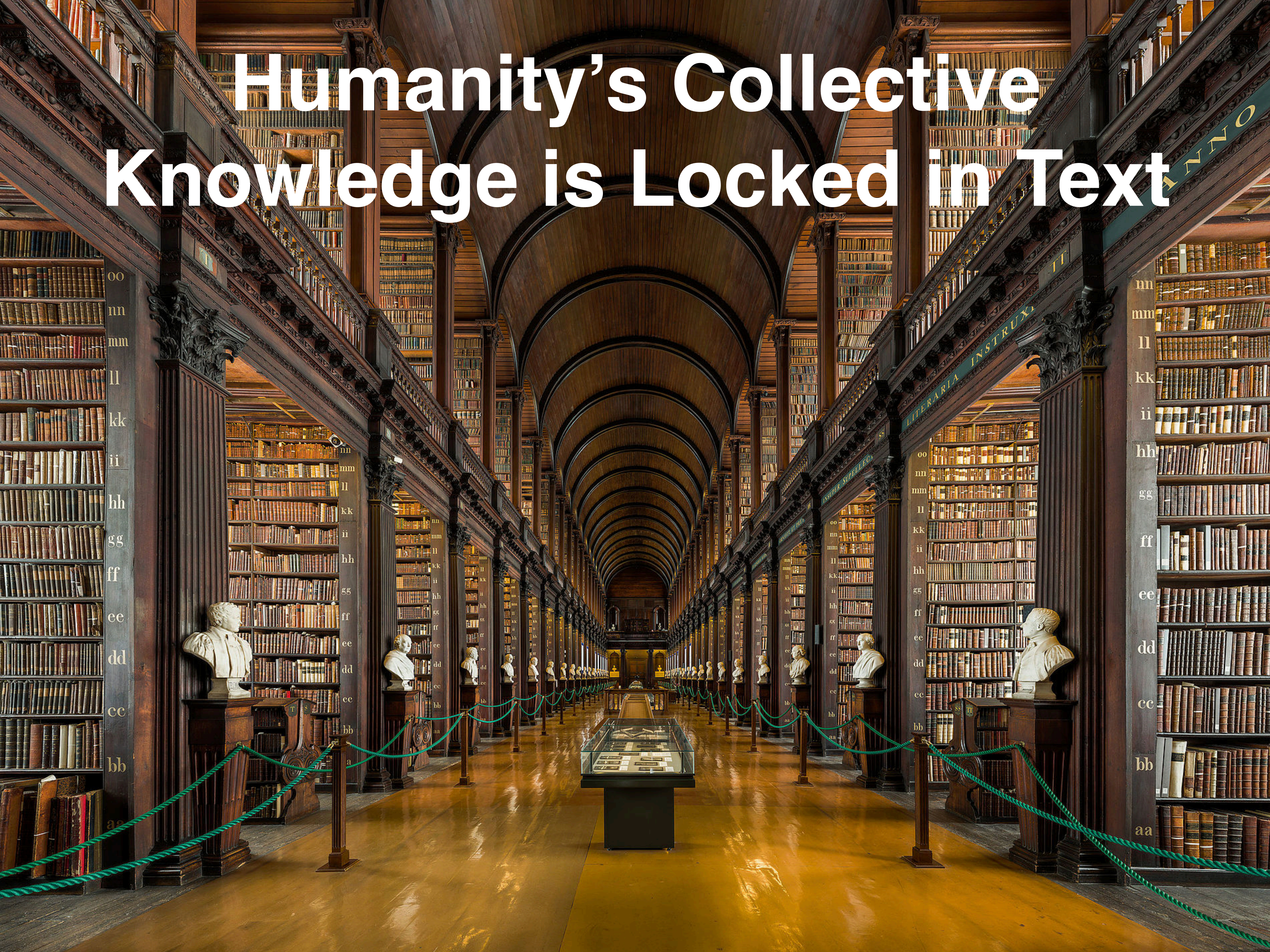
Large-Scale Learning for Information Extraction

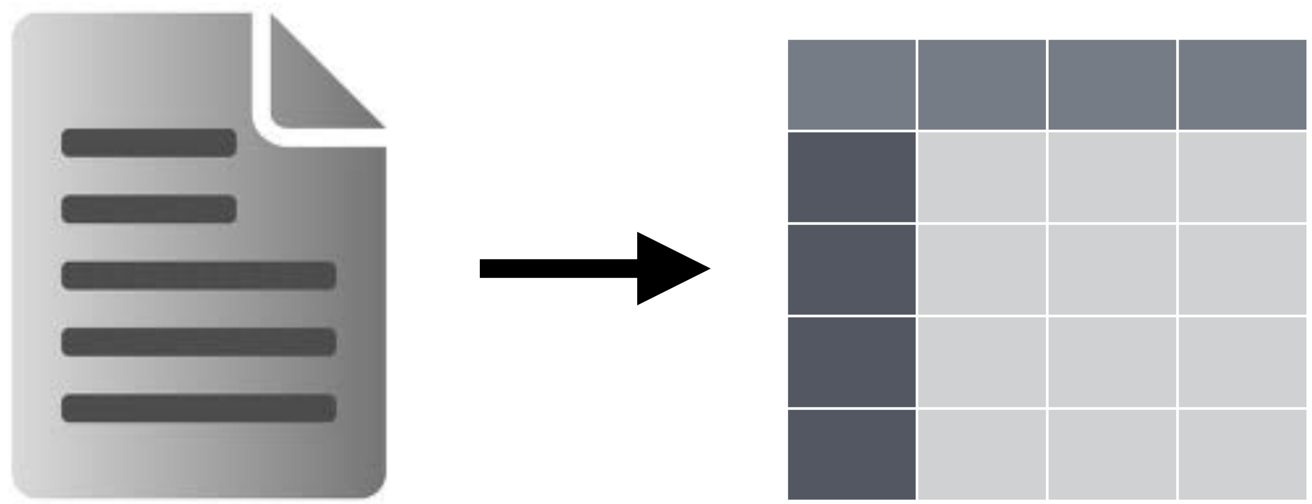
Alan Ritter
Computer Science and Engineering
Ohio State University
@alan_ritter

THE OHIO STATE UNIVERSITY

Humanity's Collective Knowledge is Locked in Text



Information Extraction



Text Structured Data

Traditional Information Extraction

1) Humans Annotate Text

Traditional Information Extraction

1) Humans Annotate Text

2) Supervised Machine Learning

$$\frac{1}{Z(w_1, \ldots, w_n, \theta)} \prod_{i=1}^{n} e^{\theta \cdot f(t_i, t_{i-1}, w_1, \ldots, w_n, i)}$$

