

Large-Scale Learning for Information Extraction

Alan Ritter
Computer Science and Engineering
Ohio State University
 @alan_ritter

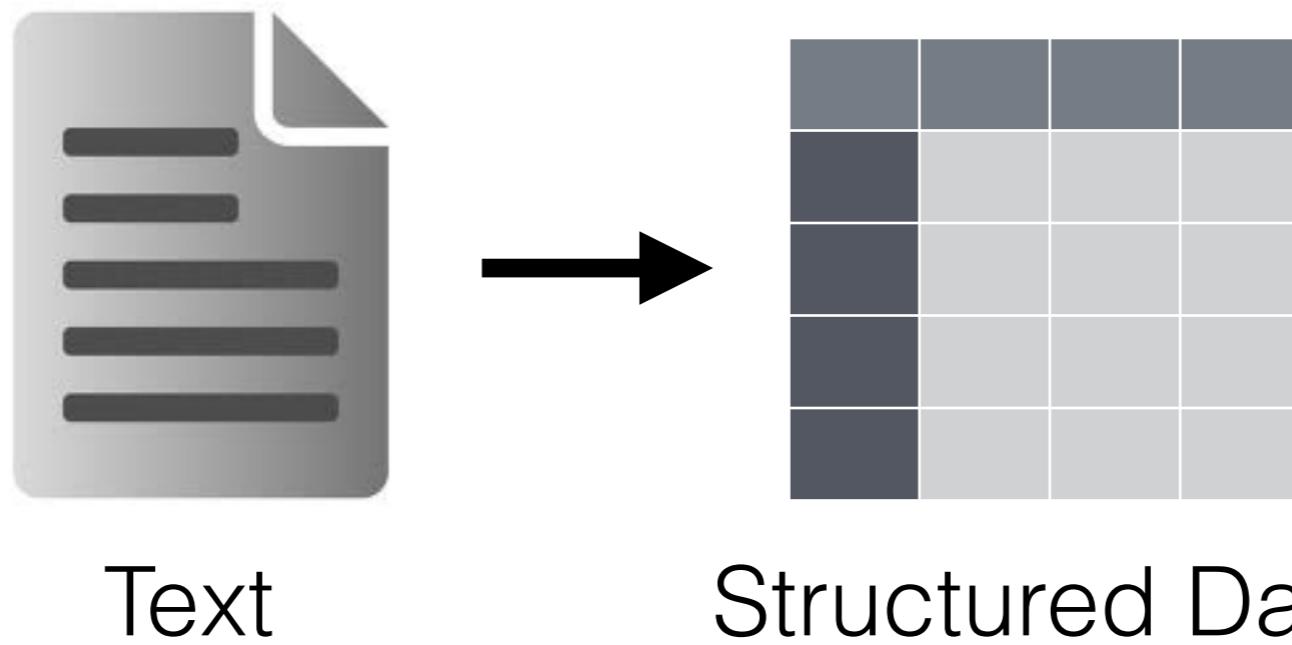


THE OHIO STATE UNIVERSITY

Humanity's Collective Knowledge is Locked in Text



Information Extraction



Traditional Information Extraction



1) Humans Annotate Text

Traditional Information Extraction



1) Humans Annotate Text

2) Supervised
Machine Learning

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

Traditional Information Extraction

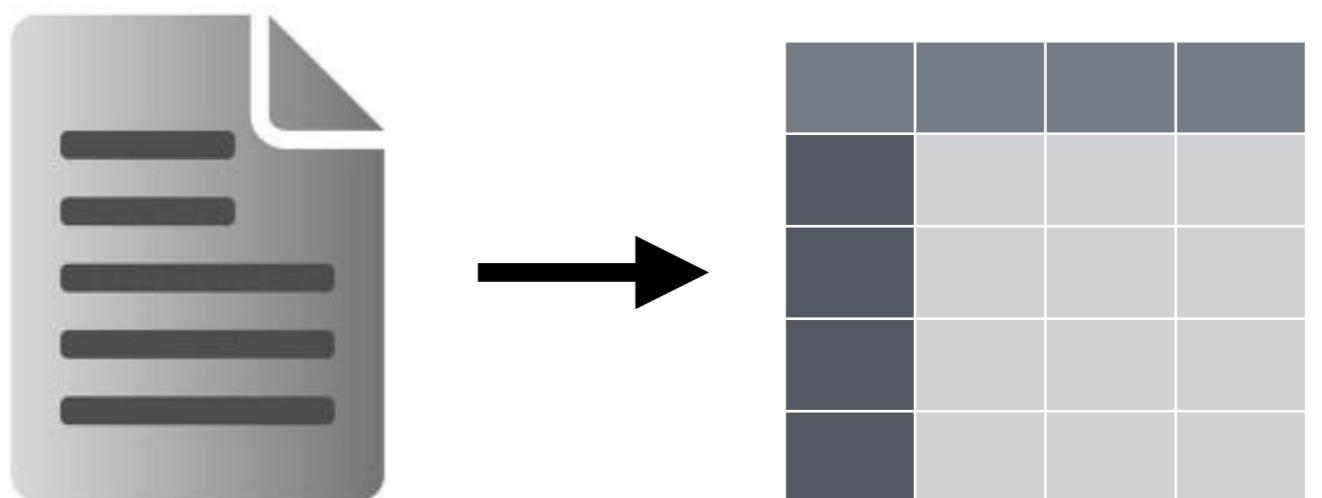


1) Humans Annotate Text

2) Supervised
Machine Learning

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

