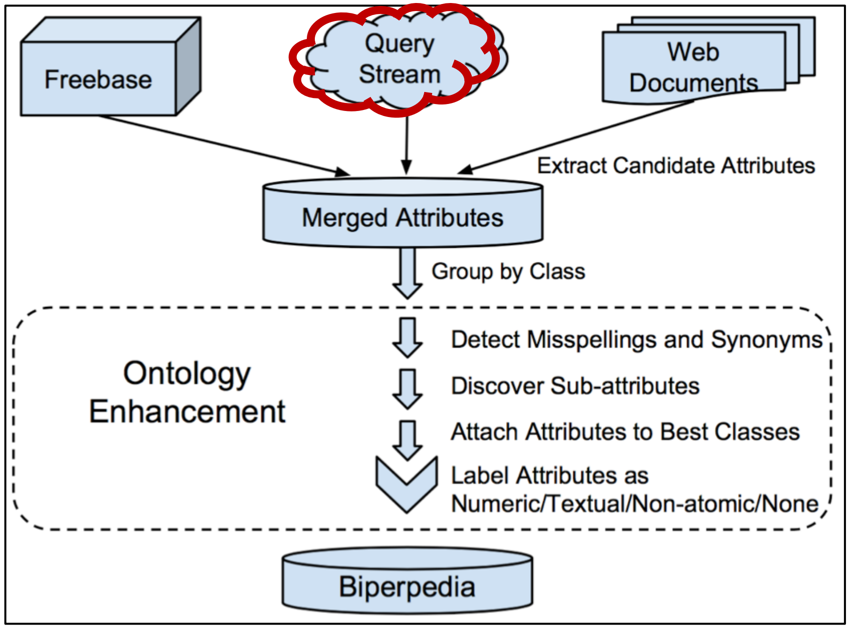






Entity-based Textual Patterns: Google's Advantages

Biperpedia [Gupta et al. 2014]: Pipeline and **E-A patterns**



Pattern	Example
E A	[Google] (CEO) Larry Page
A , E	Larry Page, (CEO), [Google]
(E = his) A	[his] (wife)
E 's A	[Google] 's (CEO)
A of E	(CEO) of [Google]
A in E	(urban population) in [Kingston]
A of the E	(captain) of the [Australian cricket team]
(E = its) A	[its] (fire-chief)
A at E	(CEO) at [Google]
A for E	(spokesman) for [gun control]

PRO: 36 billion anonymized unique queries

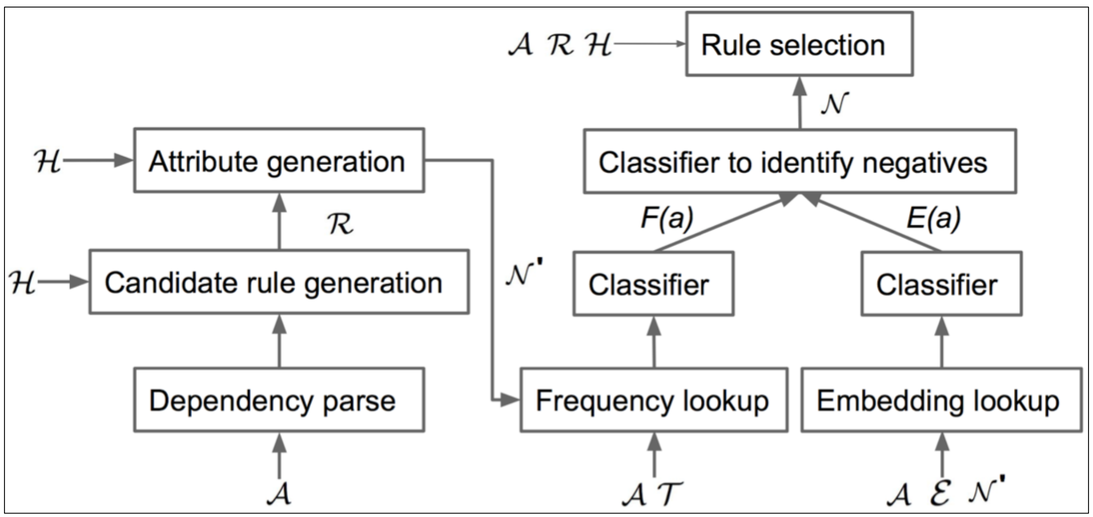
Ontology with 1.6M (CLASS, ATTRIBUTE) pairs and 67K attribute names