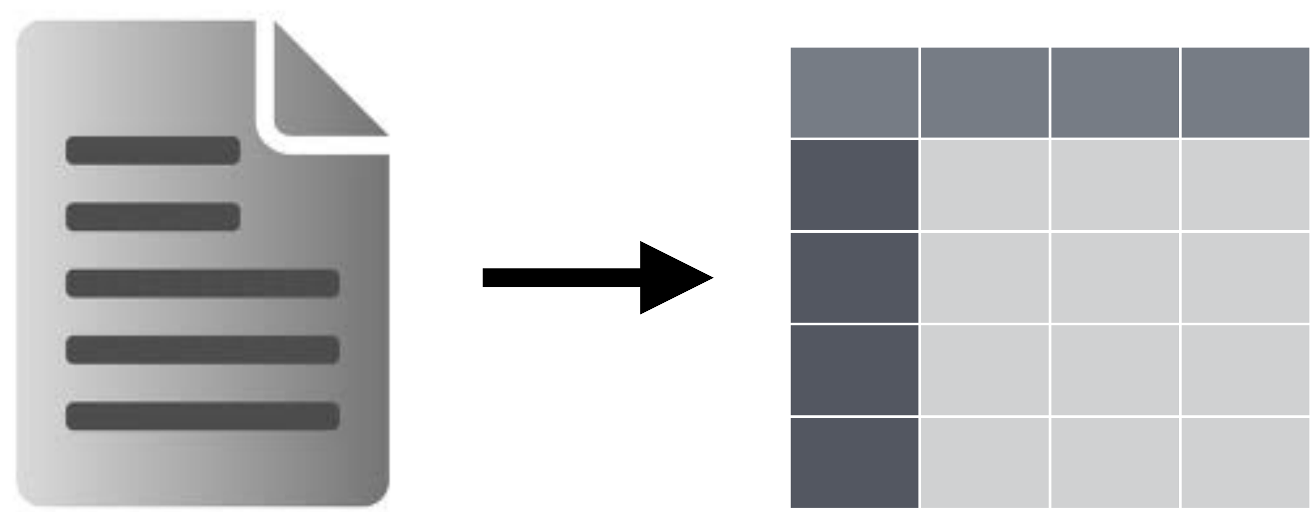
Traditional Information Extraction

1) Humans Annotate Text

2) Supervised Machine Learning

$$\frac{1}{Z(w_1, \ldots, w_n, \theta)} \prod_{i=1}^{n} e^{\theta \cdot f(t_i, t_{i-1}, w_1, \ldots, w_n, i)}$$

3) Apply Models to New Documents



Traditional Information Extraction: **Key Limitations**

DATA

Edited Text

LEARNING

Supervised Learning (small datasets)

KNOWLEDGE

Narrow Pre-specified domains

Traditional Information Extraction: **Key Limitations**

DATA

Edited Text

Newswire

LEARNING

Supervised Learning (small datasets)

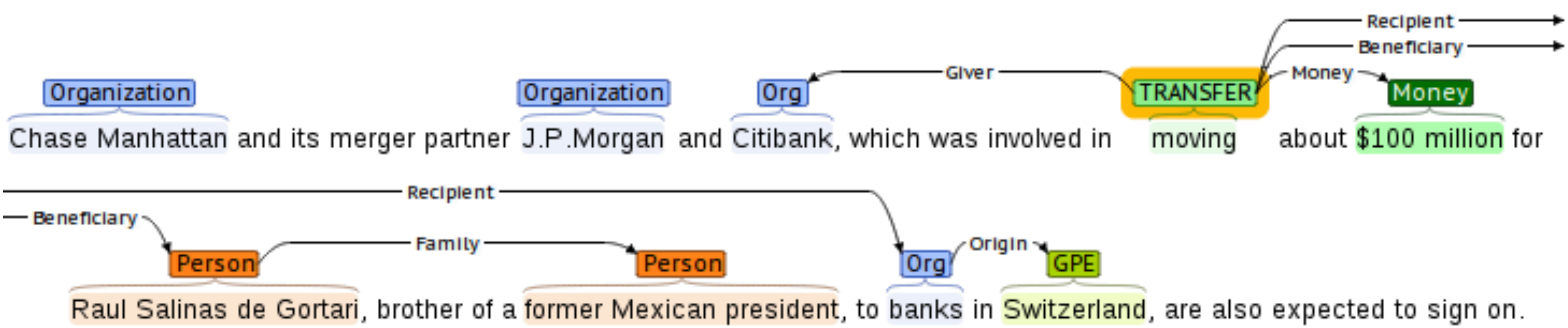
~400 docs.

KNOWLEDGE

Narrow Pre-specified domains

17 relations

Benchmark: Automatic Content Extraction (ACE)



Traditional Information Extraction

DATA

Edited Text

LEARNING

Supervised Learning (small datasets)

KNOWLEDGE

Narrow domains

Goals of my lab's research

DATA

User-Generated Text

LEARNING

Minimal Supervision **(large datasets)**

KNOWLEDGE

Open Domain

Traditional Information Extraction

DATA

Edited Text

LEARNING

Supervised Learning (small datasets)

KNOWLEDGE

Narrow domains

Opportunistically Gathered Data!

Goals of my lab's research

DATA

User-Generated Text

LEARNING

Minimal Supervision

(large datasets)

KNOWLEDGE

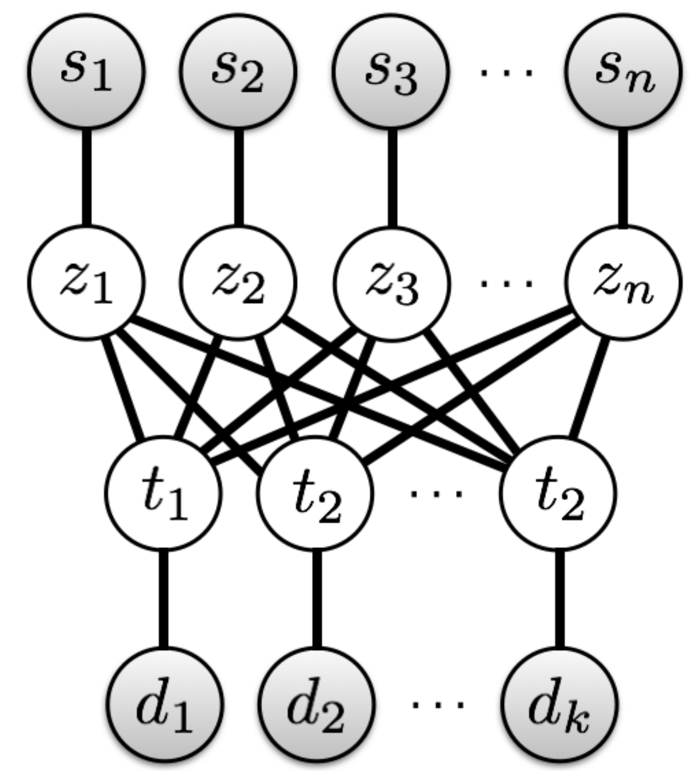
Open Domain

Weakly Supervised Learning for Information Extraction

1) Named Entity Recognition

Challenge: highly ambiguous labels

[Ritter, et. al. EMNLP 2011]



2) Relation Extraction

Challenge: missing data

[Ritter, et. al. TACL 2013]

3) Time Expressions

Challenge: diversity in noisy text

[Tabbasum, Ritter, Xu, EMNLP 2016]

4) Event Extraction

Challenge: lack of negative examples

[Ritter, et. al. WWW 2015]

[Konovalov, et. al. WWW 2017]