3) Apply Models to
New Documents



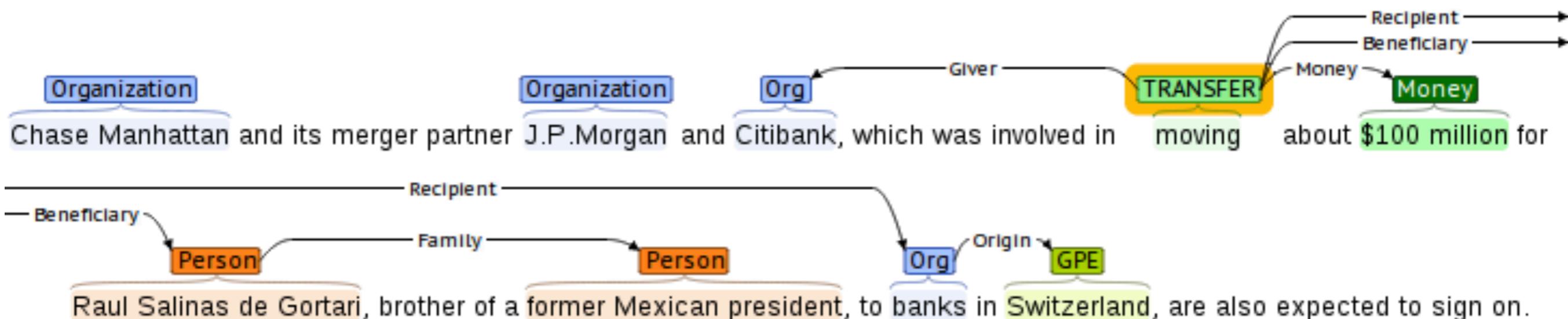
Traditional Information Extraction: Key Limitations



Traditional Information Extraction: Key Limitations



Benchmark: Automatic Content Extraction (ACE)



Traditional Information Extraction



Goals of my lab's research



Traditional Information Extraction



Goals of my lab's research



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

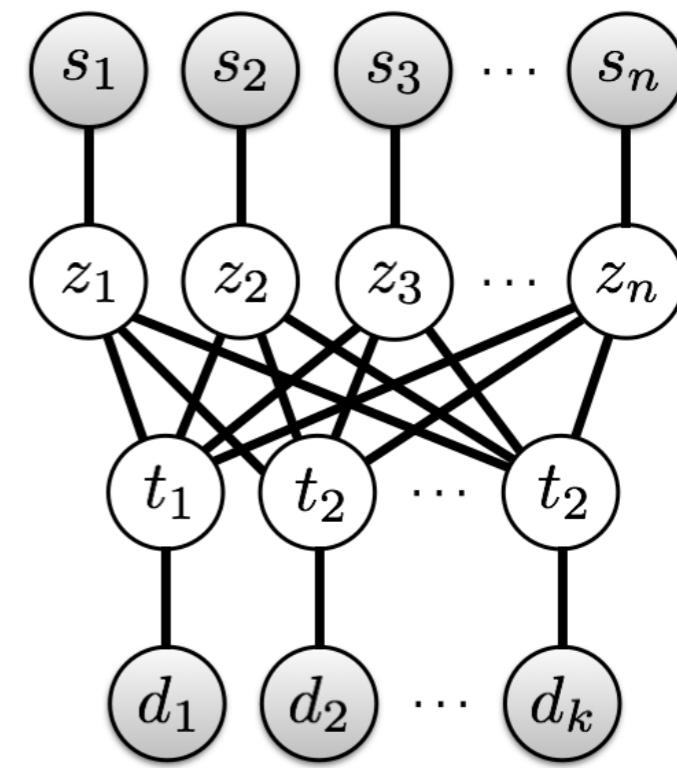
[Tabbasum, Ritter, Xu, EMNLP 2016]

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

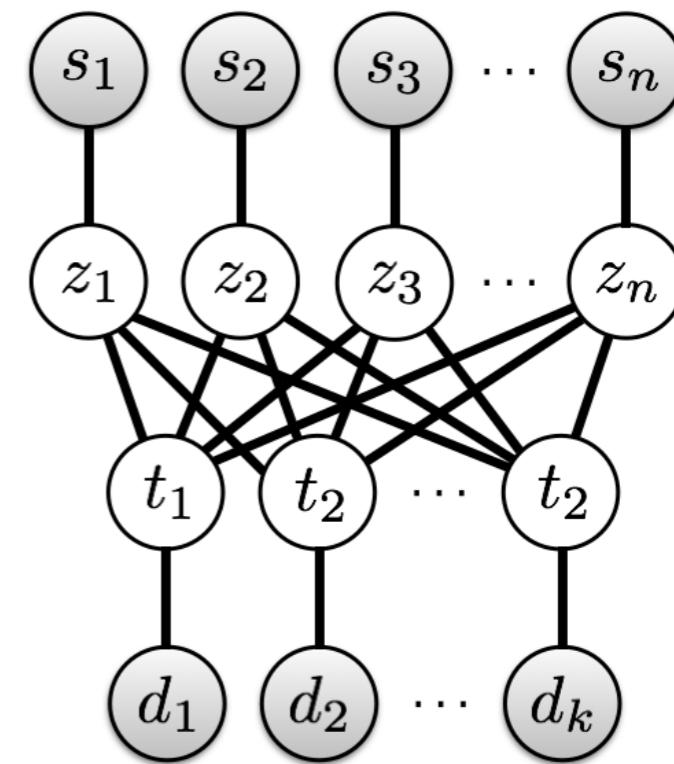
[Tabbasum, Ritter, Xu, EMNLP 2016]