CONs: (query log) Highly constrained and unavailable in academia

Entity-based Textual Patterns: Google's Advantages

ARI [Halevy et al. 2016]: Discover structure in attribute names

Learning the attribute grammar

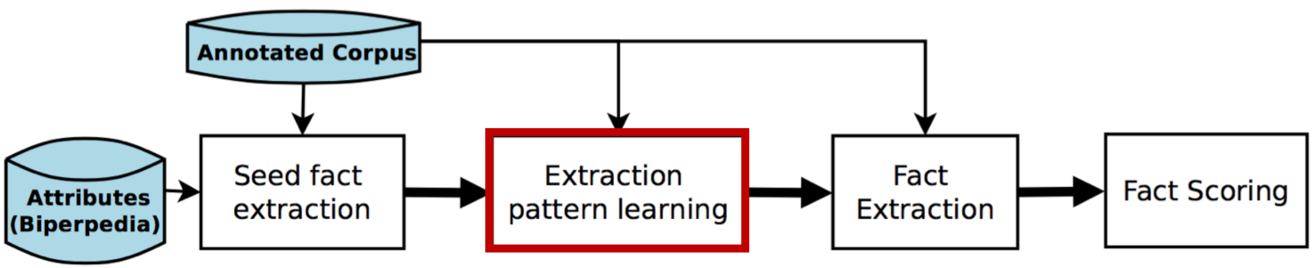


$SportsCars
tyre price in Singapore

$Attribute_price	::=	price
$Attribute_price	::=	$Nation $Attribute_price
$Attribute_price	::=	$Attribute_price in $Nation
$Attribute_price	::=	$Product $Attribute_price
$Nation	::=	Singapore | USA | UAE | UK | . . .
$Product	::=	battery | insurance | kit | door | . . .

Entity-based Textual Patterns: Google's Advantages

ReNoun [Yahya et al. 2014]: Pipeline and **S-A-O patterns**



1. *the* A *of* S, O – the CEO of Google, Larry Page
2. *the* A *of* S *is* O – the CEO of Google is Larry Page
3. O, S A – Larry Page, Google CEO
4. O, S*'s* A – Larry Page, Google's CEO
5. O, [*the*] A *of* S – Larry Page, [the] CEO of Google
6. SAO – Google CEO Larry Page
7. S A, O – Google CEO, Larry Page
8. S*'s* A, O – Google's CEO, Larry Page

PROs: 8 manually crafted high-precision patterns to find seed triples in corpus

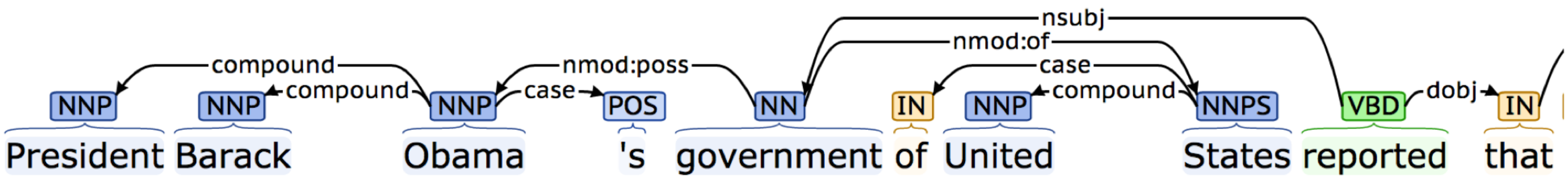
680K unique facts
400M news docs

CONs: (annotated corpus) domain-limited and expensive

Syntactic-**O**ntological-**L**exical Patterns with **Semantic Types**

PATTY [Nakashole et al. 2012]

Definition. An SOL pattern is the shortest path between two entities in the dependency parse tree.



poss (“government”, “Barack Obama”)

nmod:of (“government”, “United States”)

Output: $POLICITIAN government [of] $COUNTRY

PRO: Harnessing typing information (O) from a typing system

CONs: Relying on Stanford’s dependency parsers (S & L).

Losing pattern contexts. Lacking pattern organization.

Reminder: Attribute Discovery

Given text corpora (news, tweets, paper text, etc.), **find**

1. ⟨entity, attribute name, attribute value⟩: *instance-level*

Ex. ⟨Burkina Faso, president, Blaise Compaoré⟩

⟨Burkina Faso, population, 17 million⟩

⟨Blaise Compaoré, age, 65⟩

2. ⟨entity type, attribute name⟩: *type-level*

Ex. ⟨$COUNTRY, president⟩

⟨$LOCATION, population⟩

⟨$PERSON, age⟩

Organizing **Context-aware** Textual Patterns into **Synonymous Groups**

"President Blaise Compaoré's government of Burkina Faso …"

→ ⟨$COUNTRY, president⟩, ⟨Burkina Faso, president, Blaise Compaoré⟩

$COUNTRY President $POLITICIAN
$COUNTRY's president $POLITICIAN
President $POLITICIAN of $COUNTRY
…
President $POLITICIAN's government of $COUNTRY

A new textual pattern: **Meta Pattern**
and a **synonymous** meta pattern group
MetaPAD [Jiang et al. Submitted to KDD 2017]

The **MetaPAD** Framework: **Meta PA**tten **D**iscovery from *Massive* Text Corpora

(#1) "President Blaise Compaoré's government of Burkina Faso was founded …"

⟨$COUNTRY, {president}, $POLITICIAN⟩

⟨Burkina Faso, {president}, Blaise Compaoré⟩

The **MetaPAD** Framework:
Meta PAtten **D**iscovery from *Massive* Text Corpora

(#1) “President Blaise Compaoré’s government of Burkina Faso was founded …”
(#2) “President Barack Obama’s government of U.S. claimed that…”

Meta patterns:

⌈president \$Person.Politician ’s government of \$Location.Country⌋ was founded…

Generate patterns with <u>massive</u> instances in the data

⟨\$Country, {president}, \$Politician⟩

frequency↑

⟨U.S., {president}, Barack Obama⟩

The **MetaPAD** Framework: **Meta PA**tten **D**iscovery from *Massive* Text Corpora

(#1) "President Blaise Compaoré's government of Burkina Faso was founded …"
(#2) "President Barack Obama's government of U.S. claimed that…"

Meta patterns:

⌈president $Person.Politician 's government of $Location.Country⌋ was founded…

⟨$Country, {president}, $Politician⟩

Generate <u>massive</u> triples by matching the meta patterns

⟨Burkina Faso, {president}, Blaise Compaoré⟩
⟨U.S., {president}, Barack Obama⟩

The **MetaPAD** Framework: **Meta PA**tten **D**iscovery from *Massive* Text Corpora

(#1) "President Blaise Compaoré's government of Burkina Faso was founded …"
(#2) "President Barack Obama's government of U.S. claimed that…"
(#3) "U.S. President Barack Obama visited …"

Meta patterns:

⌈president $PERSON.POLITICIAN 's government of $LOCATION.COUNTRY⌋ was founded…
⌈$LOCATION.COUNTRY president $PERSON.POLITICIAN⌋ …

⟨$COUNTRY, {president}, $POLITICIAN⟩

frequency↑↑

Group synonymous patterns by <u>massive</u> triples

⟨Burkina Faso, {president}, Blaise Compaoré⟩
⟨U.S., {president}, Barack Obama⟩

The **MetaPAD** Framework: **Meta PA**tten **D**iscovery from *Massive* Text Corpora

(#1) “President Blaise Compaoré’s government of Burkina Faso was founded …”
(#2) “President Barack Obama’s government of U.S. claimed that…”
(#3) “U.S. President Barack Obama visited …”

Meta patterns:

⌈president $PERSON.POLITICIAN ’s government of $LOCATION.COUNTRY⌋ was founded…

⌈$LOCATION.COUNTRY president $PERSON.POLITICIAN⌋ …

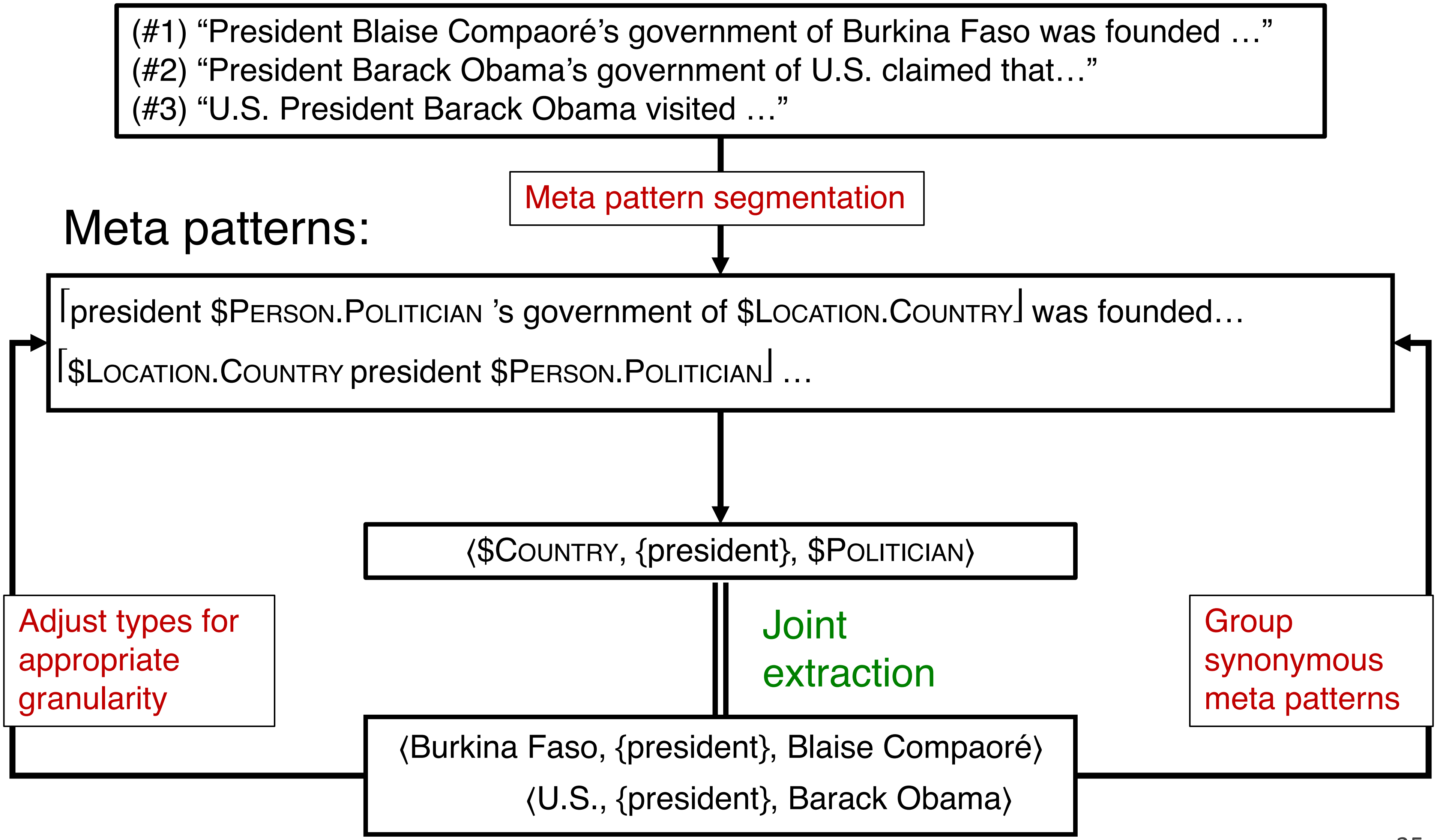
⟨$COUNTRY, {president}, $POLITICIAN⟩

Adjust entity types in meta patterns for appropriate granularity with triples

⟨Burkina Faso, {president}, Blaise Compaoré⟩

⟨U.S., {president}, Barack Obama⟩

The **MetaPAD** Framework: **Meta PA**tten **D**iscovery from *Massive* Text Corpora



The **MetaPAD** Framework: **Meta PA**tten **D**iscovery from *Massive* Text Corpora

(#1) “President Blaise Compaoré’s government of Burkina Faso was founded …”
(#2) “President Barack Obama’s government of U.S. claimed that…”
(#3) “U.S. President Barack Obama visited …”

Meta pattern segmentation

Meta patterns:

⌈president $PERSON.POLITICIAN ’s government of $LOCATION.COUNTRY⌋ was founded…
⌈$LOCATION.COUNTRY president $PERSON.POLITICIAN⌋ …

No heavy annotation required
No domain knowledge required
No query log required

if we can recognize and type the entities in the same manner…

Adjust ty
appropri
granularity

extraction

meta patterns

⟨Burkina Faso, {president}, Blaise Compaoré⟩
⟨U.S., {president}, Barack Obama⟩

Effort-Light Text Mining

"President Blaise Compaoré's government of Burkina Faso was founded …"

Phrase mining (SegPhrase Liu et al. 2015)

"president blaise_compaoré 's government of burkina_faso was founded …"

Entity recognition and typing with Distant Supervision
(ClusType Ren et al. 2015)

"president \$PERSON 's government of \$LOCATION was founded …"

Fine-grained typing (PLE Ren et al. 2016)

"president \$PERSON.POLITICIAN 's government of \$LOCATION.COUNTRY was founded …"

Meta-Pattern Quality Assessment and Segmentation

A rich set of features:

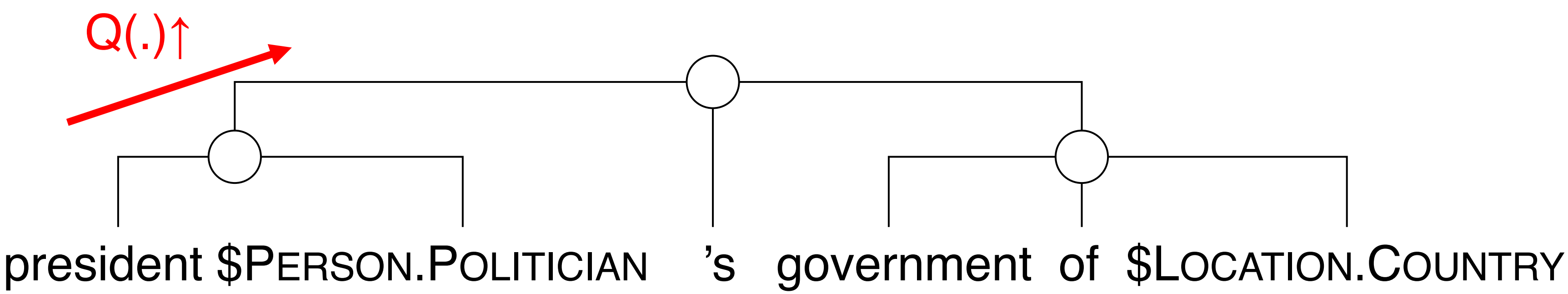
✓ **Frequency**

✓ **Concordance:** “$PERSON ’s wife”