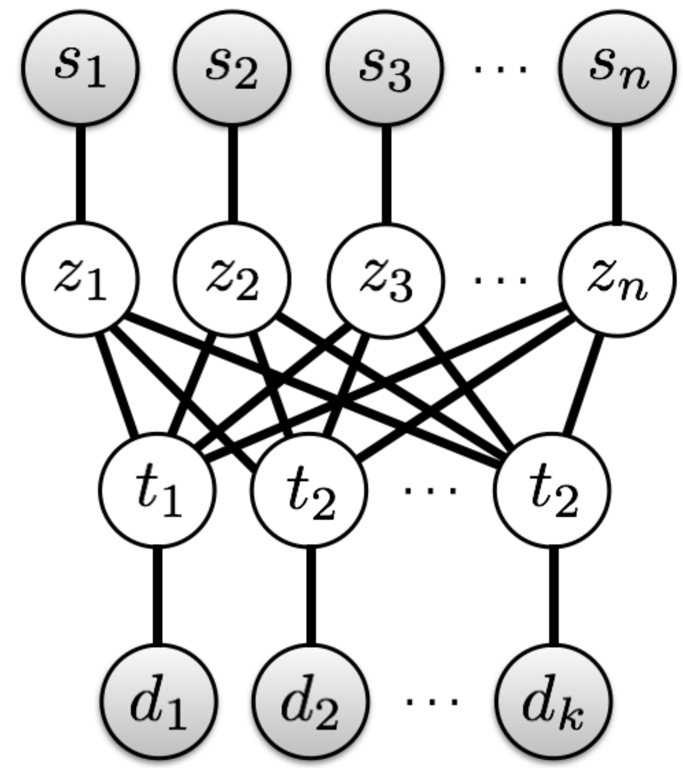
$$O(\theta) = \underbrace{\sum_{i}^{N} \log p_\theta(y_i|x_i)}_{\text{Log Likelihood}} - \underbrace{\lambda^U D(\tilde{p}||\hat{p}_\theta^{\text{unlabeled}})}_{\text{Label regularization}}$$

Weakly Supervised Learning for Information Extraction

1) Named Entity Recognition

Challenge: highly ambiguous labels

[Ritter, et. al. EMNLP 2011]



2) Relation Extraction

Challenge: missing data

[Ritter, et. al. TACL 2013]

3) Time Expressions

Challenge: diversity in noisy text

[Tabbasum, Ritter, Xu, EMNLP 2016]

4) Event Extraction

Challenge: lack of negative examples

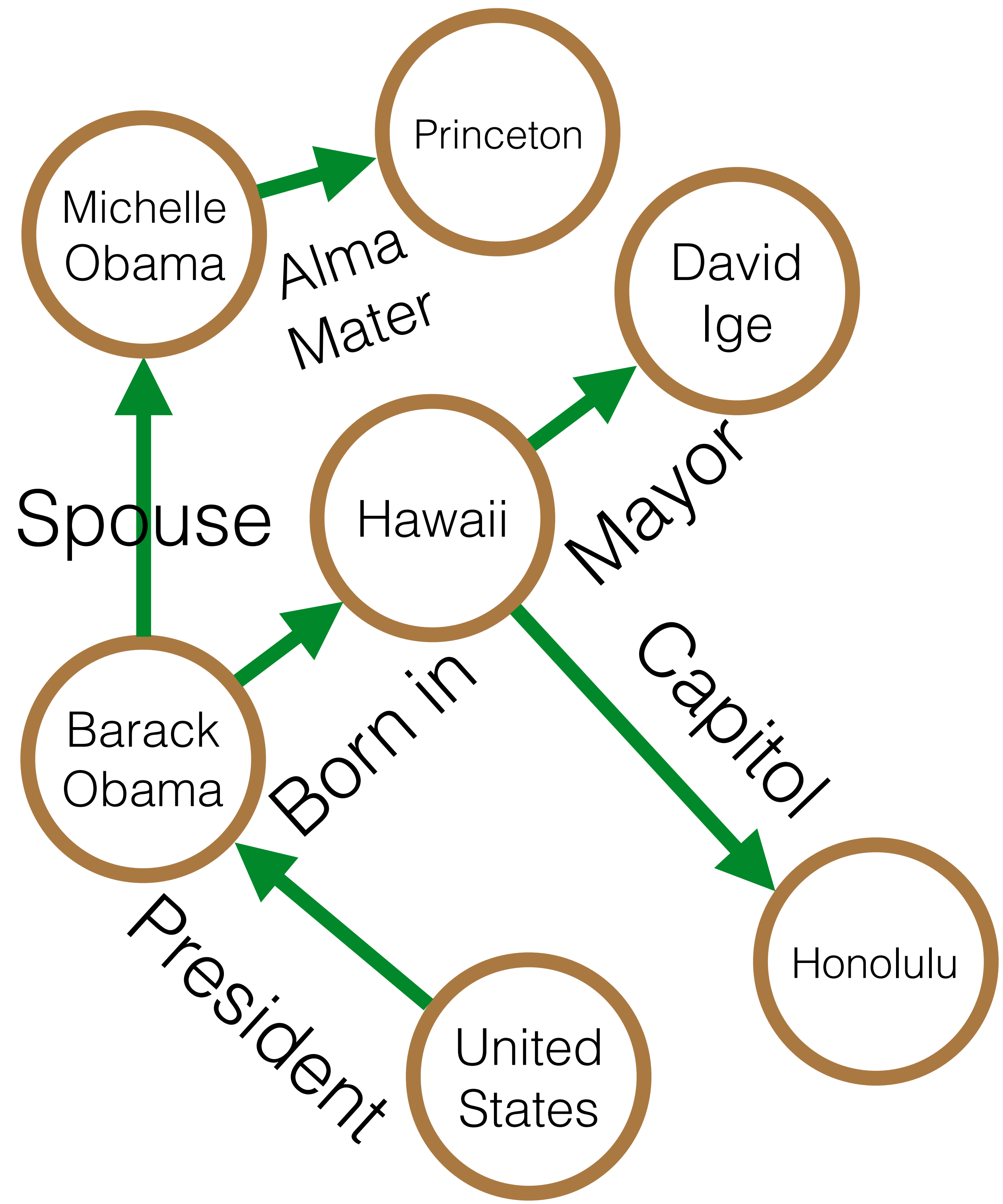
[Ritter, et. al. WWW 2015]

[Konovalov, et. al. WWW 2017]

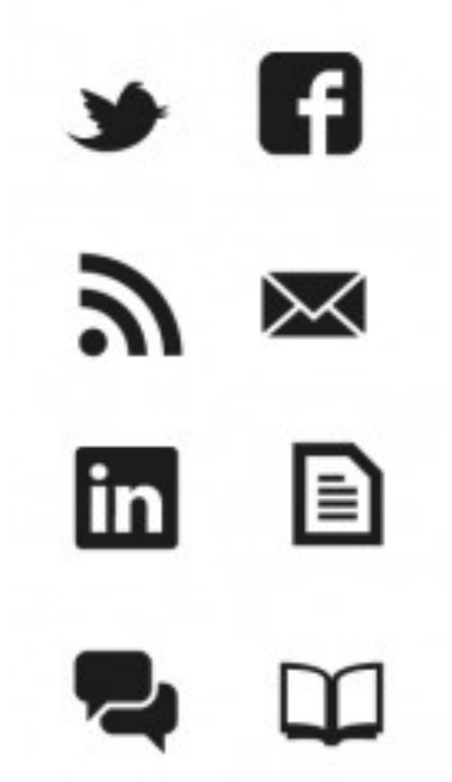
$$O(\theta) = \underbrace{\sum_{i}^{N} \log p_\theta(y_i|x_i)}_{\text{Log Likelihood}} - \underbrace{\lambda^U D(\tilde{p}||\hat{p}_\theta^{\text{unlabeled}})}_{\text{Label regularization}}$$