→ 4) Event Extraction

Challenge: lack of negative examples

[Ritter, et. al. WWW 2015]

[Konovalov, et. al. WWW 2017]

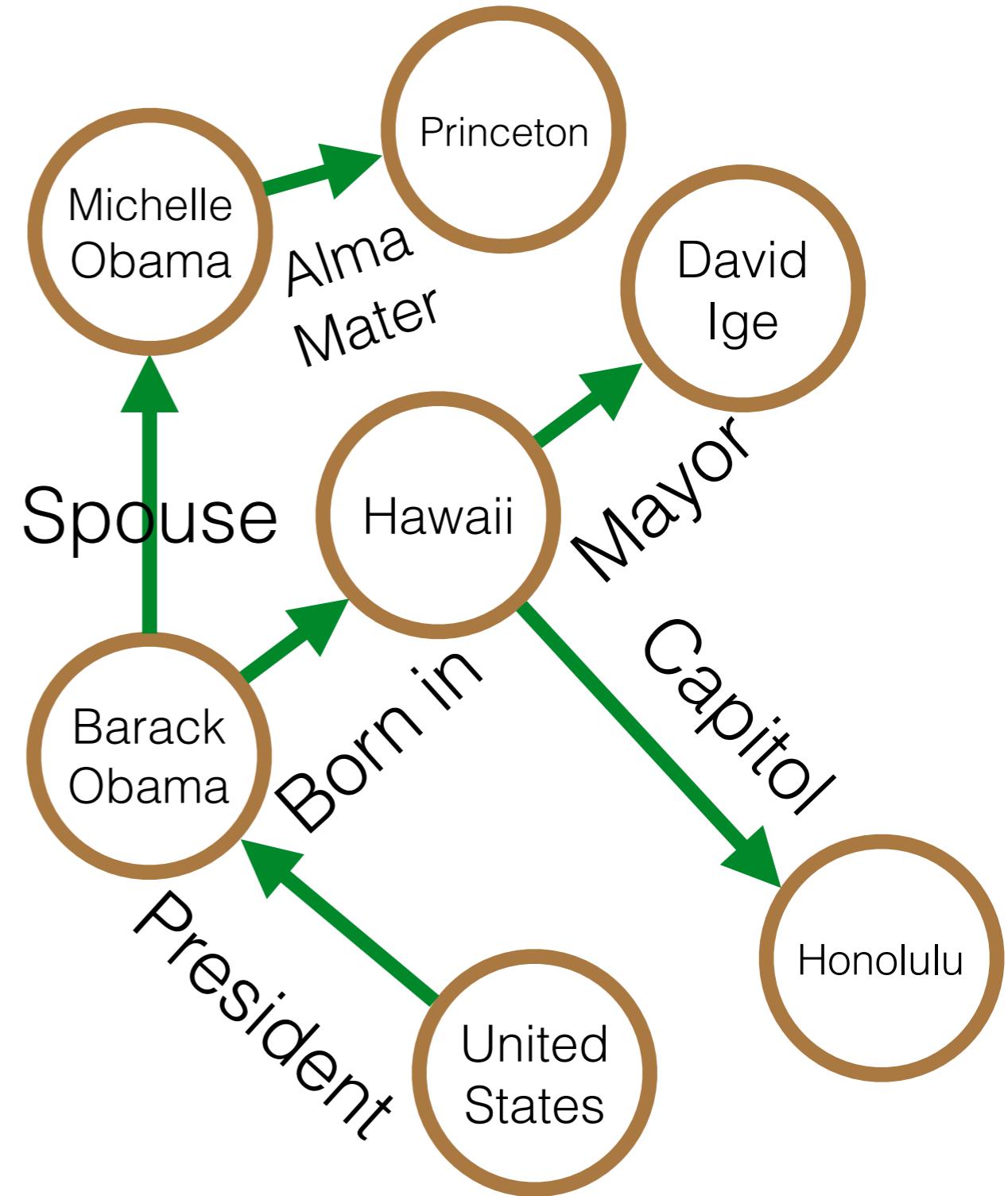


$$O(\theta) = \underbrace{\sum_i^N \log p_\theta(y_i|x_i)}_{\text{Log Likelihood}} - \underbrace{\lambda^U D(\tilde{p}||\hat{p}_\theta^{\text{unlabeled}})}_{\text{Label regularization}}$$