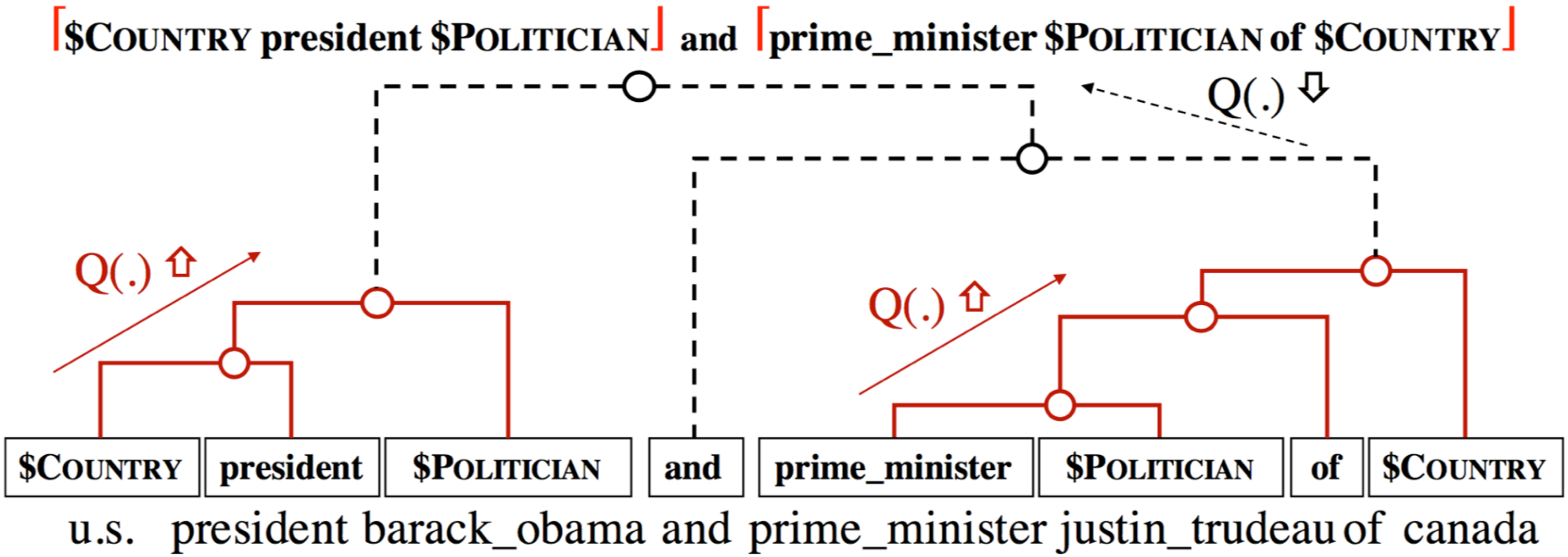
✓ **Completeness:** “$COUNTRY president” vs “$COUNTRY president $POLITICIAN”

✓ **Informativeness:** “$PERSON and $PERSON ” vs “$PERSON ’s wife, $PERSON”