Goal: Realtime Information Extraction

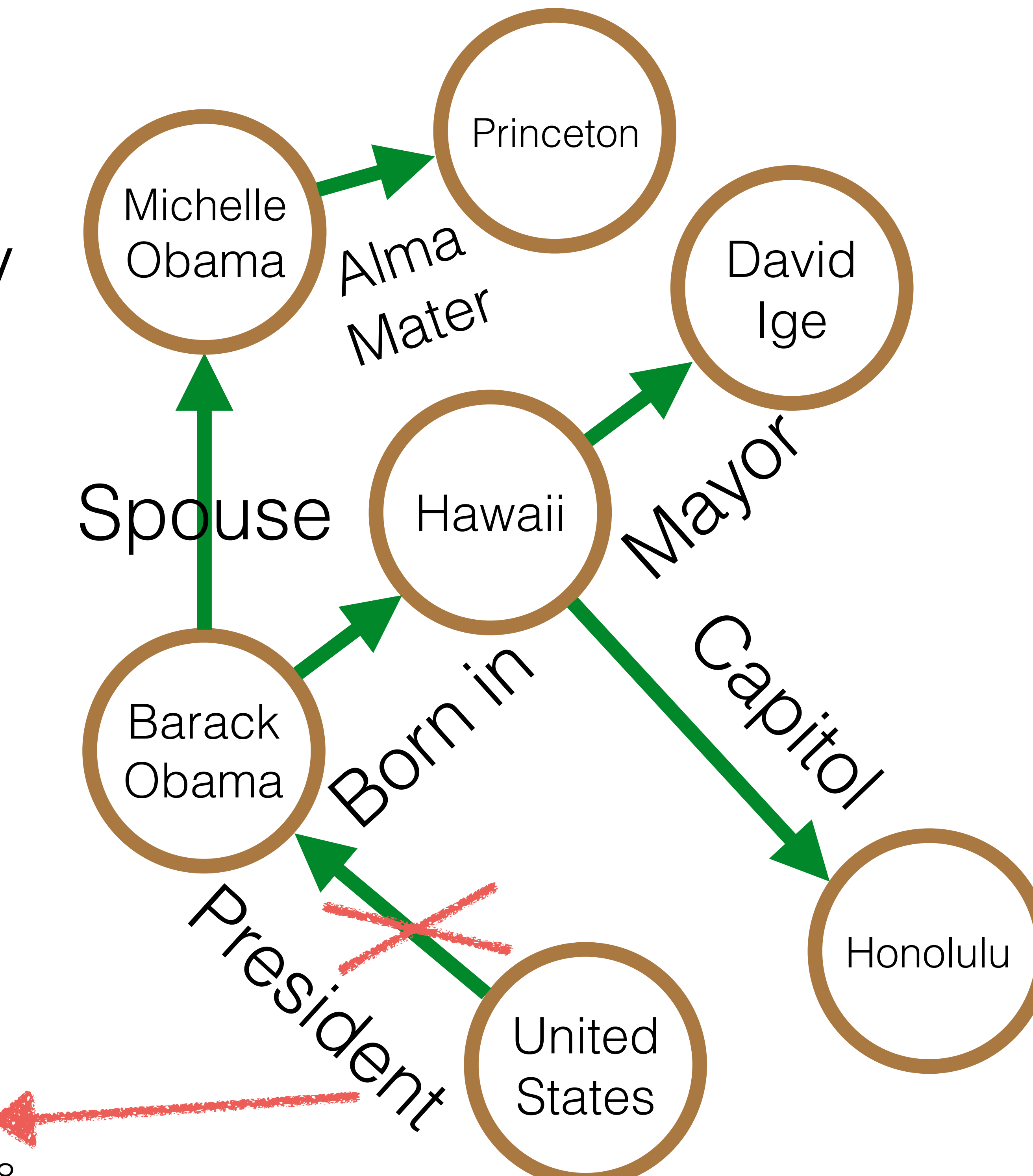
Continuously Extract new Entities, Relations and Events



Goal: Realtime Information Extraction



Continuously Extract new Entities, Relations and Events



Wikipedia: A dynamically evolving knowledge base

Jimmy Wales

Wales at the Wikimedia Conference 2013 board meeting

Board member of	Wikimedia Foundation Creative Commons
Spouse(s)	Pamela Green (m. 1986, div)
Title	President of Wikia, Inc. (2004–present)

Wikipedia: A dynamically evolving knowledge base

Wiki wedding: Wikipedia founder Jimmy Wales marries Tony Blair's former aide

Wikipedia founder Jimmy Wales married Tony Blair's former diary secretary Kate Garvey on Saturday, witnessed by guests from the world of politics and celebrity.