Goal: Realtime Information Extraction



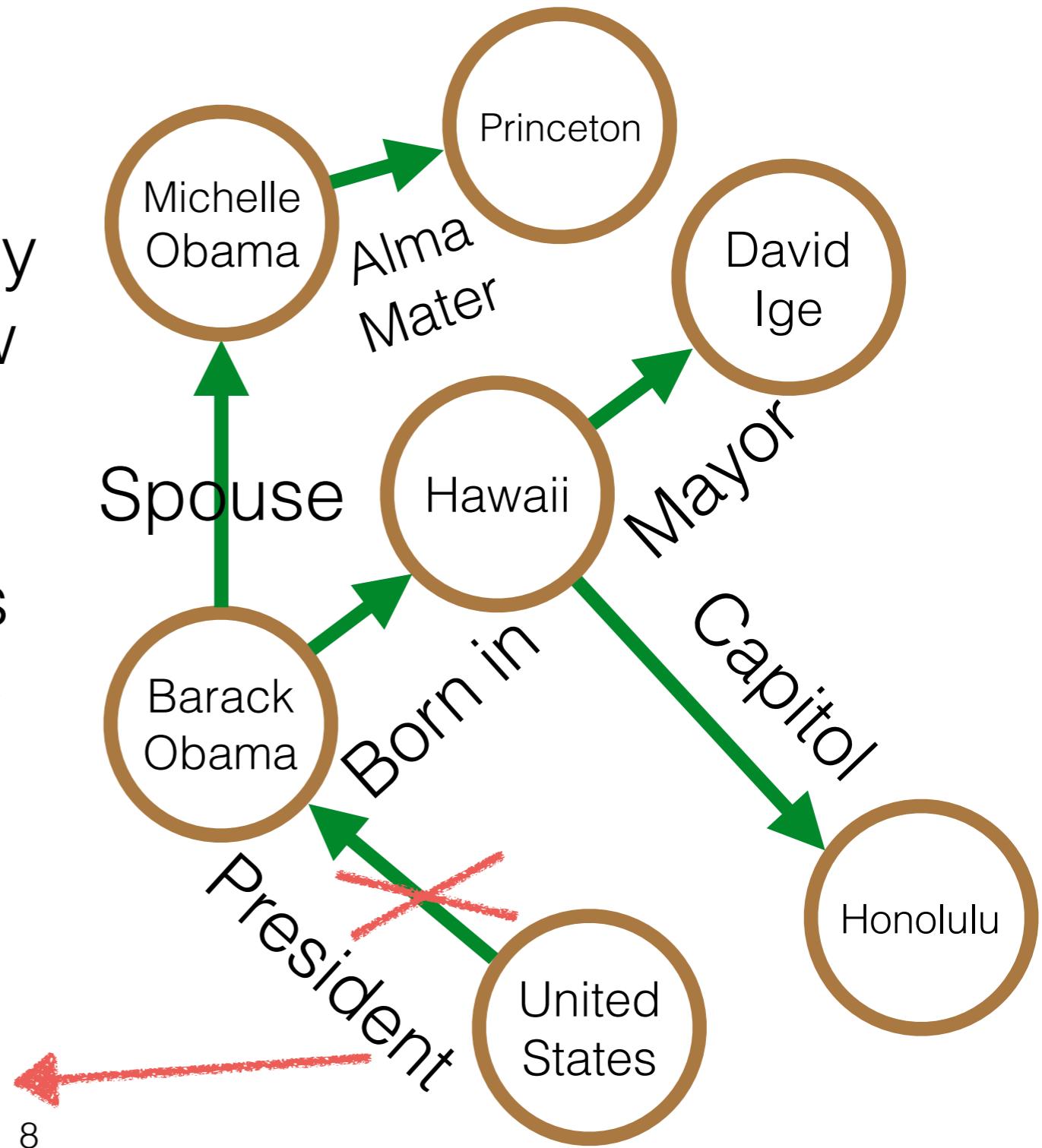
Continuously Extract new Entities, Relations and Events



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.



Jimmy Wales



Wales at the Wikimedia Conference 2013 board meeting
Wikimedia Foundation
Creative Commons

Board member

Spouse(s)