Regression Q(.): *random forest* with only 300 labels

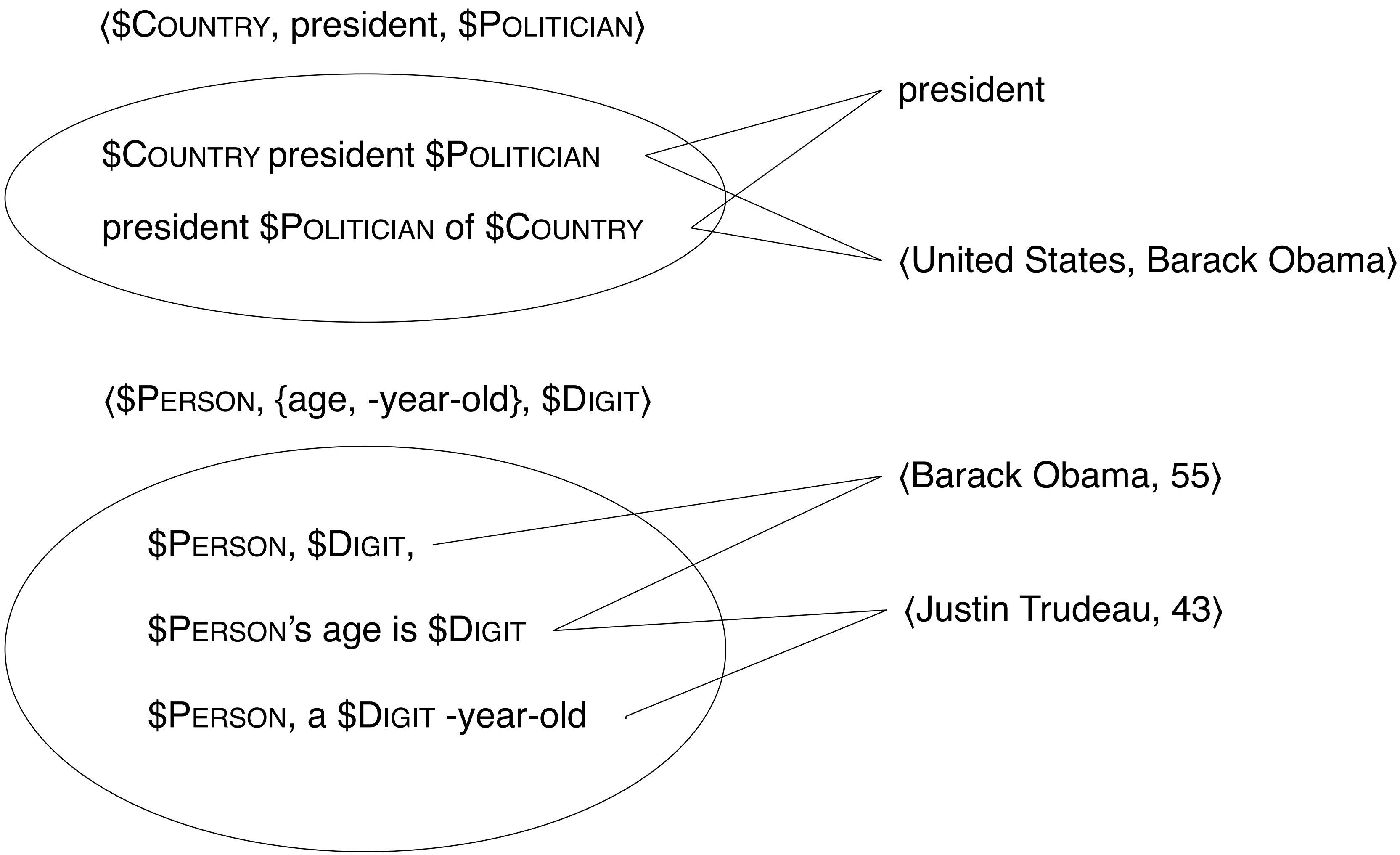


Why Segmentation?



Pattern candidate	Before segmentation Count	Before segmentation Quality	Frequency rectified after segmentation Count	Frequency rectified after segmentation Quality	Issue fixed by feedback
$COUNTRY president $POLITICIAN	2,912	0.93	2,785	0.97	N/A
prime_minister $POLITICIAN of $COUNTRY	1,285	0.84	1,223	0.92	slight underestimation
$POLITICIAN and prime_minister $POLITICIAN	532	0.70	94	0.23	overestimation

Grouping Synonymous Patterns

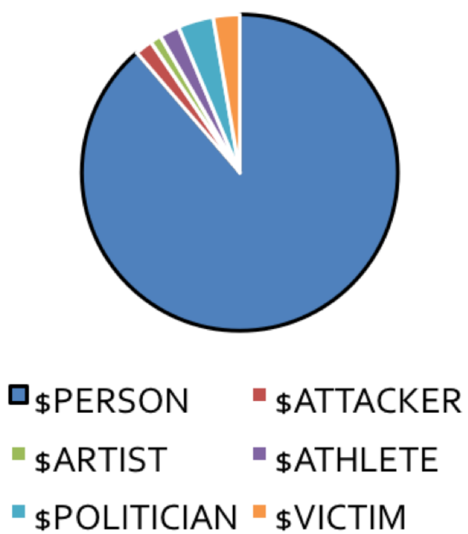


Adjusting Types in Meta Patterns for Appropriate Granularity

\$PERSON, \$DIGIT,

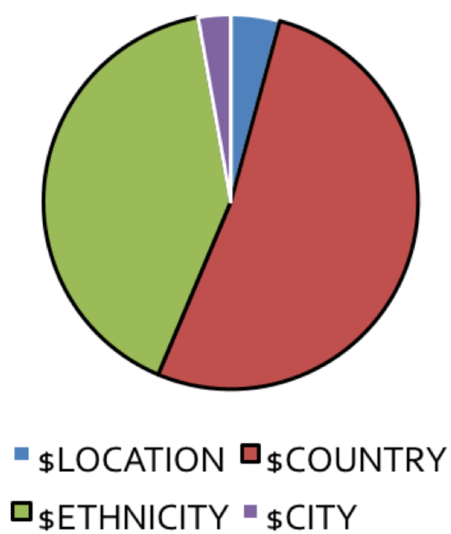
\$PERSON's age is \$DIGIT

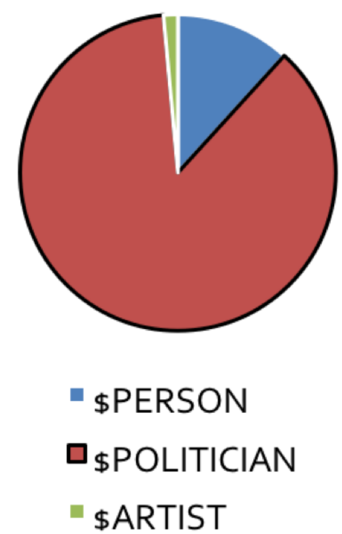
\$PERSON, a \$DIGIT -year-old



\$COUNTRY president \$POLITICIAN

president \$POLITICIAN of \$COUNTRY





Results in General Domain

Meta patterns	Entity	Attribute value
$COUNTRY President $POLITICIAN $COUNTRY's president $POLITICIAN President $POLITICIAN of $COUNTRY … President $POLITICIAN's government of $COUNTRY	United States	Barack Obama
	Russia	Vladimir Putin
	France	Francois Hollande
	…	…
	Burkina Faso	Blaise Compaoré

Meta patterns	Entity	Attribute value
$COMPANY CEO $PERSON $COMPANY chief executive $PERSON $PERSON, the $COMPANY CEO, … $COMPANY former CEO $PERSON $PERSON, the $COMPANY former CEO,	Apple	Tim Cook
	Facebook	Mark Zuckerberg
	Hewlett-Packard	Carly Fiorina
	…	…
	Infor	Charles Phillips
	Afghan Citadel	Roya Mahboob