143 65 0 0 208 Email

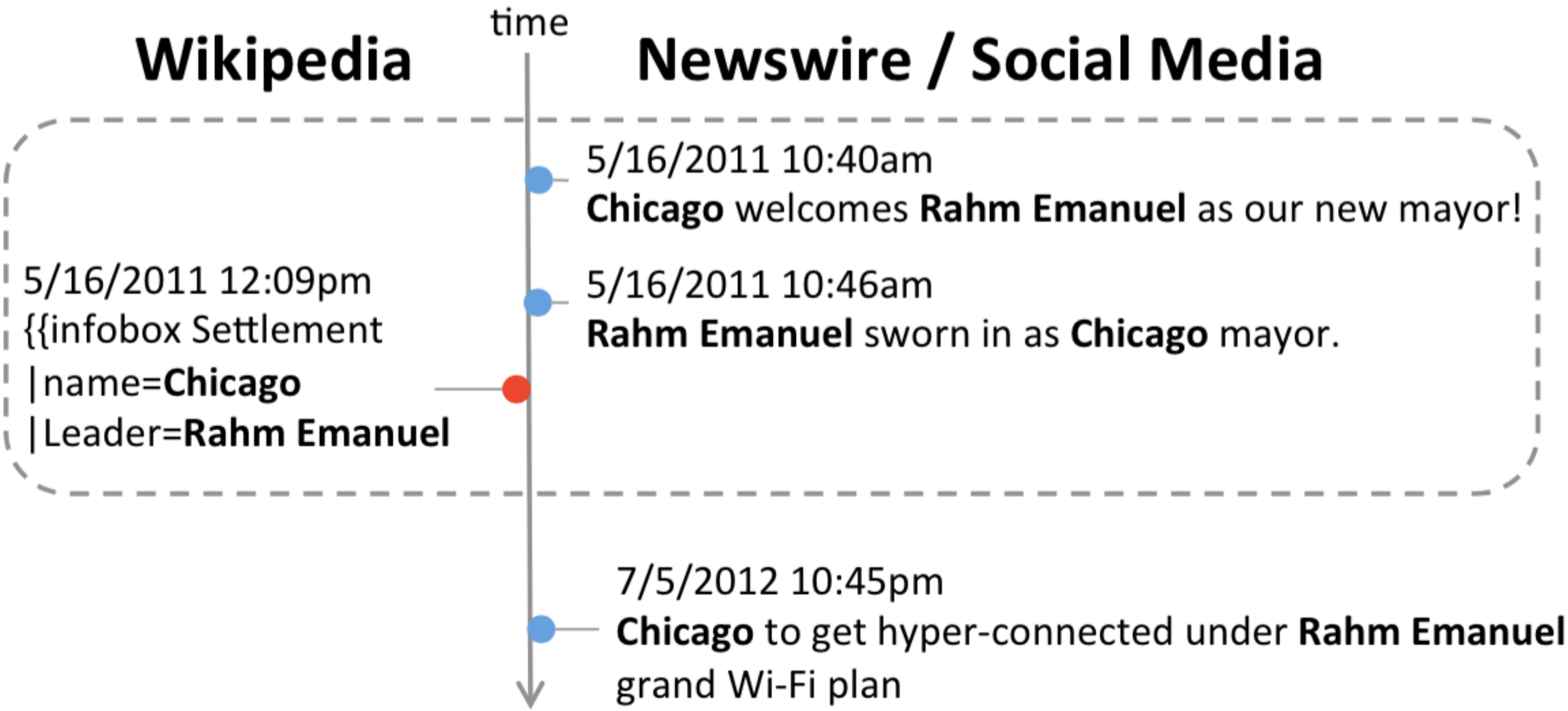
Jimmy Wales

Wales at the Wikimedia Conference 2013 board meeting

Board member	Wikimedia Foundation Creative Commons
Spouse(s)	Kate Garvey (m. 2012)
Title	President of Wikia, Inc. (2004–present)

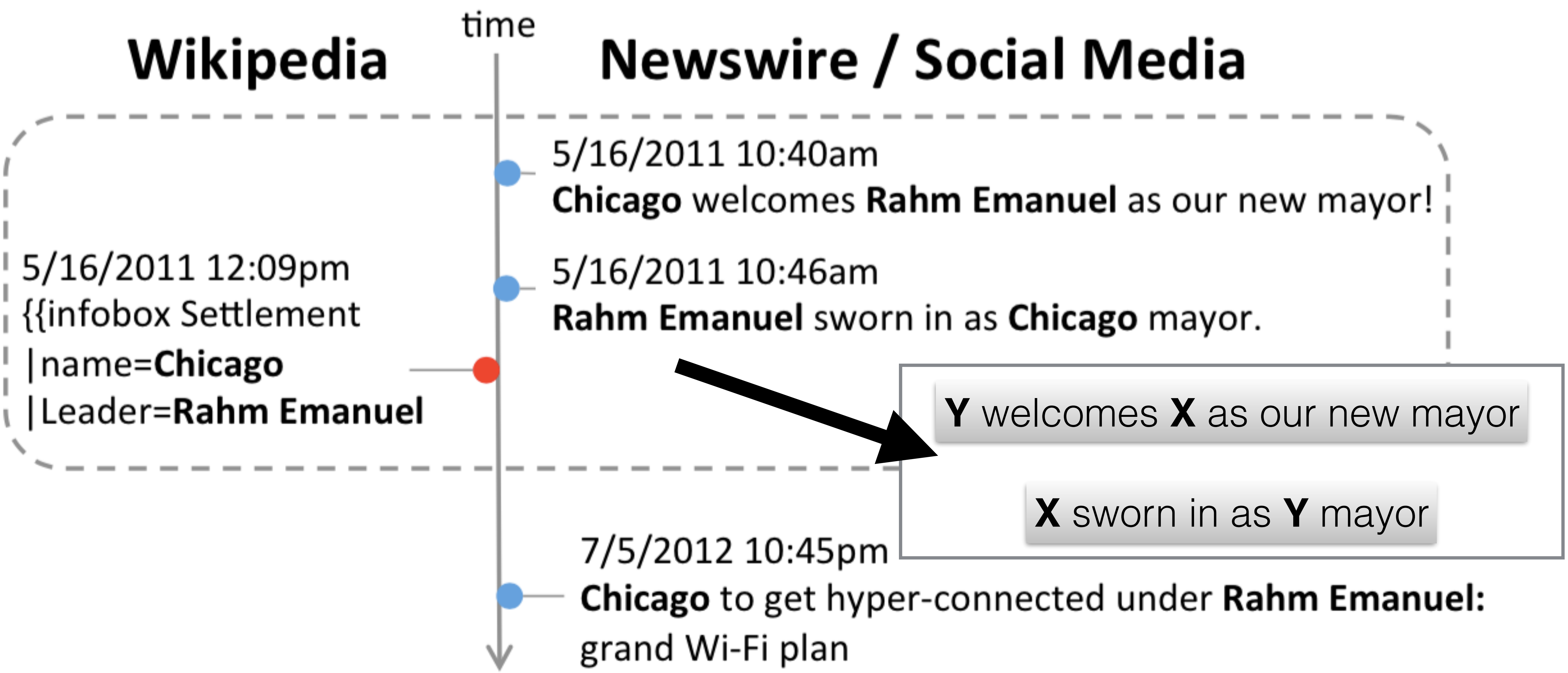
Learning to Extract Events

[Konovalov, Strauss, Ritter, O'Connor WWW 2017]



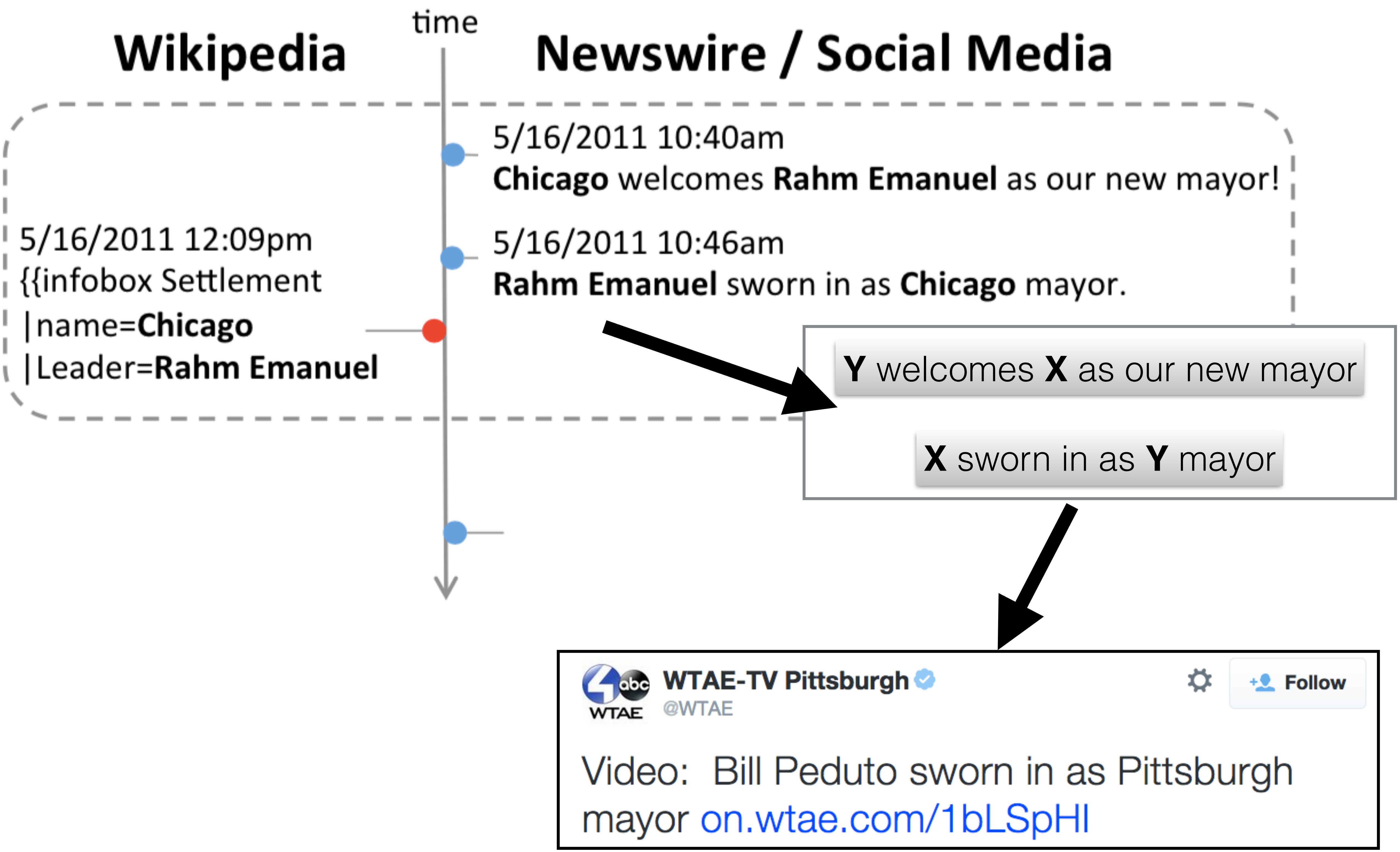
Learning to Extract Events

[Konovalov, Strauss, Ritter, O'Connor WWW 2017]



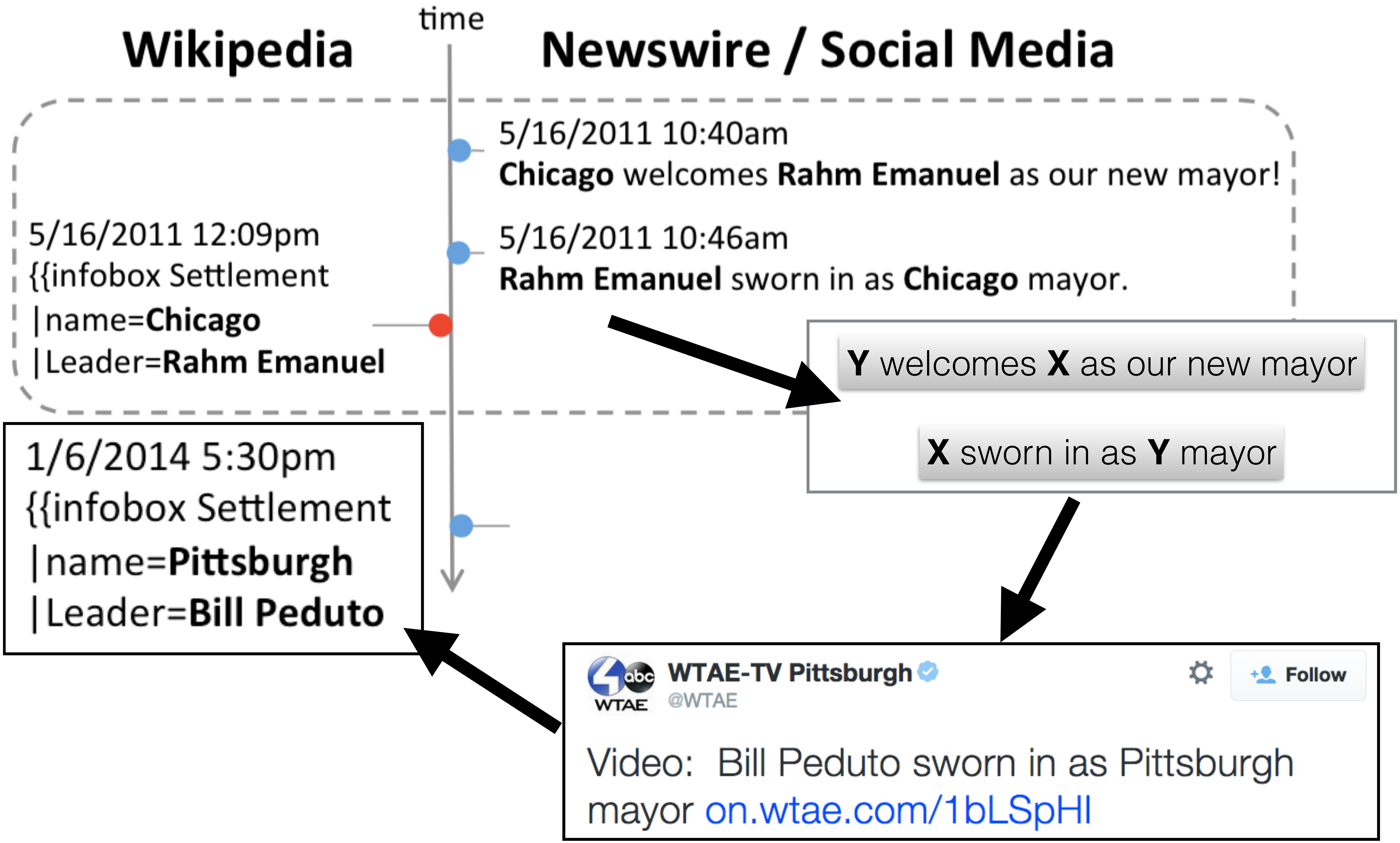
Learning to Extract Events

[Konovalov, Strauss, Ritter, O'Connor WWW 2017]