Title

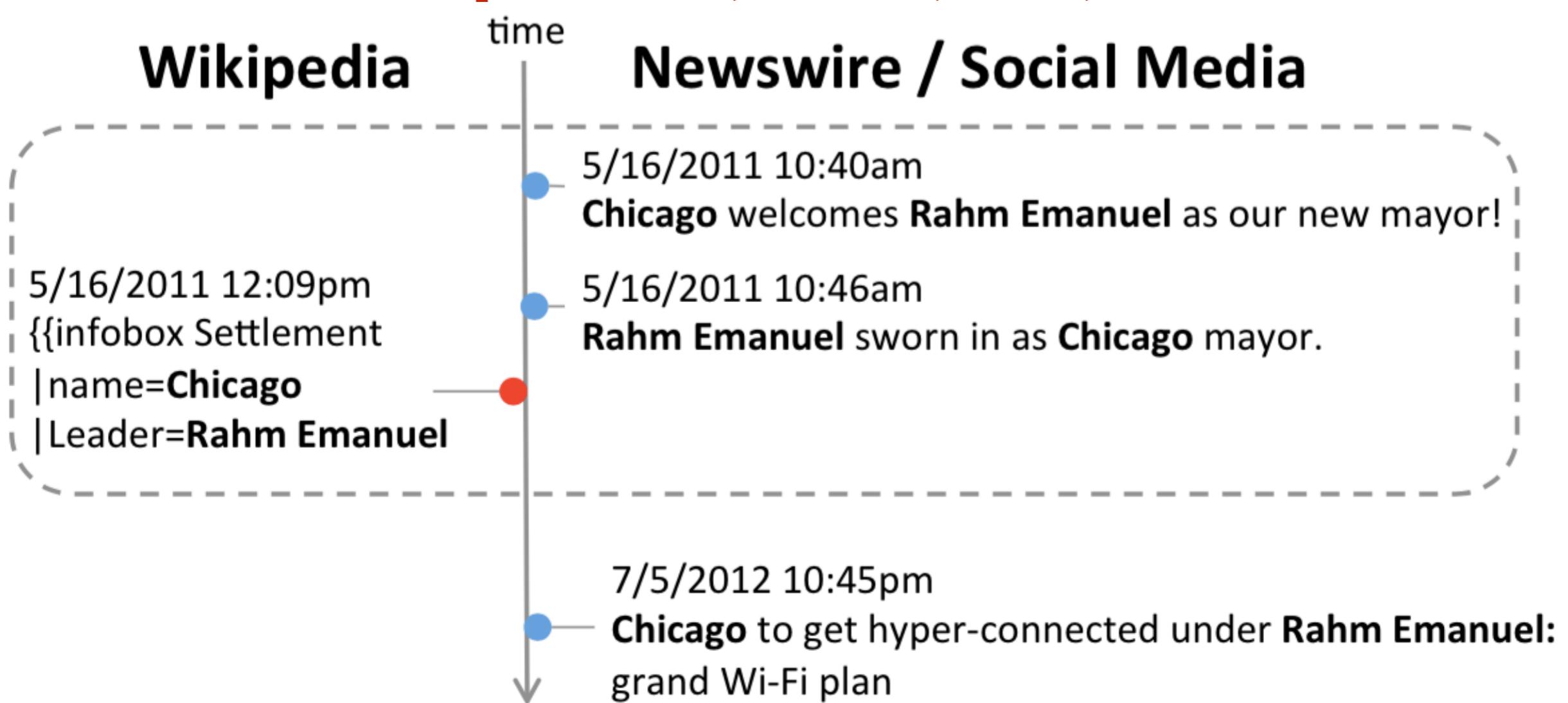
Kate Garvey
(m. 2012)

President of [Wikia, Inc.](#)
(2004–present)

A red arrow points from the text "Spouse(s)" to the "Kate Garvey" entry in the sidebar.

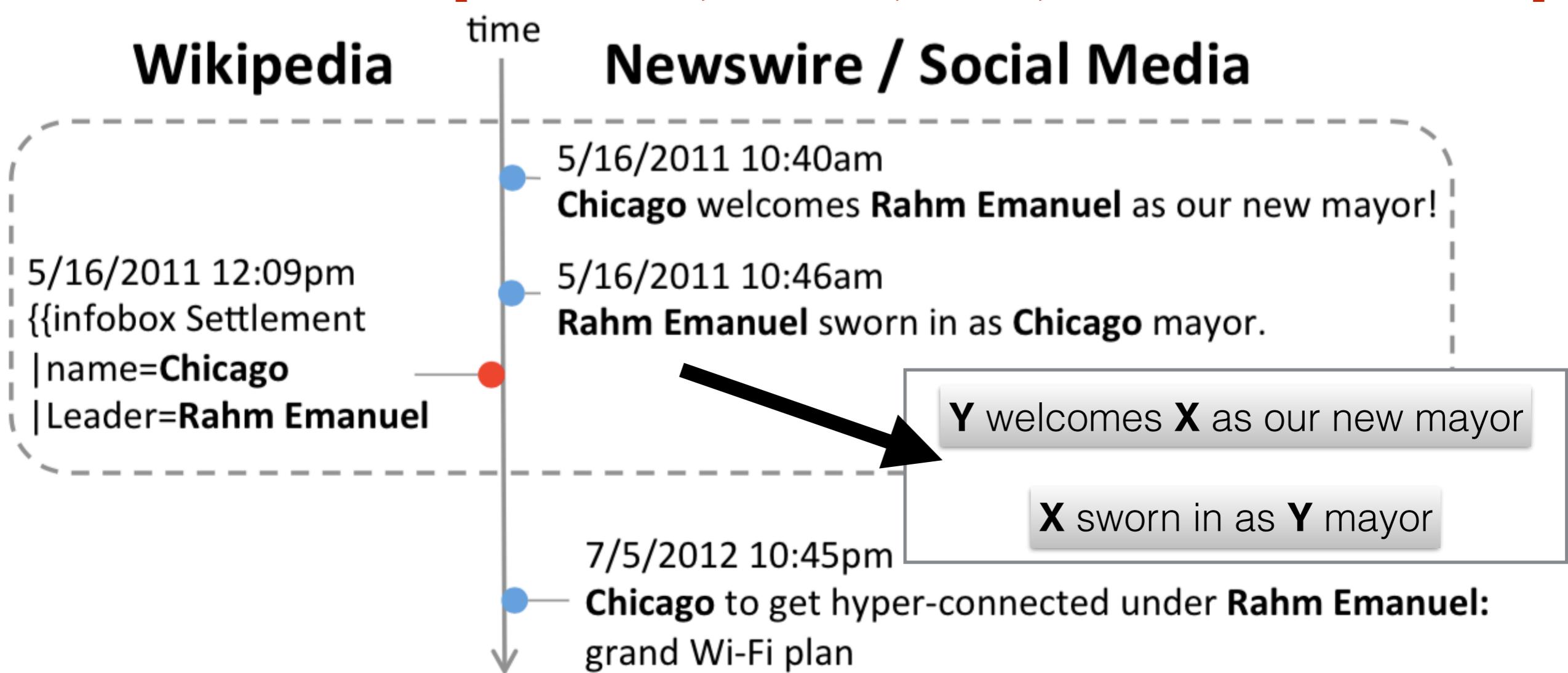
Learning to Extract Events

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



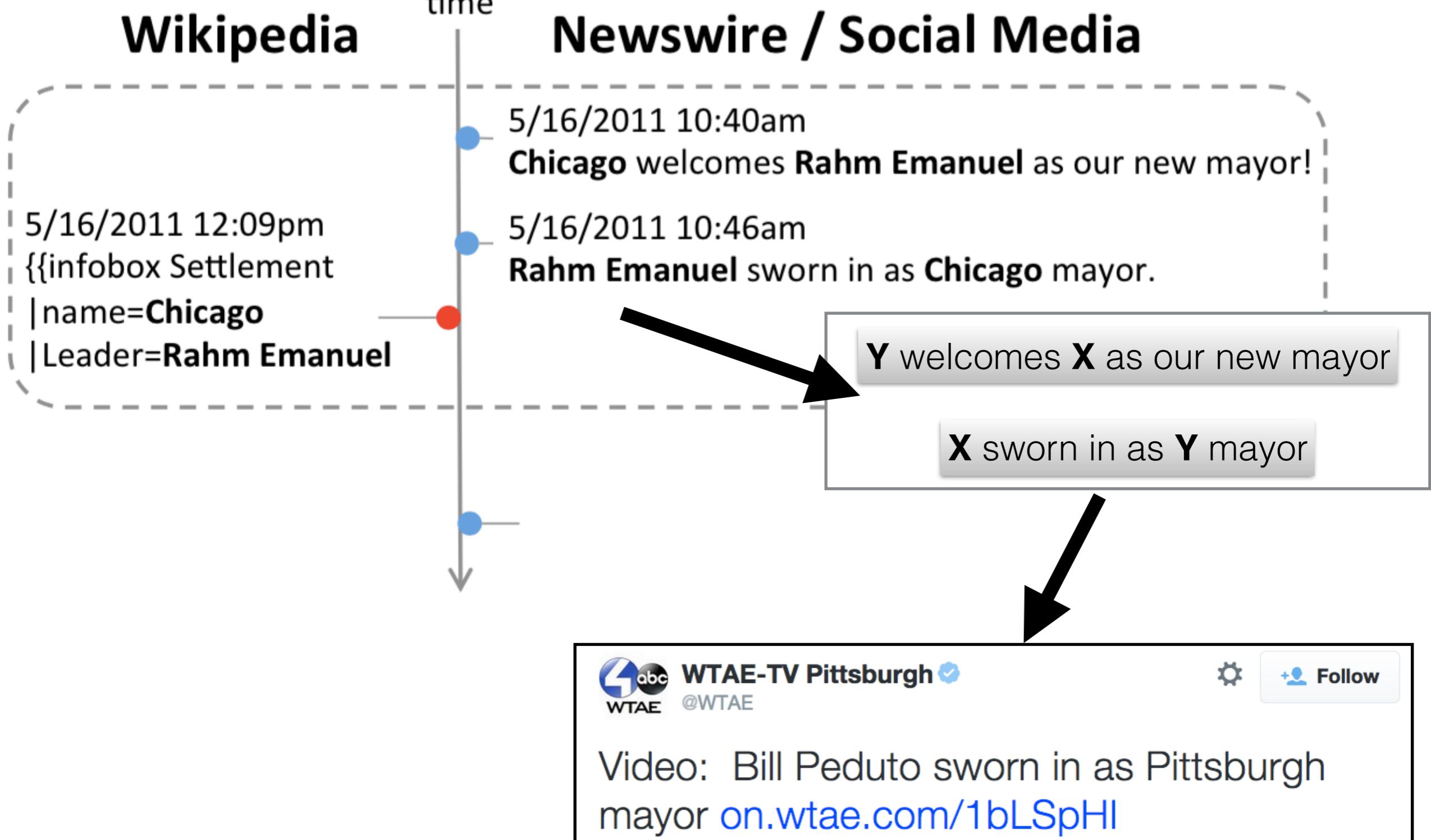
Learning to Extract Events

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



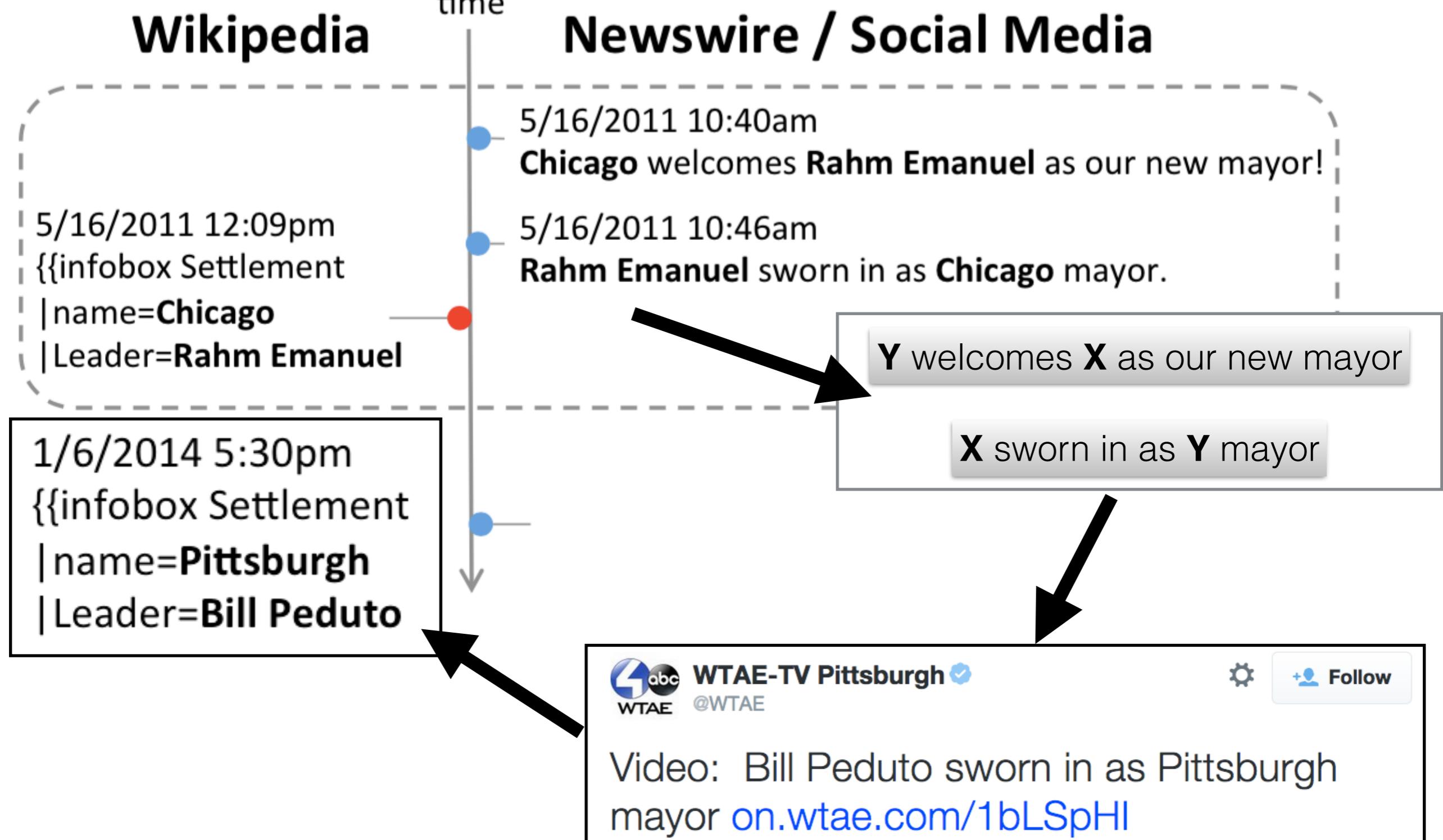
Learning to Extract Events

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

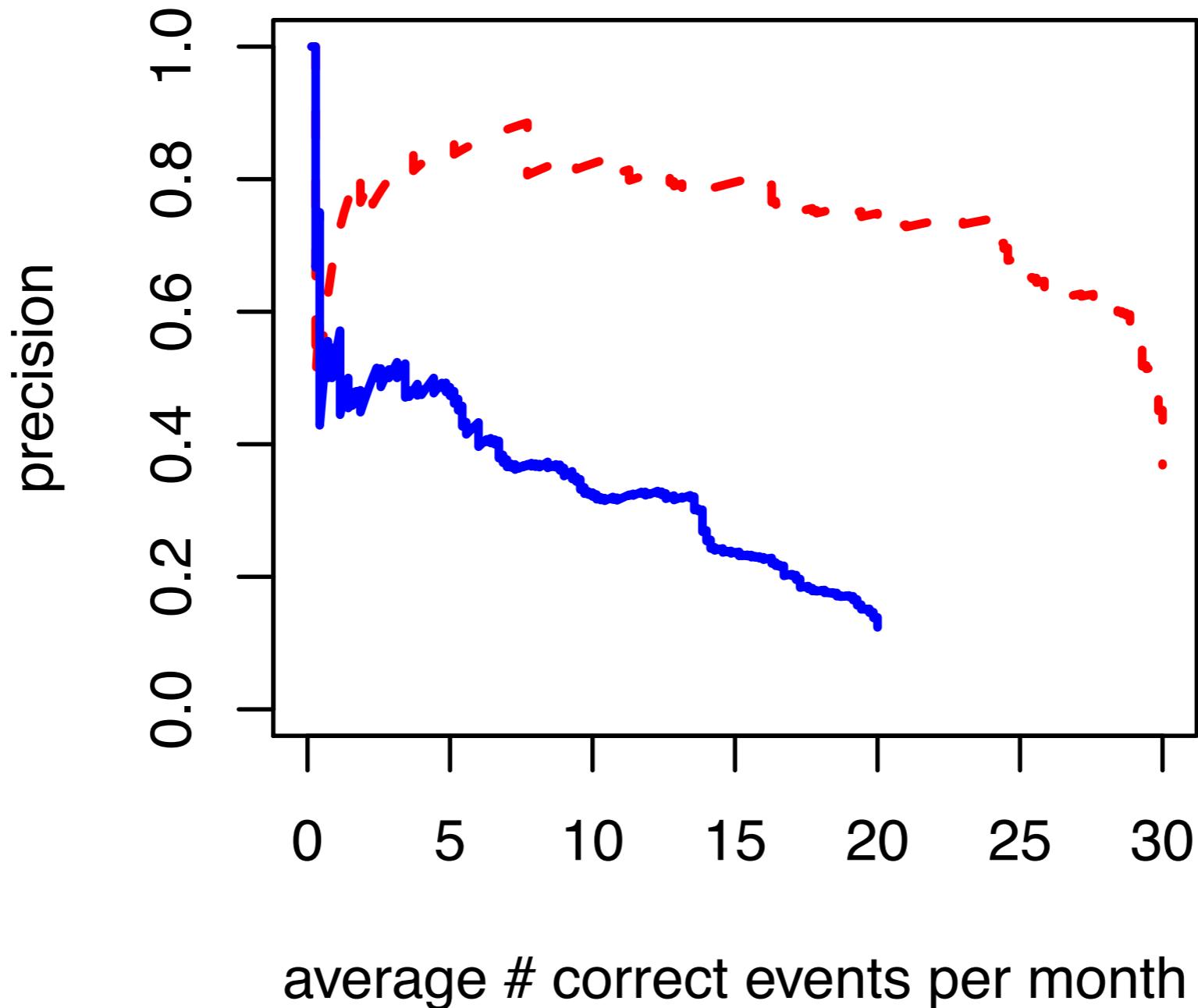


Learning to Extract Events

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



Results



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:

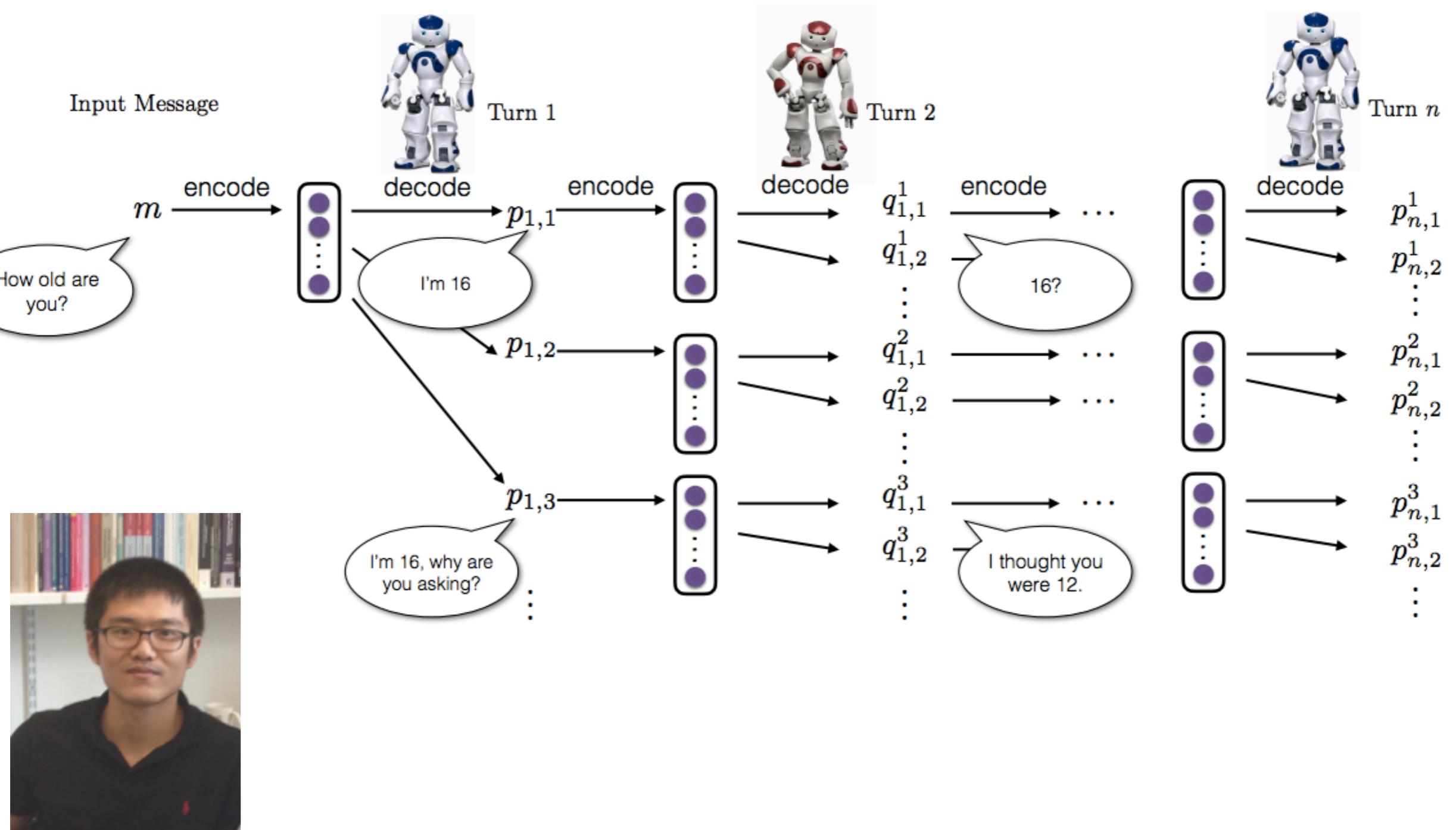
“I don’t know”

“OK”

“I’m sorry”

“I love you”

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