Results in Biomedical Domain

Meta patterns	Entity	Attribute value
$TREATMENT was used to treat $DISEASE $DISEASE using the $TREATMENT $TREATMENT has been used to treat $DISEASE $TREATMENT of patients with $DISEASE …	zoledronic acid therapy	Paget's disease of bone
	bisphosphonates	osteoporosis
	calcitonin	Paget's disease of bone
	calcitonin	osteoporosis
	…	…

Meta patterns	Entity	Attribute value
$BACTERIA was resistant to $ANTIBIOTICS $BACTERIA are resistant to $ANTIBIOTICS $BACTERIA is the most resistant to $ANTIBIOTICS … $BACTERIA, particularly those resistant to $ANTIBIOTICS	corynebacterium striatum BM4687	gentamicin
	corynebacterium striatum BM4687	tobramycin
	methicillin-susceptible S aureus	vancomycin
	multidrug-resistant enterobacteriaceae	gentamicin
	…	…

Future Directions

- Combining complementary methods towards attribute discovery from massive text corpora
  - Learning Approaches
  - Linguistic patterns using POS tagging, NP chunking, clause analysis, dependency parsing …
  - Meta pattern-driven approaches
    - Harnessing entity recognition and (fine-grained) typing systems
    - Quality assessment and meta-pattern segmentation based on contexts
    - Grouping synonymous patterns
    - Adjusting type levels for appropriate granularity

References

Cancedda, N., Gaussier, E., Goutte, C. and Renders, J.M., 2003. Word-sequence kernels. Journal of machine learning research, 3(Feb), pp.1059-1082.

Bunescu, R.C. and Mooney, R.J., 2005, October. A shortest path dependency kernel for relation extraction. In Proceedings of the conference on human language technology and empirical methods in natural language processing (pp. 724-731). Association for Computational Linguistics.

Zelenko, D., Aone, C. and Richardella, A., 2003. Kernel methods for relation extraction. Journal of machine learning research, 3(Feb), pp.1083-1106.

Collins, M. and Duffy, N., 2001, December. Convolution kernels for natural language. In NIPS (Vol. 14, pp. 625-632).

Moschitti, A., 2006, September. Efficient convolution kernels for dependency and constituent syntactic trees. In European Conference on Machine Learning (pp. 318-329). Springer Berlin Heidelberg.

Suzuki, J., Hirao, T., Sasaki, Y. and Maeda, E., 2003, July. Hierarchical directed acyclic graph kernel: Methods for structured natural language data. In Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1 (pp. 32-39). Association for Computational Linguistics.

Bikel, D.M., Schwartz, R. and Weischedel, R.M., 1999. An algorithm that learns what's in a name. Machine learning, 34(1-3), pp.211-231.

Lafferty, J., McCallum, A. and Pereira, F., 2001, June. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In Proceedings of the eighteenth international conference on machine learning, ICML (Vol. 1, pp. 282-289).

McCallum, A. and Li, W., 2003, May. Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons. In Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003-Volume 4 (pp. 188-191). Association for Computational Linguistics.

Culotta, A., Wick, M., Hall, R. and McCallum, A., 2006. First-order probabilistic models for coreference resolution.

References

Bundschus, M., Dejori, M., Stetter, M., Tresp, V. and Kriegel, H.P., 2008. Extraction of semantic biomedical relations from text using conditional random fields. BMC bioinformatics, 9(1), p.207.

Rosario, B. and Hearst, M.A., 2004, July. Classifying semantic relations in bioscience texts. In Proceedings of the 42nd annual meeting on association for computational linguistics (p. 430). Association for Computational Linguistics.

Socher, R., Huval, B., Manning, C.D. and Ng, A.Y., 2012, July. Semantic compositionality through recursive matrix-vector spaces. In Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (pp. 1201-1211). Association for Computational Linguistics.

Zeng, D., Liu, K., Lai, S., Zhou, G. and Zhao, J., 2014, August. Relation Classification via Convolutional Deep Neural Network. In COLING (pp. 2335-2344).

Zeng, D., Liu, K., Chen, Y. and Zhao, J., 2015, September. Distant Supervision for Relation Extraction via Piecewise Convolutional Neural Networks. In EMNLP (pp. 1753-1762).

Santos, C.N.D., Xiang, B. and Zhou, B., 2015. Classifying relations by ranking with convolutional neural networks. In ACL.

Li, J. and Jurafsky, D., 2015. Do multi-sense embeddings improve natural language understanding? In ACL.

Li, J., Luong, M.T., Jurafsky, D. and Hovy, E., 2015. When are tree structures necessary for deep learning of representations? In ACL.

Hearst, M.A., 1992, August. Automatic acquisition of hyponyms from large text corpora. In Proceedings of the 14th conference on Computational linguistics-Volume 2 (pp. 539-545). Association for Computational Linguistics.

Brin, S., 1998, March. Extracting patterns and relations from the world wide web. In International Workshop on The World Wide Web and Databases (pp. 172-183). Springer Berlin Heidelberg.