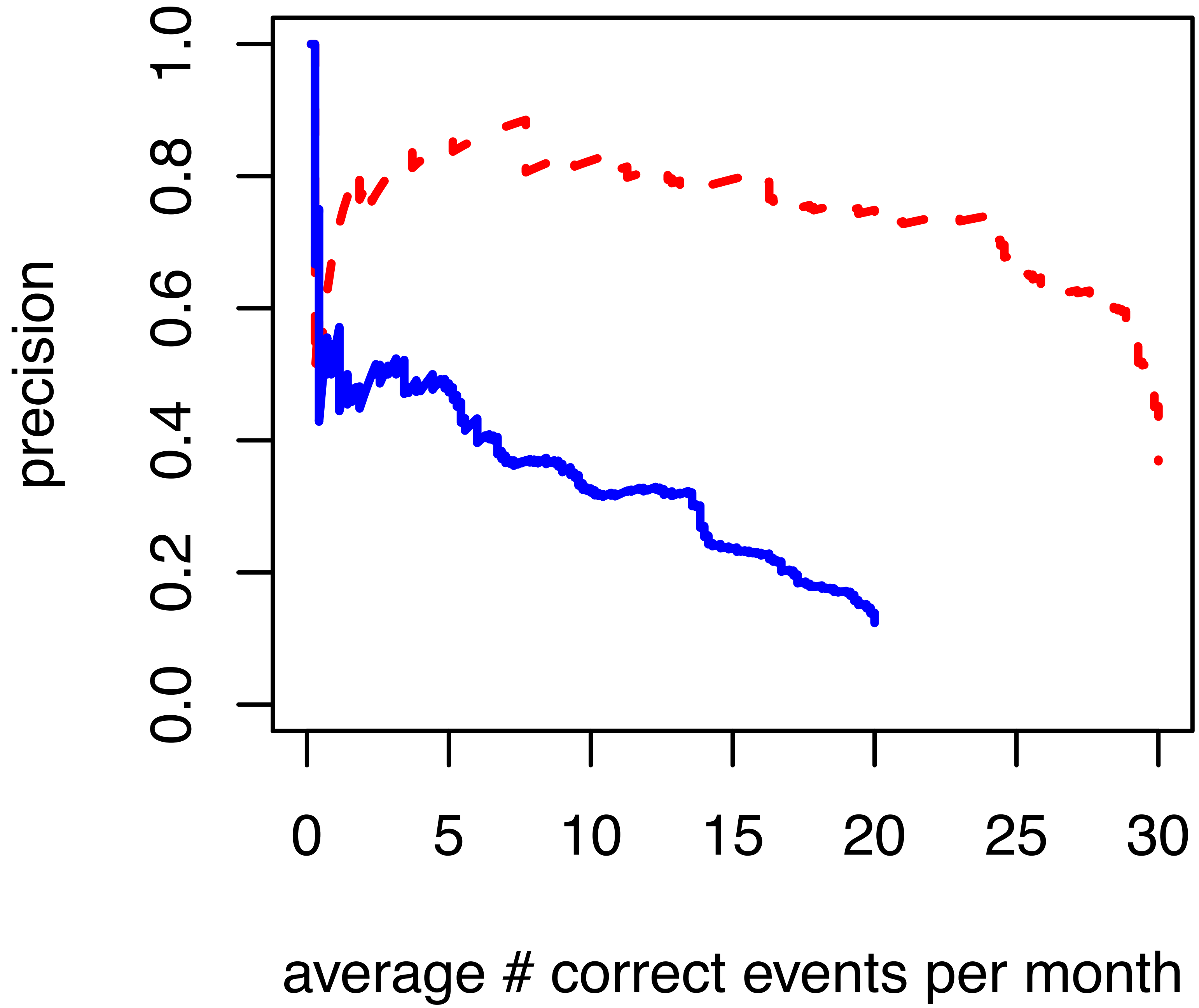


Learning to Extract Events

[Konovalov, Strauss, Ritter, O’Connor WWW 2017]



Results



1.0
0.8
0.6
0.4
0.2
0.0
precision
0
5
10
15
20
25
30
average # correct events per month

[Ritter, Cherry, Dolan EMNLP 2011b]

Data-Driven Conversation

- Twitter: ~ 1/2 Billion Public SMS-Style Conversations per Month
- **Goal:** Learn conversational agents directly from massive volumes of data.

[Ritter, Cherry, Dolan EMNLP 2011b]

Follow-Up Work:
Data-Driven Conversation

2015:

- O. Vinyals, Q.V. Le. A Neural Conversational Model. **ICML Deep Learning Workshop 2015**
- Alessandro Sordoni, Michel Galley, Michael Auli, Chris Brockett, Yangfeng Ji, Meg Mitchell, Jian-Yun Nie, Jianfeng Gao, and Bill Dolan, A Neural Network Approach to Context-Sensitive Generation of Conversational Responses. **NAACL 2015**
- Lifeng Shang, Zhengdong Lu, Hang Li. Neural Responding Machine for Short Text Conversation. **ACL 2015**

2016:

- I. Serban, A. Sordoni, Y. Bengio, A. Courville and J. Pineau. Building End-To-End Dialogue Systems Using Generative Hierarchical Neural Networks. In Proc of **AAAI, 2016**.
- Jesse Dodge, Andreea Gane, Xiang Zhang, Antoine Bordes, Sumit Chopra, Alexander Miller, Arthur Szlam, Jason Weston. Evaluating Prerequisite Qualities for Learning End-to-end Dialog Systems, **ICLR 2016**
- Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao and Bill Dolan. A Diversity-Promoting Objective Function for Neural Conversation Models. **NAACL 2016**

[Ritter, Cherry, Dolan EMNLP 2011b]

Follow-Up Work:

Data-Driven Conversation

2015:

- O. Vinyals, Q.V. Le. A Neural Conversational Model. **ICML Deep Learning Workshop 2015**
- Alessandro Sordoni, Michel Galley, Michael Auli, Chris Brockett, Yangfeng Ji, Meg Mitchell, Jian-Yun Nie, Jianfeng Gao, and Bill Dolan, A Neural Network Approach to Context-Sensitive Generation of Conversational Responses. **NAACL 2015**
- Lifeng Shang, Zhengdong Lu, Hang Li. Neural Responding Machine for Short Text Conversation. **ACL 2015**

2016:

- I. Serban, A. Sordoni, Y. Bengio, A. Courville and J. Pineau. Building End-To-End Dialogue Systems Using Generative Hierarchical Neural Networks. In Proc of **AAAI, 2016**.
- Jesse Dodge, Andreea Gane, Xiang Zhang, Antoine Bordes, Sumit Chopra, Alexander Miller, Arthur Szlam, Jason Weston. Evaluating Prerequisite Qualities for Learning End-to-end Dialog Systems, **ICLR 2016**
- Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao and Bill Dolan. A Diversity-Promoting Objective Function for Neural Conversation Models. **NAACL 2016**

Challenge: Some replies have high probability given any input:

"I don't know" *"OK"*

"I'm sorry" *"I love you"*

Smart Reply: Automated Response Suggestion for Emai

Anjuli Kannan* Karol Kurach* Sujith Ravi*

Tobias Kaufmann* Andrew Tomkins Balint Miklos

Greg Corrado László Lukács Marina Ganea

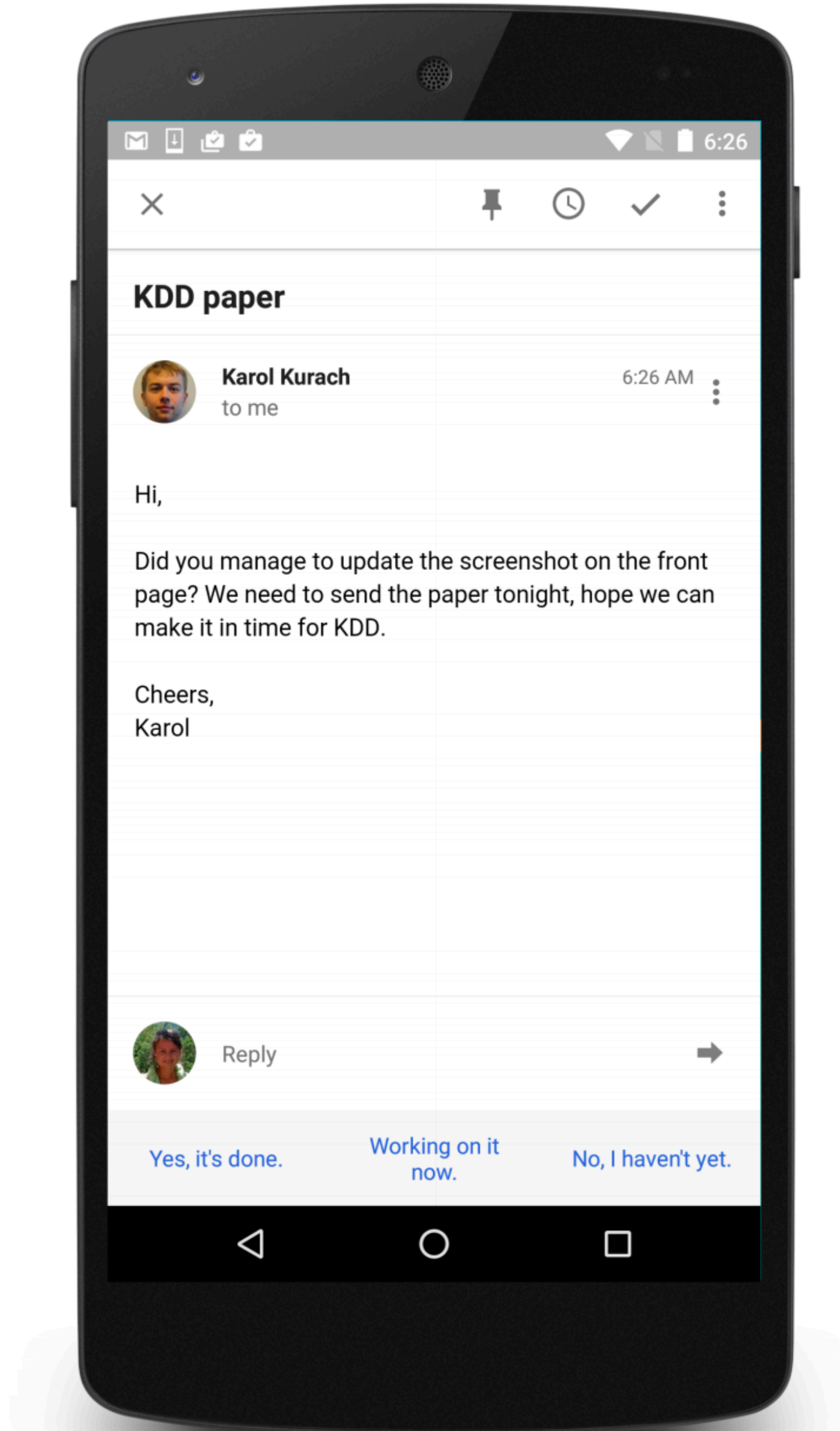
Peter Young Vivek Ramavajjala

Google
{anjuli, kkurach, sravi, snufkin}@google.com

ABSTRACT

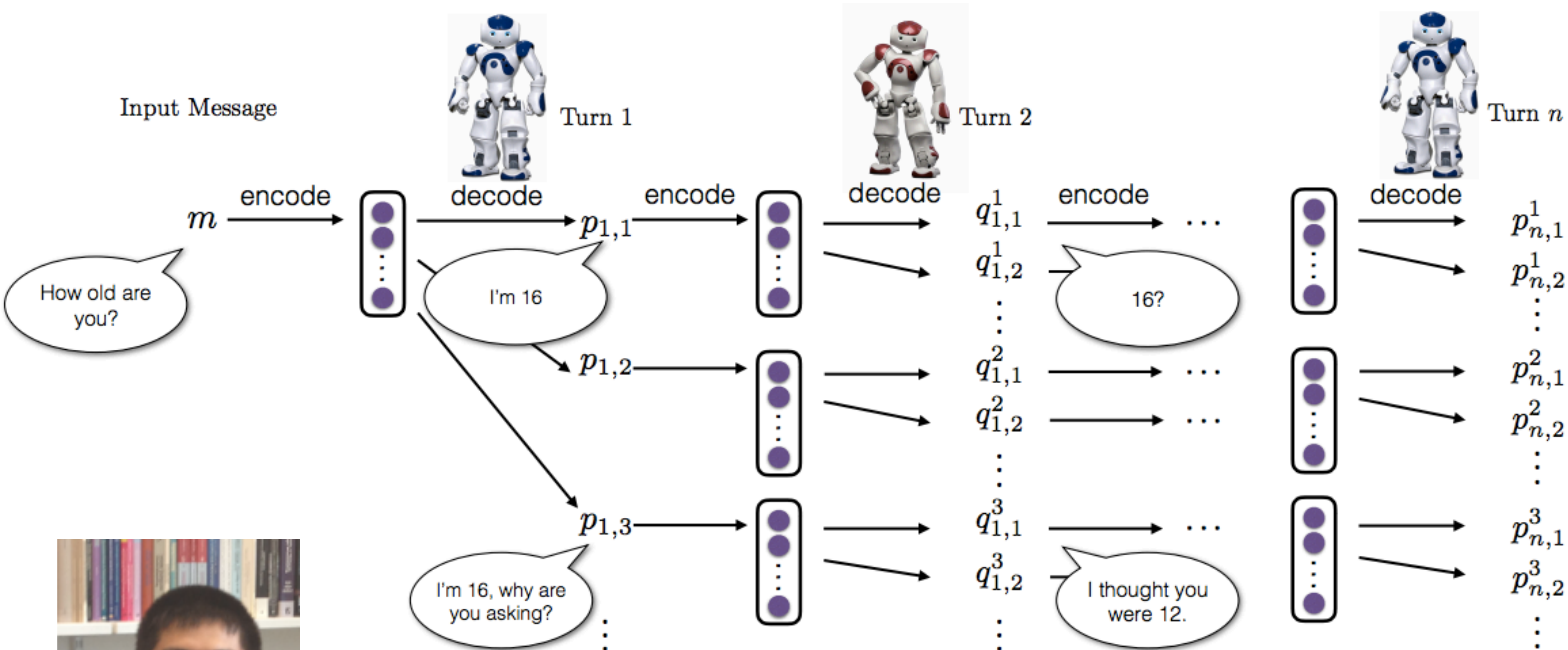
In this paper we propose and investigate a novel end-to-end method for automatically generating short email responses, called Smart Reply. It generates semantically diverse suggestions that can be used as complete email responses with just one tap on mobile. The system is currently used in *Inbox by Gmail* and is responsible for assisting with 10% of all mobile responses. It is designed to work at very high throughput and process hundreds of millions of messages daily. The system exploits state-of-the-art, large-scale deep learning.

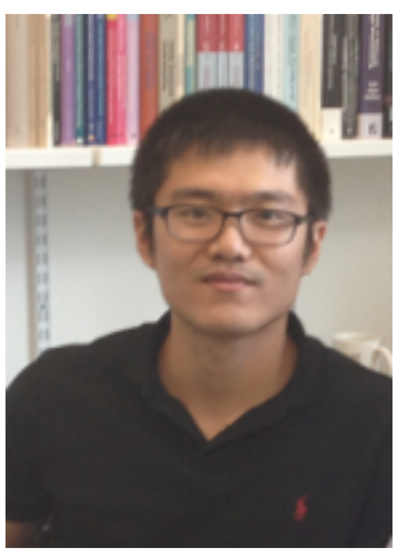
We describe the architecture of the system as well as the challenges that we faced while building it, like response diversity and scalability. We also introduce a new method for semantic clustering of user-generated content that requires only a modest amount of explicitly labeled data.



[Li, Monroe, Ritter, Galley, Gao, Jurafsky EMNLP 2016b]

Deep Reinforcement Learning





Thanks!

Justin Betteridge	(CMU)
William Casey	(CMU)
Xinlei Chen	(CMU)
Colin Cherry	(NRC)
Sam Clark	(eBay)
Marie de Marneffe	(OSU)
Bill Dolan	(MSR)
Oren Etzioni	(AI2)
Abhinav Gupta	(CMU)

Ed Hovy	(CMU)
Dan Jurafsky	(Stanford)
Jiwei Li	(Stanford)
Mausam	(IIT-D)
Tom Mitchell	(CMU)
Brendan O'Connor	(UMass)
Evan Wright	(CMU)
Wei Xu	(OSU)
Luke Zettlemoyer	(UW)

http://aritter.github.io/

@alan_ritter