Agichtein, E. and Gravano, L., 2000, June. Snowball: Extracting relations from large plain-text collections. In Proceedings of the fifth ACM conference on Digital libraries (pp. 85-94). ACM.

References

Carlson, A., Betteridge, J., Kisiel, B., Settles, B., Hruschka Jr, E.R. and Mitchell, T.M., 2010, July. Toward an Architecture for Never-Ending Language Learning. In AAAI (Vol. 5, p. 3).

Mitchell, T. and Fredkin, E., 2014, October. Never ending language learning. In Big Data (Big Data), 2014 IEEE International Conference on (pp. 1-1). IEEE.

Etzioni, O., Cafarella, M., Downey, D., Popescu, A.M., Shaked, T., Soderland, S., Weld, D.S. and Yates, A., 2005. Unsupervised named-entity extraction from the web: An experimental study. Artificial intelligence, 165(1), pp.91-134.

Yates, A., Cafarella, M., Banko, M., Etzioni, O., Broadhead, M. and Soderland, S., 2007, April. Textrunner: open information extraction on the web. In Proceedings of Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations (pp. 25-26). Association for Computational Linguistics.

Wu, F. and Weld, D.S., 2007, November. Autonomously semantifying wikipedia. In Proceedings of the sixteenth ACM conference on Conference on information and knowledge management (pp. 41-50). ACM.

Hoffman, M., Bach, F.R. and Blei, D.M., 2010. Online learning for latent dirichlet allocation. In advances in neural information processing systems (pp. 856-864).

Wu, F. and Weld, D.S., 2010, July. Open information extraction using Wikipedia. In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (pp. 118-127). Association for Computational Linguistics.

Fader, A., Soderland, S. and Etzioni, O., 2011, July. Identifying relations for open information extraction. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (pp. 1535-1545). Association for Computational Linguistics.

Schmitz, M., Bart, R., Soderland, S. and Etzioni, O., 2012, July. Open language learning for information extraction. In Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (pp. 523-534). Association for Computational Linguistics.

References

Del Corro, L. and Gemulla, R., 2013, May. Clausie: clause-based open information extraction. In Proceedings of the 22nd international conference on World Wide Web (pp. 355-366). ACM.

Angeli, G., Premkumar, M.J. and Manning, C.D., 2015, July. Leveraging linguistic structure for open domain information extraction. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (ACL 2015).

Angeli, G., Gupta, S., Premkumar, M.J., Manning, C.D., Ré, C., Tibshirani, J., Wu, J.Y., Wu, S. and Zhang, C., 2015. Stanford's distantly supervised slot filling systems for KBP 2014. In Text Analysis Conference (TAC-KBP).

Gupta, R., Halevy, A., Wang, X., Whang, S.E. and Wu, F., 2014. Biperpedia: An ontology for search applications. Proceedings of the VLDB Endowment, 7(7), pp.505-516.

Halevy, A., Noy, N., Sarawagi, S., Whang, S.E. and Yu, X., 2016, April. Discovering structure in the universe of attribute names. In Proceedings of the 25th International Conference on World Wide Web (pp. 939-949). International World Wide Web Conferences Steering Committee.

Yahya, M., Whang, S., Gupta, R. and Halevy, A.Y., 2014, October. ReNoun: Fact Extraction for Nominal Attributes. In EMNLP (pp. 325-335).

Nakashole, N., Weikum, G. and Suchanek, F., 2012, July. PATTY: a taxonomy of relational patterns with semantic types. In Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (pp. 1135-1145). Association for Computational Linguistics.

Liu, J., Shang, J., Wang, C., Ren, X. and Han, J., 2015, May. Mining quality phrases from massive text corpora. In Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data (pp. 1729-1744). ACM.

Ren, X., El-Kishky, A., Wang, C. and Han, J., 2016, April. Automatic entity recognition and typing in massive text corpora. In Proceedings of the 25th International Conference Companion on World Wide Web (pp. 1025-1028). International World Wide Web Conferences Steering Committee.

References

Ren, X., El-Kishky, A., Ji, H. and Han, J., 2016, June. Automatic Entity Recognition and Typing in Massive Text Data. In Proceedings of the 2016 International Conference on Management of Data (pp. 2235-2239). ACM.

Shang, J., Liu, J., Jiang, M., Ren, X., Voss, C.R. and Han, J., 2017. Automated Phrase Mining from Massive Text Corpora. arXiv preprint arXiv:1702.04457.

El-Kishky, A., Song, Y., Wang, C., Voss, C.R. and Han, J., 2014. Scalable topical phrase mining from text corpora. Proceedings of the VLDB Endowment, 8(3), pp.305-316.

Ren, X., El-Kishky, A., Wang, C., Tao, F., Voss, C.R. and Han, J., 2015, August. Clustype: Effective entity recognition and typing by relation phrase-based clustering. In Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 995-1004). ACM.

Jiang, M., Shang, J., Cassidy, T., Ren, X., Kaplan, L.M., Hanratty, T.P. and Han, J., 2017. MetaPAD: Meta Pattern Discovery from Massive Text Corpora. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM.