Building and Leveraging Knowledge Bases for Science

Andrew McCallum

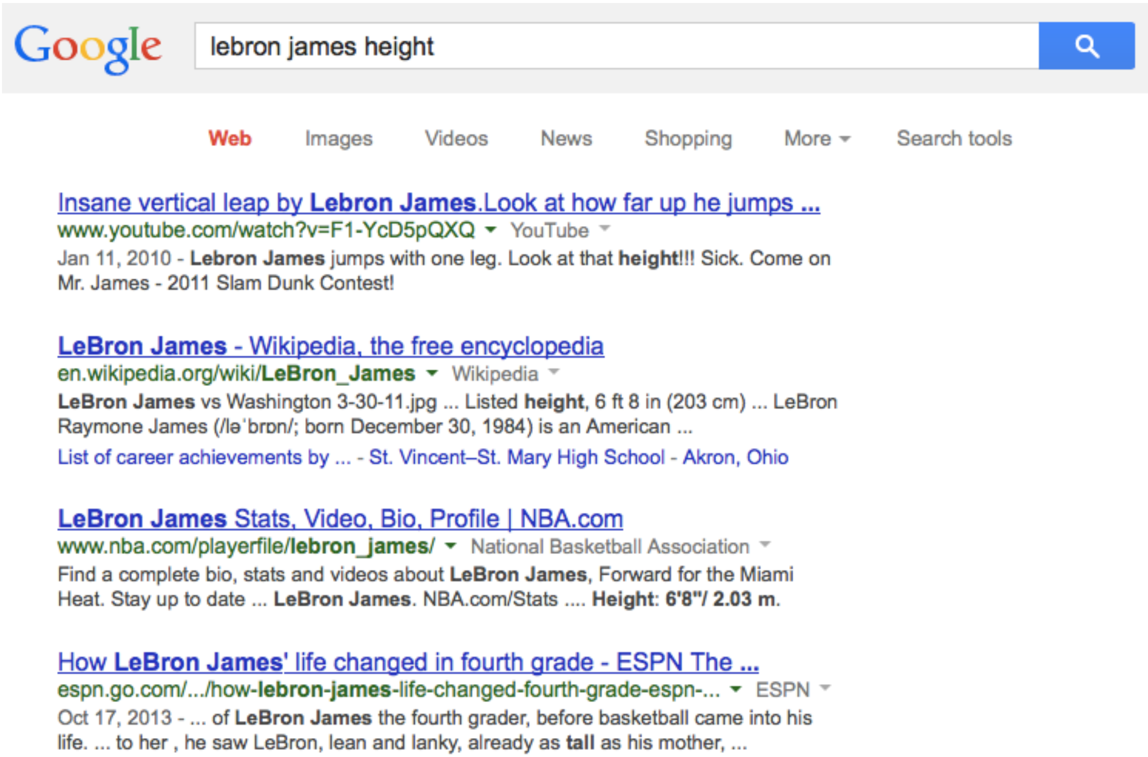
Center for Data Science
College of Information and Computer Sciences
University of Massachusetts Amherst

Joint work with
Sebastian Riedel, Limin Yao, Arvind Neelakantan, Patrick Verga, Rajarshi Das.

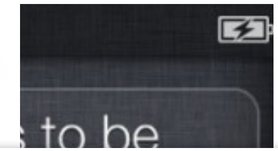
Web page search

~~Web page search~~ → dialog, QA, KB

Structured knowledge of world.

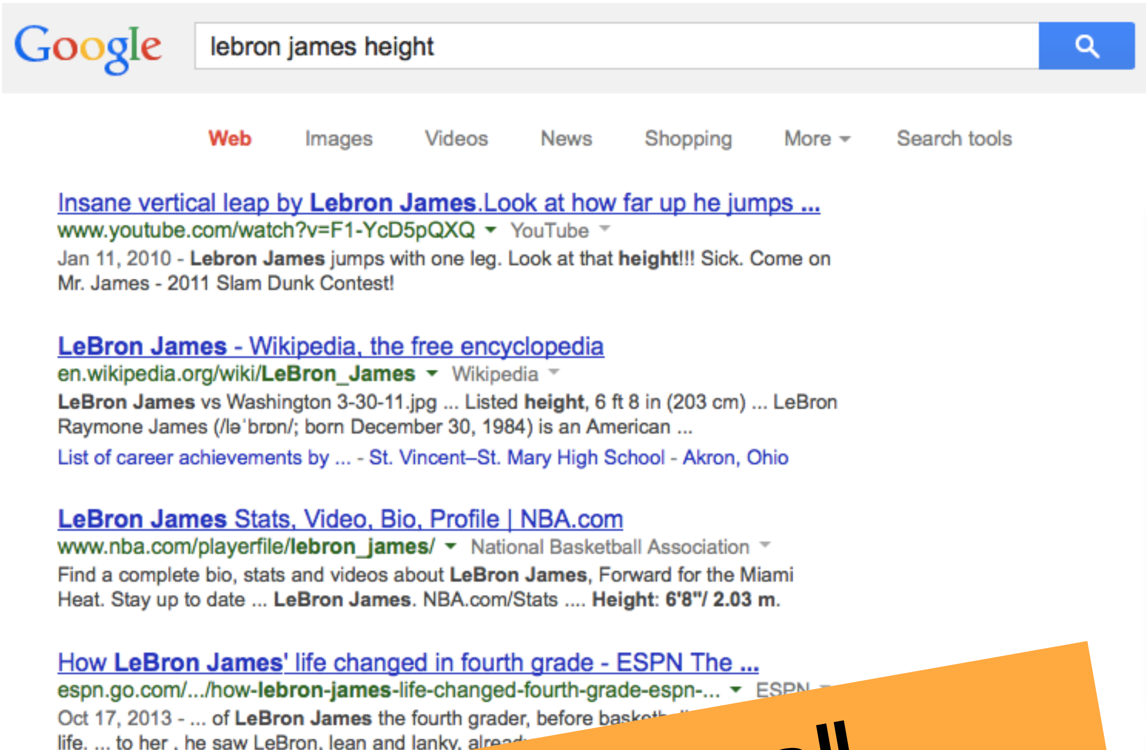


Freebase

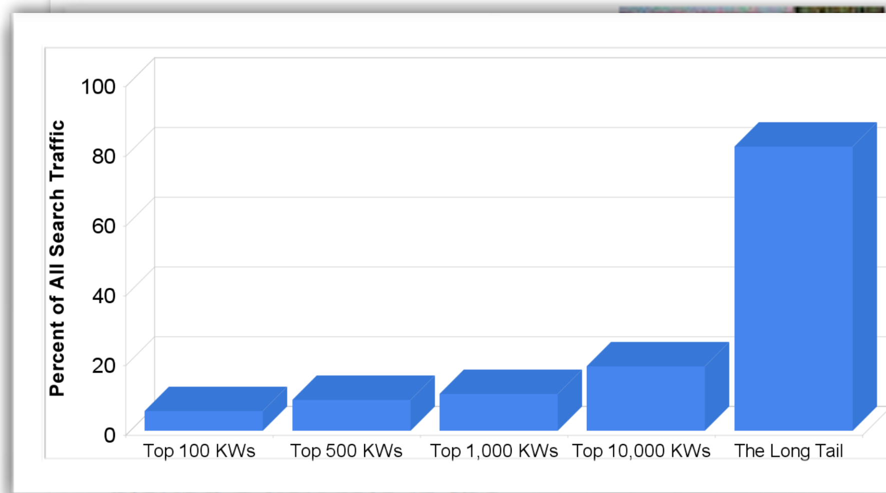


~~Web page search~~ → dialog, QA, KB

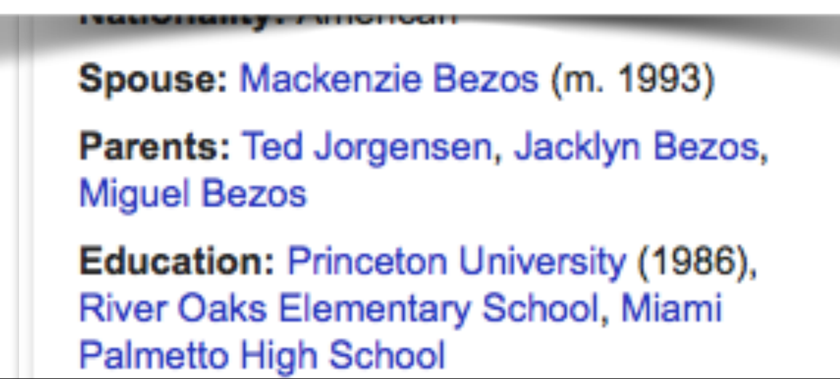
Structured knowledge of world.



Freebase

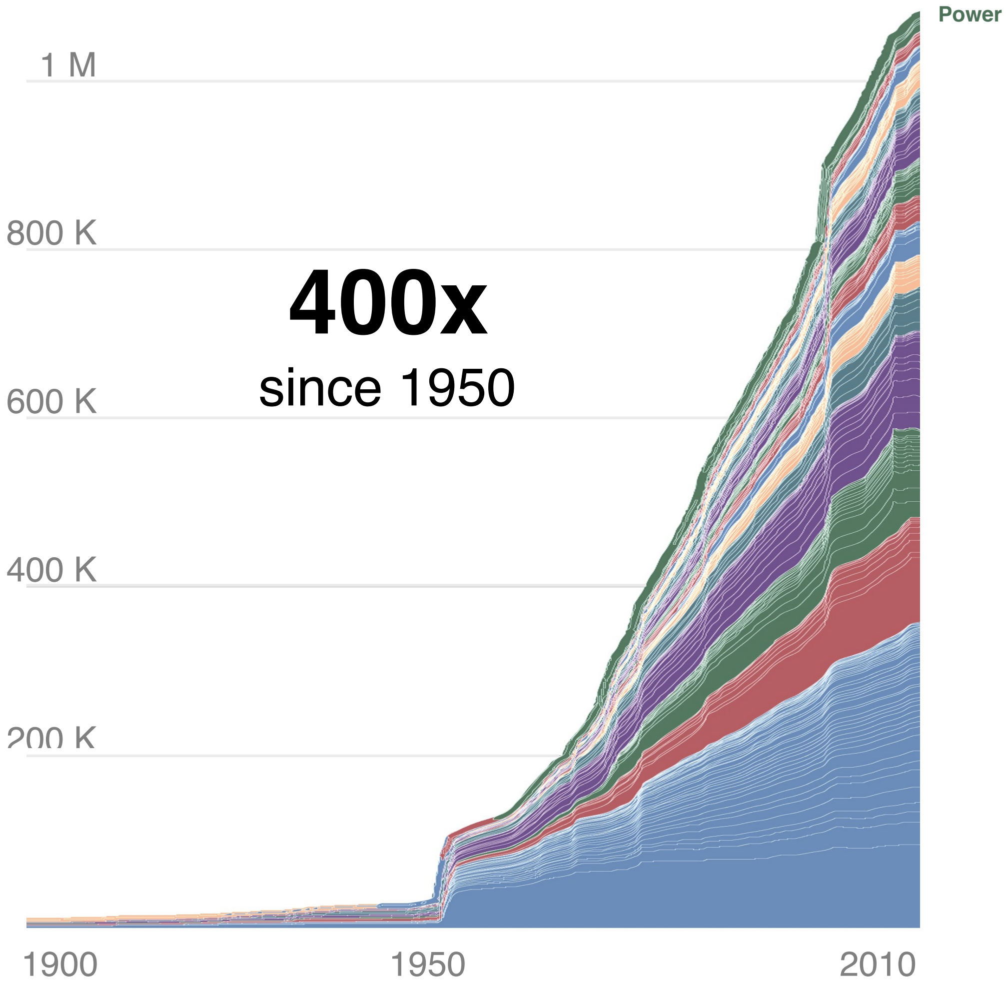


"Who did Bezos criticize?"

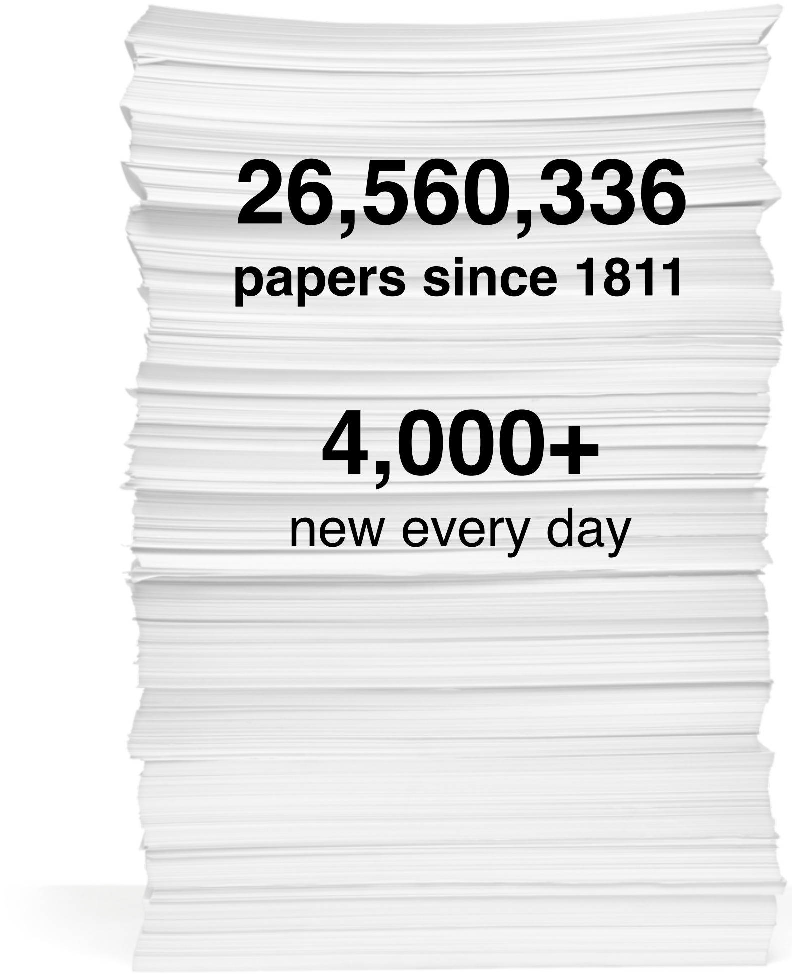


"open schema"
KB (of Science!)
with entity-relation structure?
...and reasoning?

Scientific Literature Growth



Graph: ReLX Group



Text

Knowledge Base

Reasoning

Scientific Text

↓

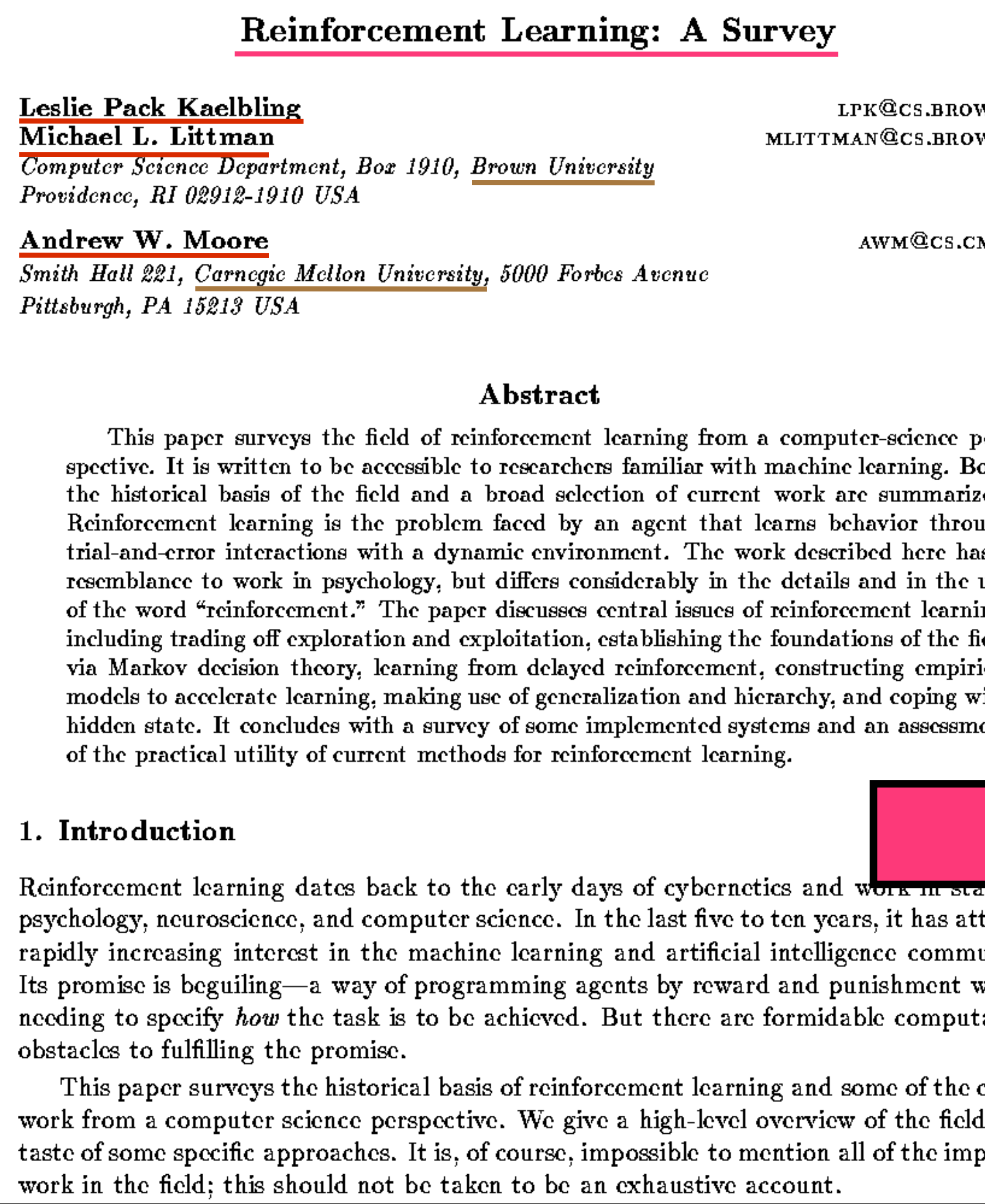
Scientific Knowledge Base

↓

Scientific Reasoning

Cora: KB of Research Papers

[McCallum et al 1996]



Reinforcement Learning: A Survey

Leslie Pack Kaelbling LPK@CS.BROW
Michael L. Littman MLITTMAN@CS.BROW
Computer Science Department, Box 1910, Brown University
Providence, RI 02912-1910 USA

Andrew W. Moore AWM@CS.CM
Smith Hall 221, Carnegie Mellon University, 5000 Forbes Avenue
Pittsburgh, PA 15213 USA

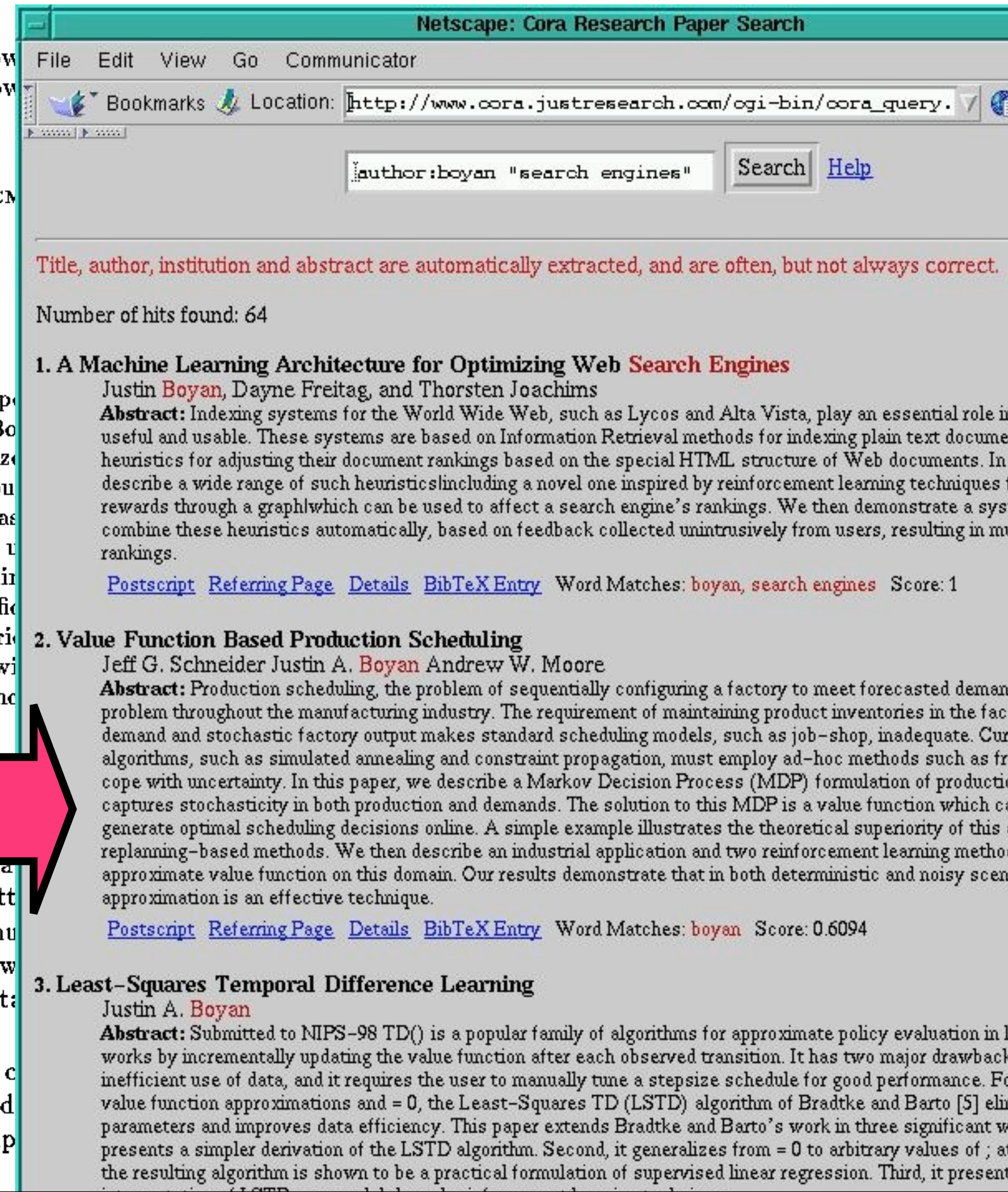
Abstract

This paper surveys the field of reinforcement learning from a computer-science pe spective. It is written to be accessible to researchers familiar with machine learning. Bo the historical basis of the field and a broad selection of current work are summariz Reinforcement learning is the problem faced by an agent that learns behavior throu trial-and-error interactions with a dynamic environment. The work described here has resemblance to work in psychology, but differs considerably in the details and in the u of the word "reinforcement." The paper discusses central issues of reinforcement learni including trading off exploration and exploitation, establishing the foundations of the fie via Markov decision theory, learning from delayed reinforcement, constructing empiri models to accelerate learning, making use of generalization and hierarchy, and coping wi hidden state. It concludes with a survey of some implemented systems and an assessme of the practical utility of current methods for reinforcement learning.

1. Introduction

Reinforcement learning dates back to the early days of cybernetics and work in sta psychology, neuroscience, and computer science. In the last five to ten years, it has att rapidly increasing interest in the machine learning and artificial intelligence commu Its promise is beguiling—a way of programming agents by reward and punishment w needing to specify *how* the task is to be achieved. But there are formidable computa obstacles to fulfilling the promise.

This paper surveys the historical basis of reinforcement learning and some of the c work from a computer science perspective. We give a high-level overview of the field taste of some specific approaches. It is, of course, impossible to mention all of the imp work in the field; this should not be taken to be an exhaustive account.



Netscape: Cora Research Paper Search

File Edit View Go Communicator

Bookmarks Location: http://www.cora.justresearch.com/cgi-bin/cora_query.

author:boyan "search engines" Search Help

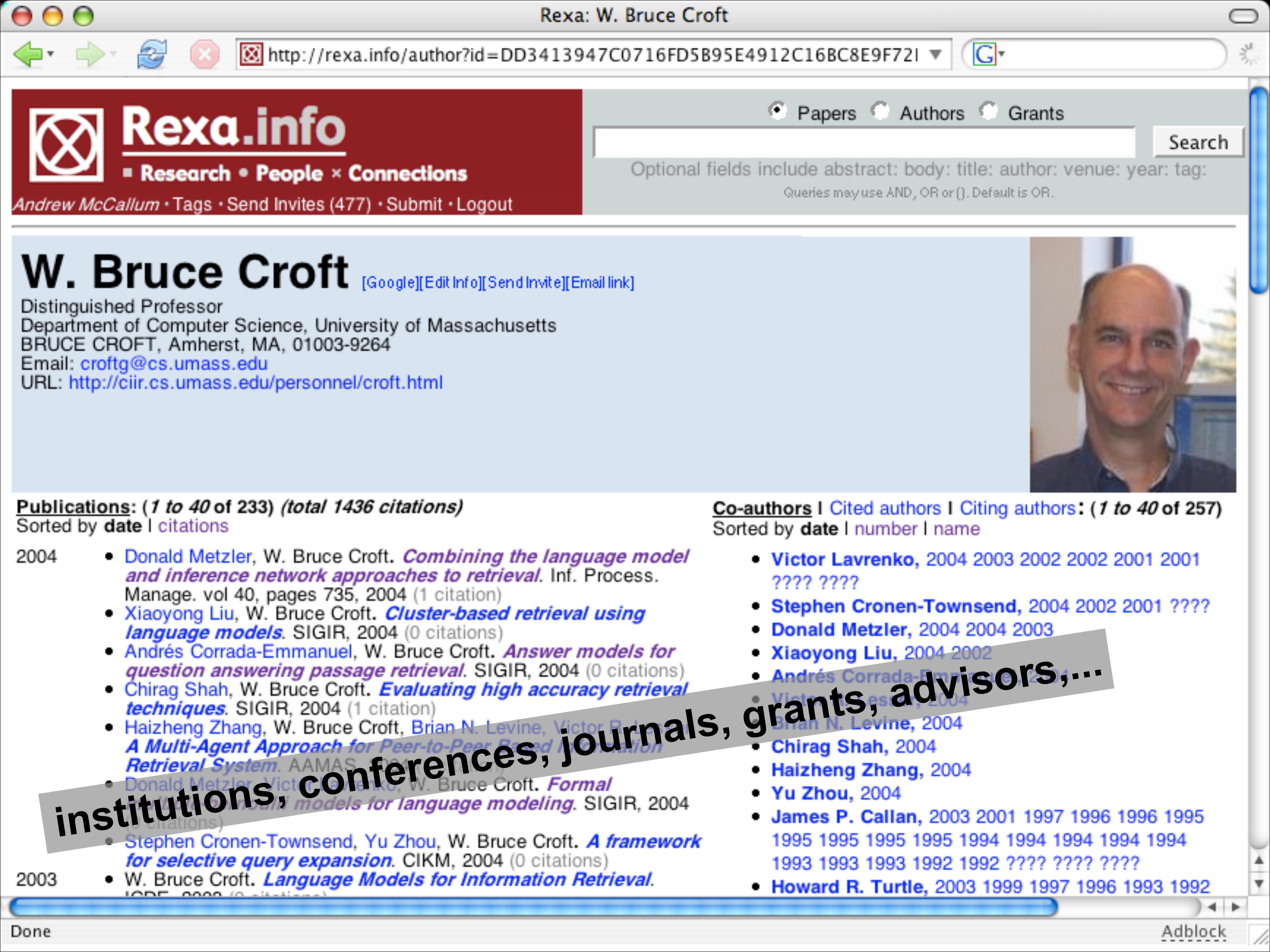
Title, author, institution and abstract are automatically extracted, and are often, but not always correct.

Number of hits found: 64

1. A Machine Learning Architecture for Optimizing Web Search Engines
Justin Boyan, Dayne Freitag, and Thorsten Joachims
Abstract: Indexing systems for the World Wide Web, such as Lycos and Alta Vista, play an essential role in m useful and usable. These systems are based on Information Retrieval methods for indexing plain text document heuristics for adjusting their document rankings based on the special HTML structure of Web documents. In th describe a wide range of such heuristicslincluding a novel one inspired by reinforcement learning techniques for rewards through a graphlwhich can be used to affect a search engine's rankings. We then demonstrate a syste combine these heuristics automatically, based on feedback collected unintrusively from users, resulting in muc rankings.
Postscript Referring Page Details BibTeX Entry Word Matches: boyan, search engines Score: 1

2. Value Function Based Production Scheduling
Jeff G. Schneider Justin A. Boyan Andrew W. Moore
Abstract: Production scheduling, the problem of sequentially configuring a factory to meet forecasted demands problem throughout the manufacturing industry. The requirement of maintaining product inventories in the face demand and stochastic factory output makes standard scheduling models, such as job-shop, inadequate. Curre algorithms, such as simulated annealing and constraint propagation, must employ ad-hoc methods such as freq cope with uncertainty. In this paper, we describe a Markov Decision Process (MDP) formulation of production captures stochasticity in both production and demands. The solution to this MDP is a value function which can generate optimal scheduling decisions online. A simple example illustrates the theoretical superiority of this ap replanning-based methods. We then describe an industrial application and two reinforcement learning methods approximate value function on this domain. Our results demonstrate that in both deterministic and noisy scenar approximation is an effective technique.
Postscript Referring Page Details BibTeX Entry Word Matches: boyan Score: 0.6094

3. Least-Squares Temporal Difference Learning
Justin A. Boyan
Abstract: Submitted to NIPS-98 TD() is a popular family of algorithms for approximate policy evaluation in lar works by incrementally updating the value function after each observed transition. It has two major drawbacks: inefficient use of data, and it requires the user to manually tune a stepsize schedule for good performance. For value function approximations and = 0, the Least-Squares TD (LSTD) algorithm of Bradtke and Barto [5] elimi parameters and improves data efficiency. This paper extends Bradtke and Barto's work in three significant way presents a simpler derivation of the LSTD algorithm. Second, it generalizes from = 0 to arbitrary values of ; at t the resulting algorithm is shown to be a practical formulation of supervised linear regression. Third, it presents



Rexa: W. Bruce Croft

http://rexa.info/author?id=DD3413947C0716FD5B95E4912C16BC8E9F72I

Rexa.info

Research • People × Connections

Andrew McCallum • Tags • Send Invites (477) • Submit • Logout

Papers Authors Grants

Search

Optional fields include abstract: body: title: author: venue: year: tag:

Queries may use AND, OR or (). Default is OR.

W. Bruce Croft [Google][Edit Info][Send Invite][Email link]

Distinguished Professor
Department of Computer Science, University of Massachusetts
BRUCE CROFT, Amherst, MA, 01003-9264
Email: croftg@cs.umass.edu
URL: http://ciir.cs.umass.edu/personnel/croft.html

Publications: (*1 to 40* of 233) *(total 1436 citations)*
Sorted by **date** | citations

2004
- Donald Metzler, W. Bruce Croft. ***Combining the language model and inference network approaches to retrieval***. Inf. Process. Manage. vol 40, pages 735, 2004 (1 citation)
- Xiaoyong Liu, W. Bruce Croft. ***Cluster-based retrieval using language models***. SIGIR, 2004 (0 citations)
- Andrés Corrada-Emmanuel, W. Bruce Croft. ***Answer models for question answering passage retrieval***. SIGIR, 2004 (0 citations)
- Chirag Shah, W. Bruce Croft. ***Evaluating high accuracy retrieval techniques***. SIGIR, 2004 (1 citation)
- Haizheng Zhang, W. Bruce Croft, Brian N. Levine, Victor R. ... ***A Multi-Agent Approach for Peer-to-Peer Based Information Retrieval System***. AAMAS, ...
- Donald Metzler, Victor Lavrenko, W. Bruce Croft. ***Formal ... models for language modeling***. SIGIR, 2004 (0 citations)
- Stephen Cronen-Townsend, Yu Zhou, W. Bruce Croft. ***A framework for selective query expansion***. CIKM, 2004 (0 citations)

2003
- W. Bruce Croft. ***Language Models for Information Retrieval***. ICDE, 2003 (0 citations)

Co-authors | Cited authors | Citing authors: (*1 to 40* of 257)
Sorted by **date** | number | name

- **Victor Lavrenko**, 2004 2003 2002 2002 2001 2001 ???? ????
- **Stephen Cronen-Townsend**, 2004 2002 2001 ????
- **Donald Metzler**, 2004 2004 2003
- **Xiaoyong Liu**, 2004 2002
- **Andrés Corrada-Emmanuel**, 2004
- **Victor ...**, 2004
- **Brian N. Levine**, 2004
- **Chirag Shah**, 2004
- **Haizheng Zhang**, 2004
- **Yu Zhou**, 2004
- **James P. Callan**, 2003 2001 1997 1996 1996 1995 1995 1995 1995 1995 1994 1994 1994 1994 1994 1993 1993 1993 1992 1992 ???? ???? ????
- **Howard R. Turtle**, 2003 1999 1997 1996 1993 1992

Done Adblock

institutions, conferences, journals, grants, advisors,...

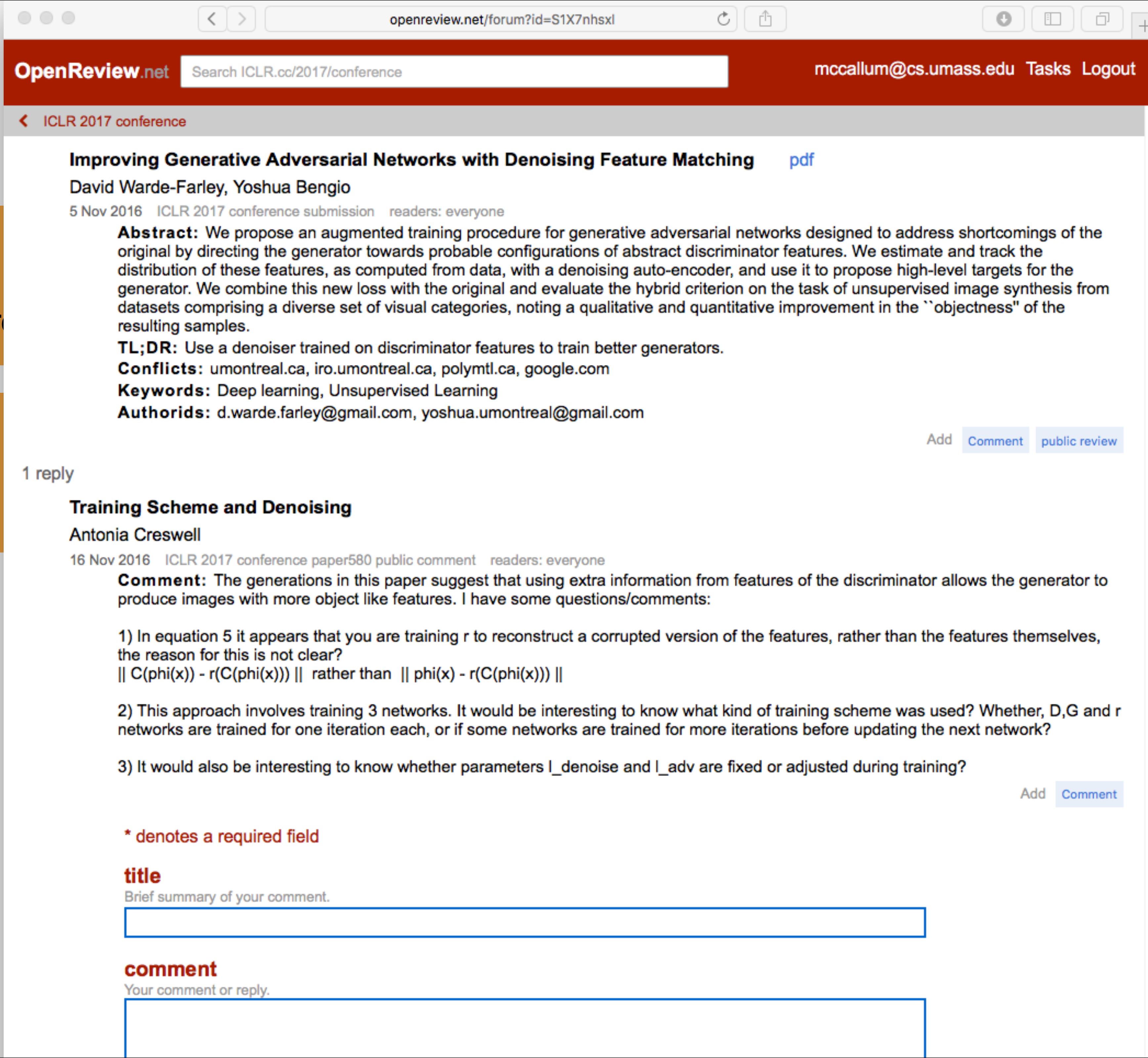
Application Goals

A KB of all scientists in the world
from papers, patents, web pages, newswire, press releases, tweets, blogs,...

A KB of scientific entities & relations
materials, equipment, organisms, processes, tasks, methods,...

- Better tools → Accelerate progress of science.
- Revolutionize peer review
  - “open peer review”
  - Submission, reviews & comments public.

fr …



openreview.net/forum?id=S1X7nhsxl

OpenReview.net Search ICLR.cc/2017/conference mccallum@cs.umass.edu Tasks Logout

ICLR 2017 conference

Improving Generative Adversarial Networks with Denoising Feature Matching pdf

David Warde-Farley, Yoshua Bengio

5 Nov 2016 ICLR 2017 conference submission readers: everyone

Abstract: We propose an augmented training procedure for generative adversarial networks designed to address shortcomings of the original by directing the generator towards probable configurations of abstract discriminator features. We estimate and track the distribution of these features, as computed from data, with a denoising auto-encoder, and use it to propose high-level targets for the generator. We combine this new loss with the original and evaluate the hybrid criterion on the task of unsupervised image synthesis from datasets comprising a diverse set of visual categories, noting a qualitative and quantitative improvement in the ``objectness'' of the resulting samples.

TL;DR: Use a denoiser trained on discriminator features to train better generators.

Conflicts: umontreal.ca, iro.umontreal.ca, polymtl.ca, google.com

Keywords: Deep learning, Unsupervised Learning

Authorids: d.warde.farley@gmail.com, yoshua.umontreal@gmail.com

Add Comment public review

1 reply

Training Scheme and Denoising

Antonia Creswell

16 Nov 2016 ICLR 2017 conference paper580 public comment readers: everyone

Comment: The generations in this paper suggest that using extra information from features of the discriminator allows the generator to produce images with more object like features. I have some questions/comments:

1) In equation 5 it appears that you are training r to reconstruct a corrupted version of the features, rather than the features themselves, the reason for this is not clear?
|| C(phi(x)) - r(C(phi(x))) || rather than || phi(x) - r(C(phi(x))) ||

2) This approach involves training 3 networks. It would be interesting to know what kind of training scheme was used? Whether, D,G and r networks are trained for one iteration each, or if some networks are trained for more iterations before updating the next network?

3) It would also be interesting to know whether parameters l_denoise and l_adv are fixed or adjusted during training?

Add Comment

* denotes a required field

title

Brief summary of your comment.

comment

Your comment or reply.

ce.

Text

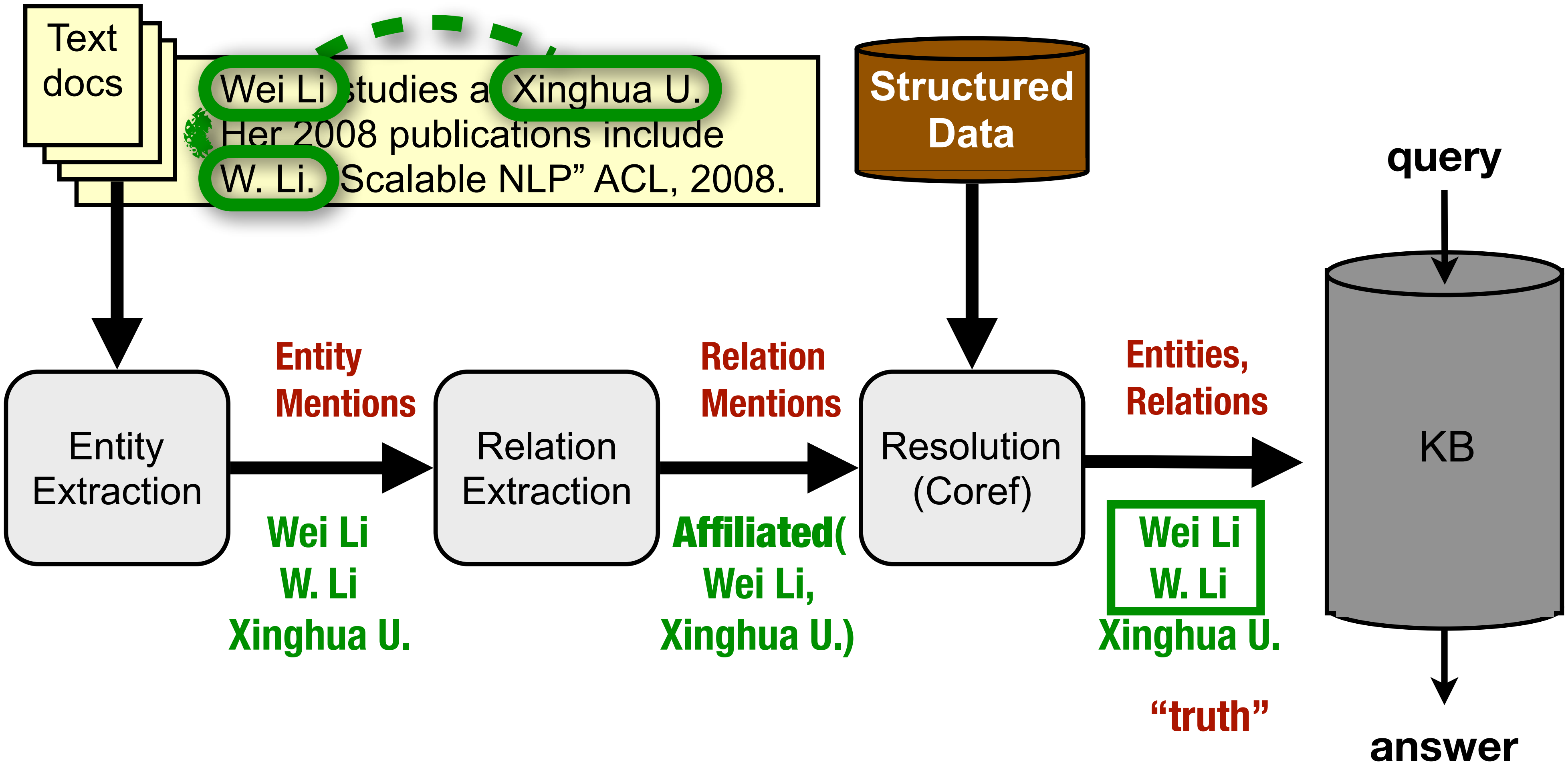
↓

Knowledge Base

↓

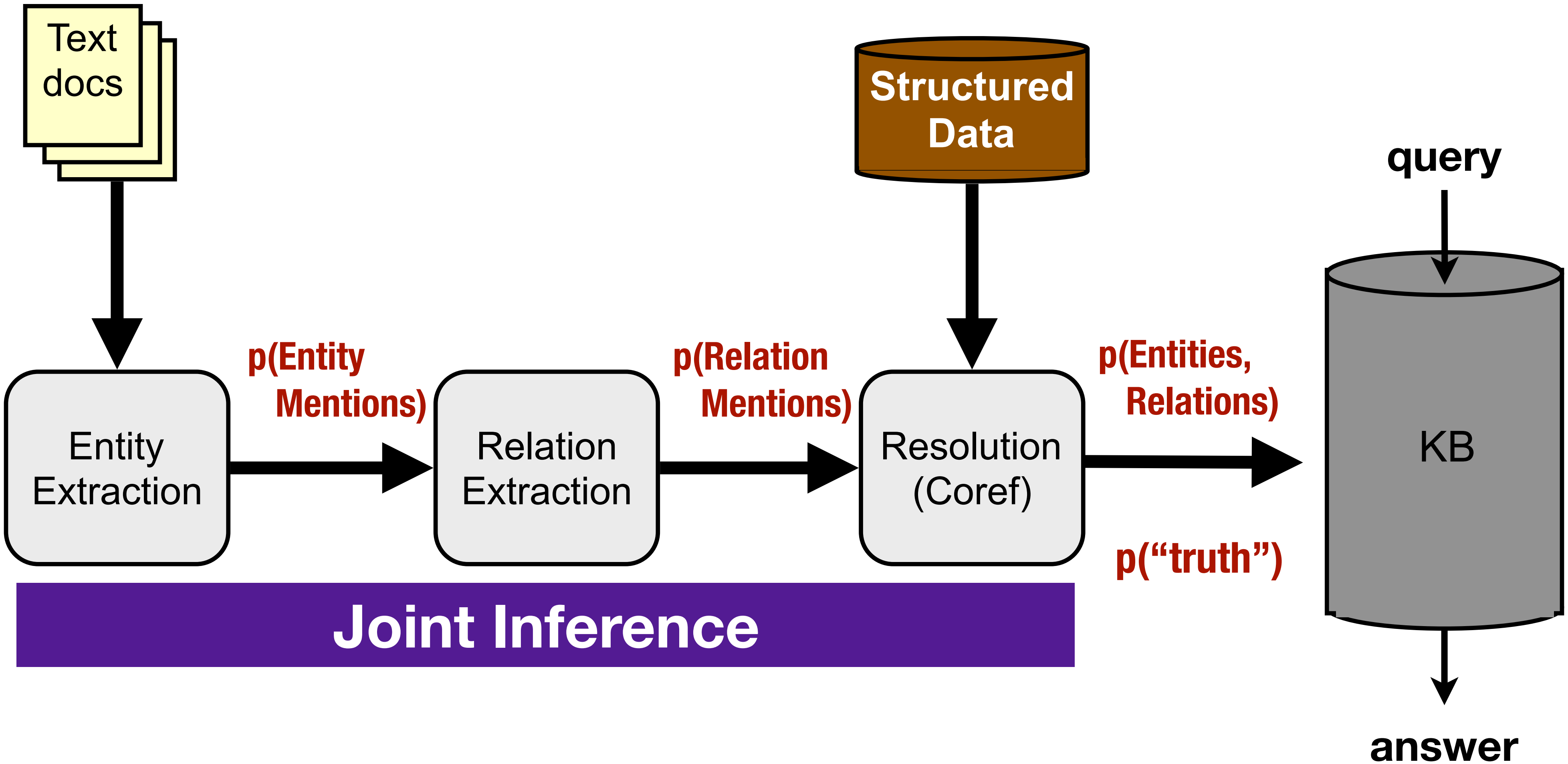
Reasoning

Knowledge Base Construction



Information Extraction components aren't perfect.
Errors snowball.

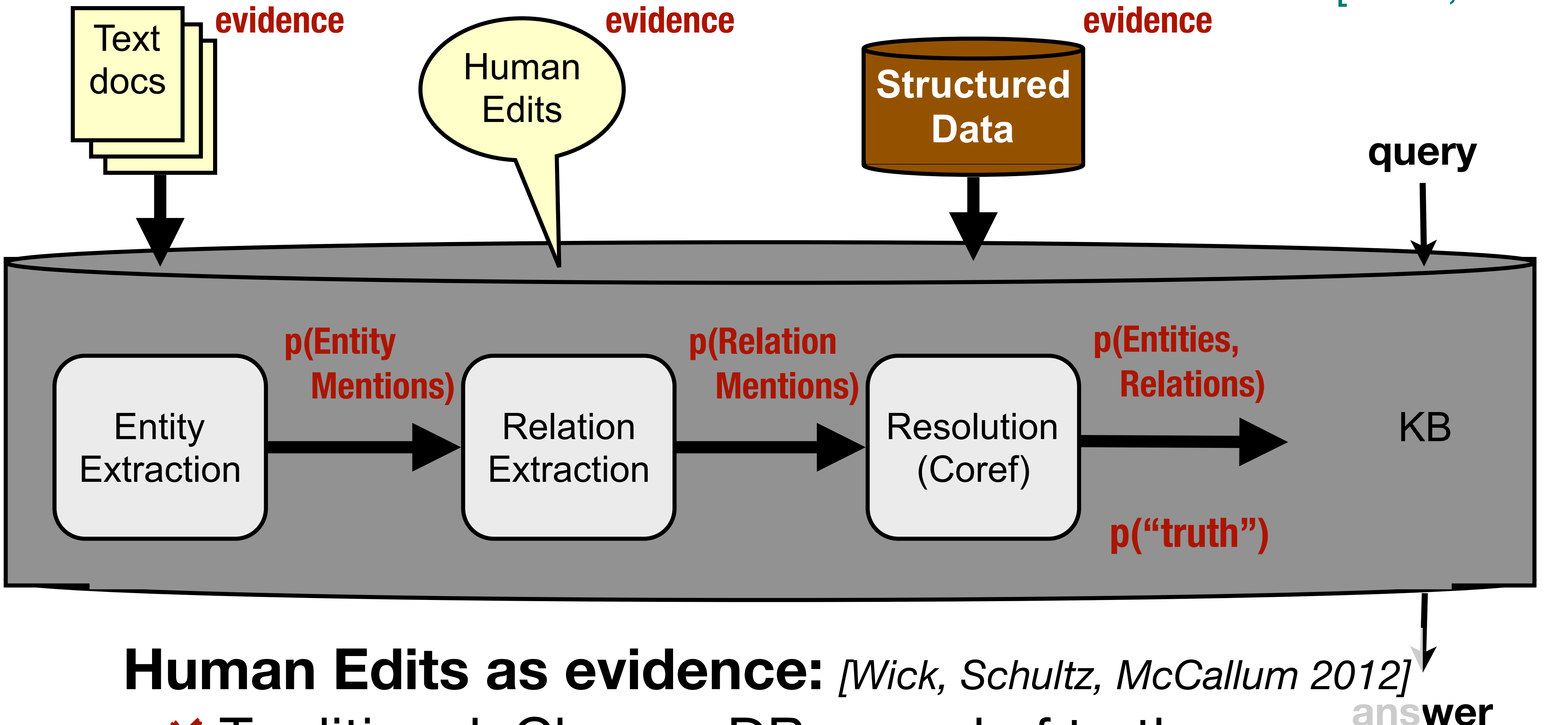
Knowledge Base Construction



1. How to represent & inject uncertainty from IE into KB?
2. How to use KB contents to aid IE?
3. IE isn't "one-shot." Add new data later; redo inference. Want KB infrastructure to manage IE.

“Knowledge Base Construction”
“Epistemological Knowledge Base”

[2010, 2012]



Human Edits as evidence: *[Wick, Schultz, McCallum 2012]*

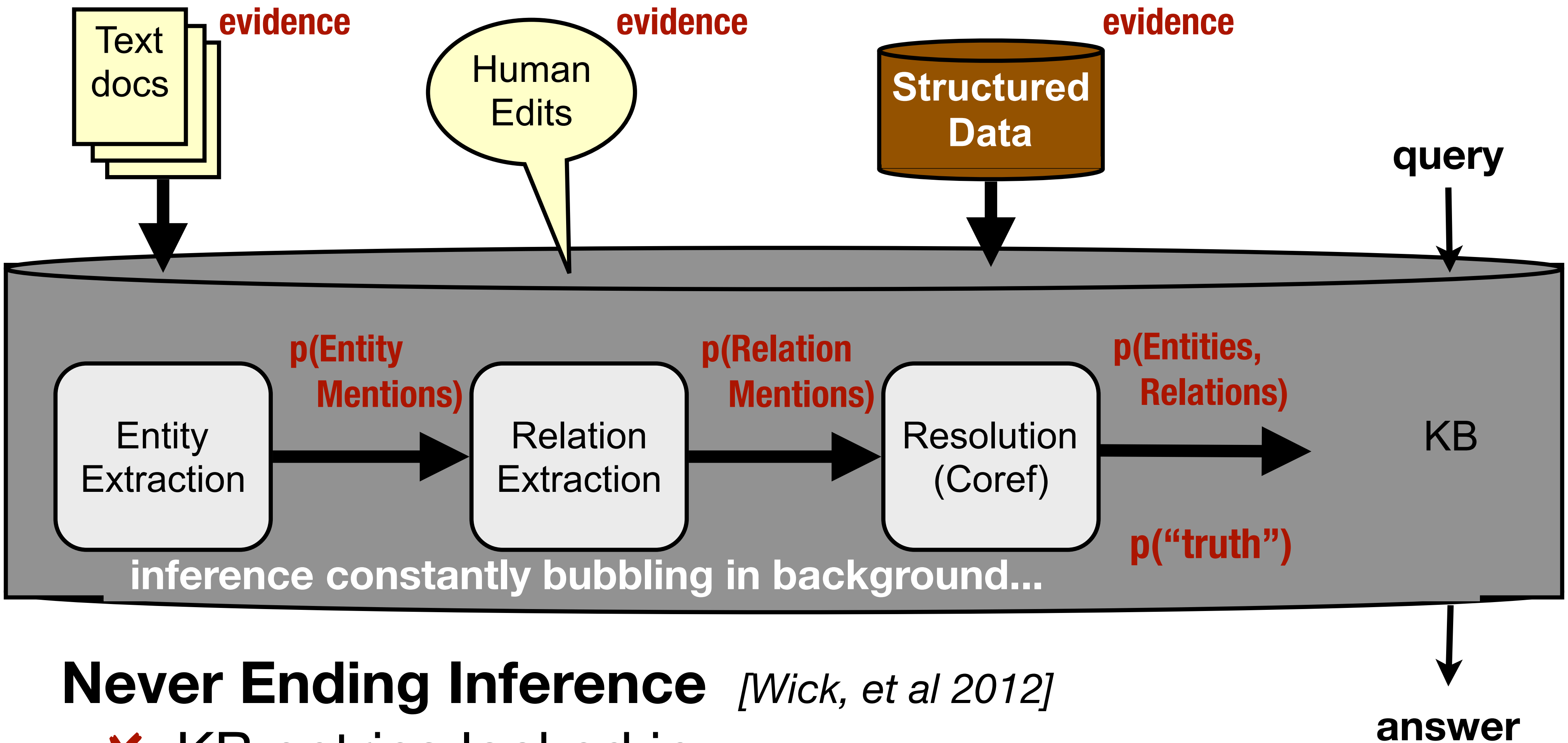
- ✘ Traditional: Change DB record of truth
- ✔ Mini-document *“Nov 15: Scott said this was true”*

- Sometimes humans are wrong, disagree, out-of-date.
- Jointly reason about truth & editors' reliability/reputation.

Epistemological Philosophy

“Truth is inferred, not observed.”

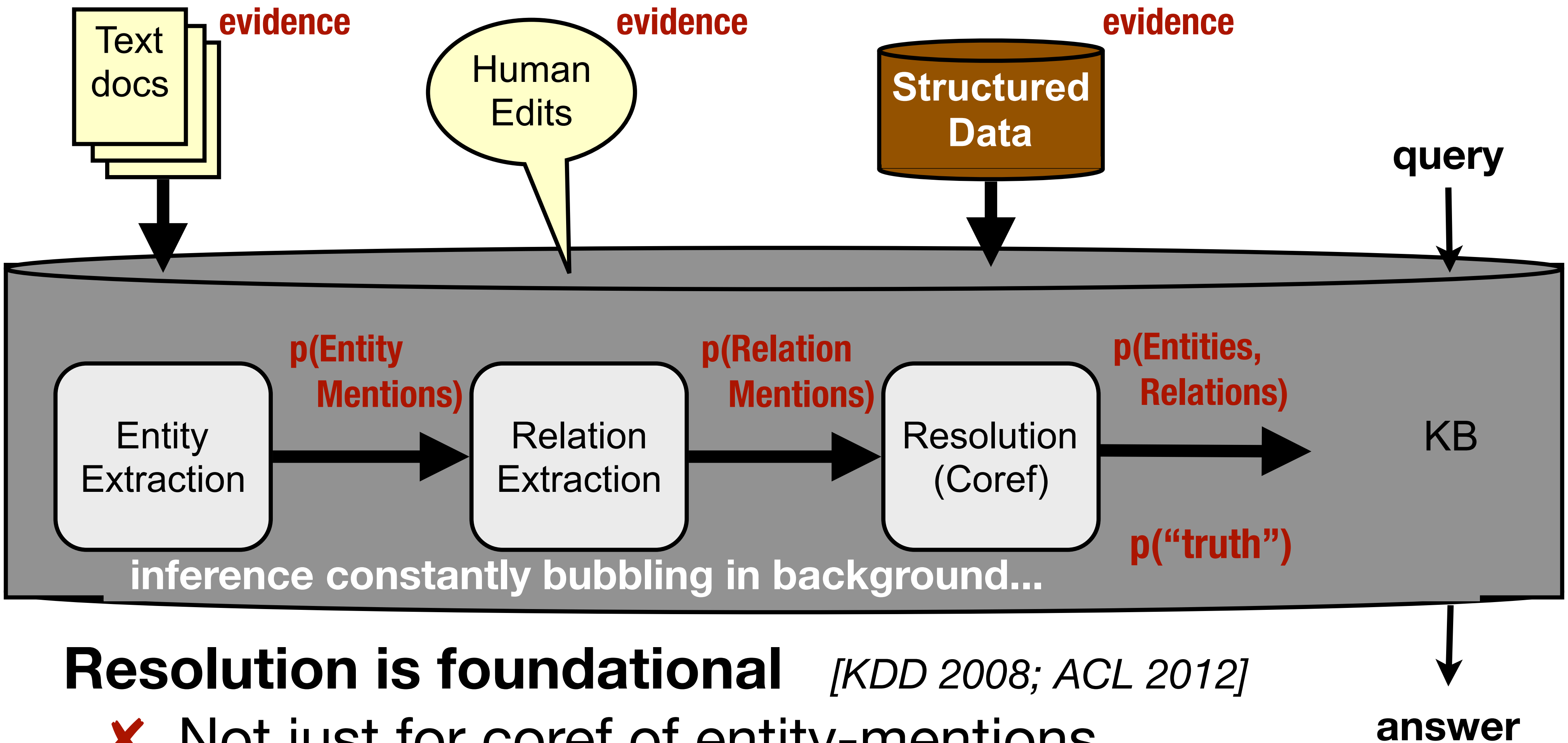
"Epistemological KnowledgeBase"



Never Ending Inference *[Wick, et al 2012]*

- ✘ KB entries locked in
- ✔ KB entries always reconsidered with more evidence, time,...

“Epistemological KnowledgeBase”



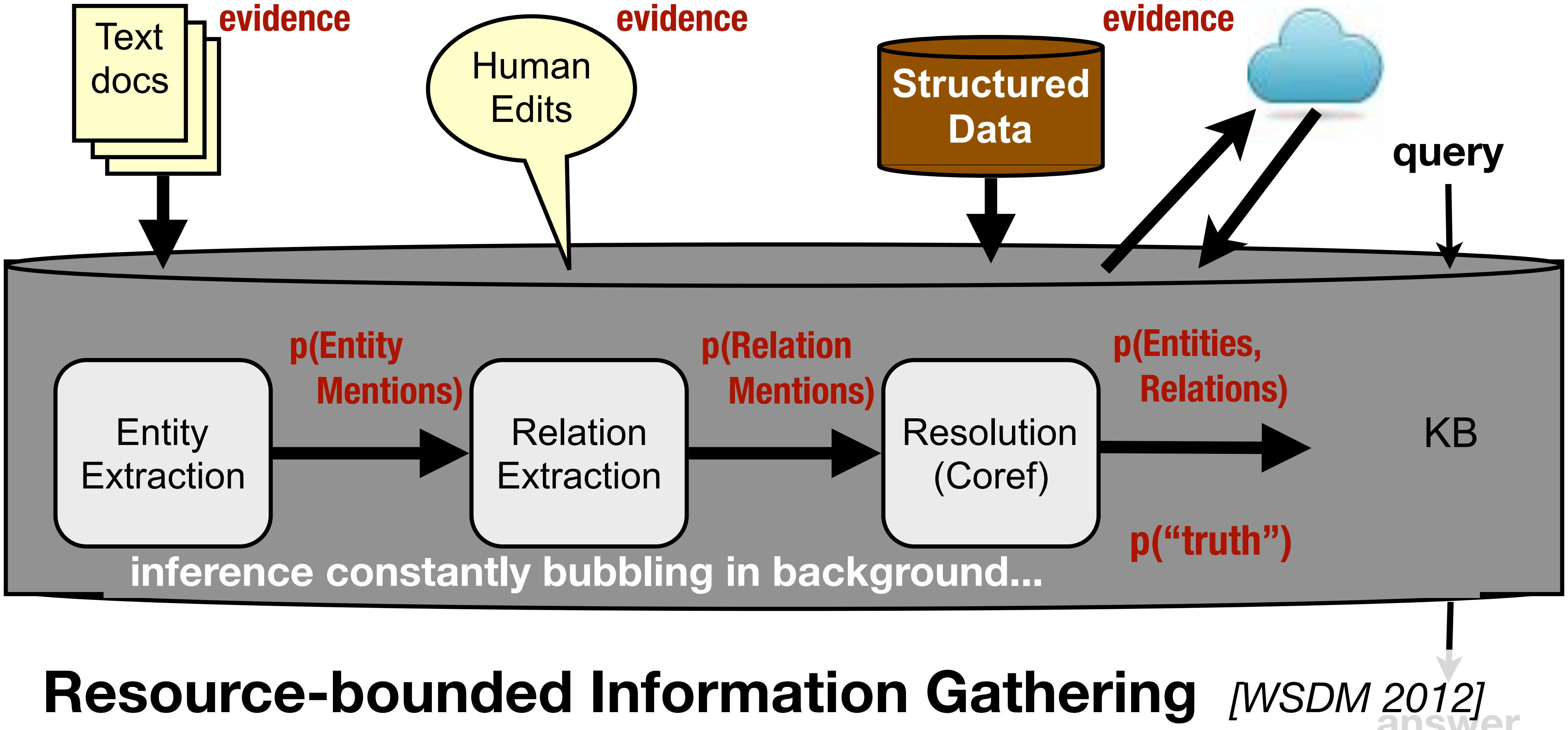
Resolution is foundational *[KDD 2008; ACL 2012]*

✘ Not just for coref of entity-mentions...

✔ Align values, ontologies, schemas, relations, events,...

Especially in Epistemological DB: entities/relations never input, only “mentions”

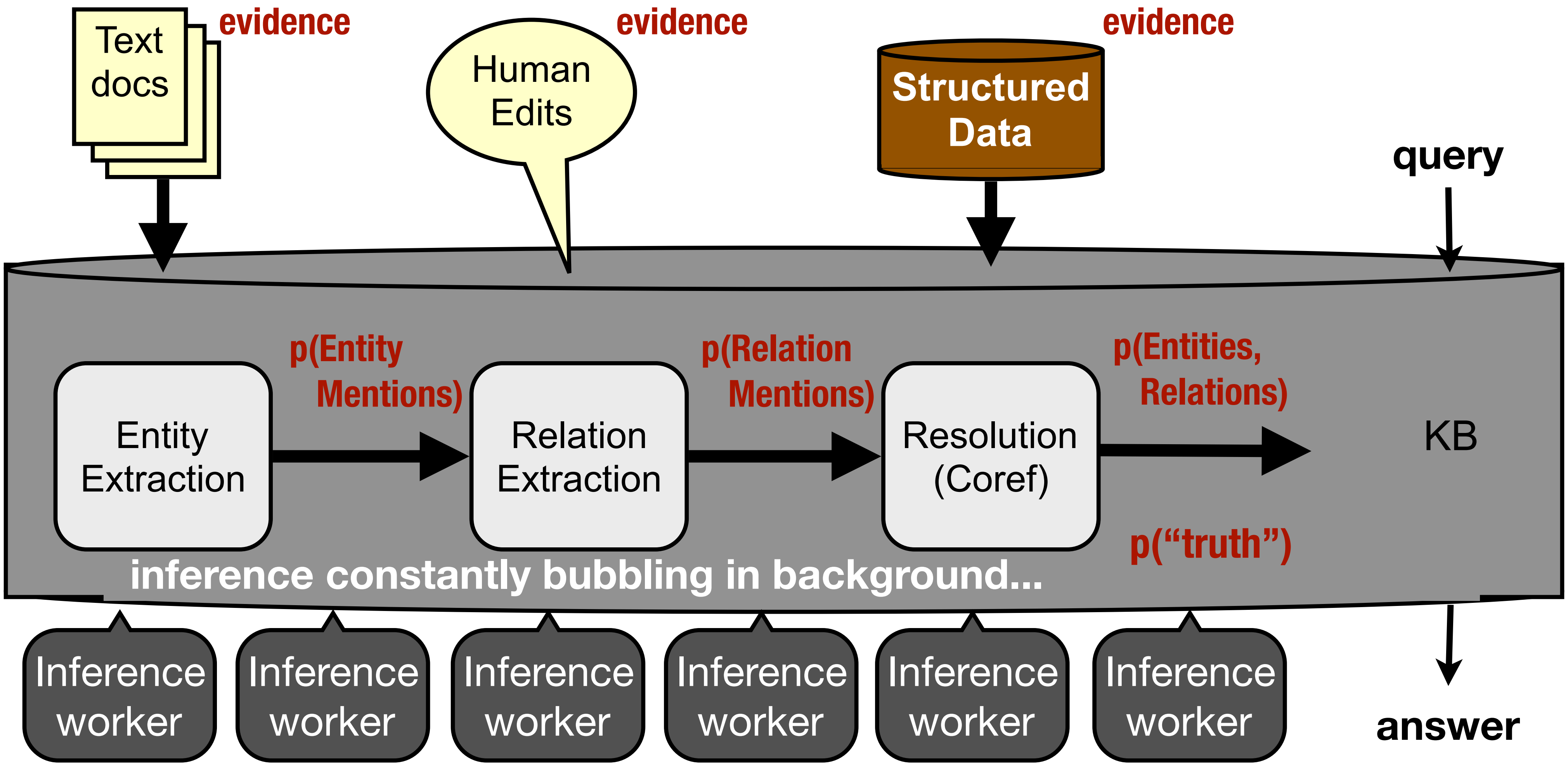
“Epistemological KnowledgeBase”



Resource-bounded Information Gathering *[WSDM 2012]*

- ✘ Full processing on whole web
- ✔ Focus queries and processing where needed & fruitful

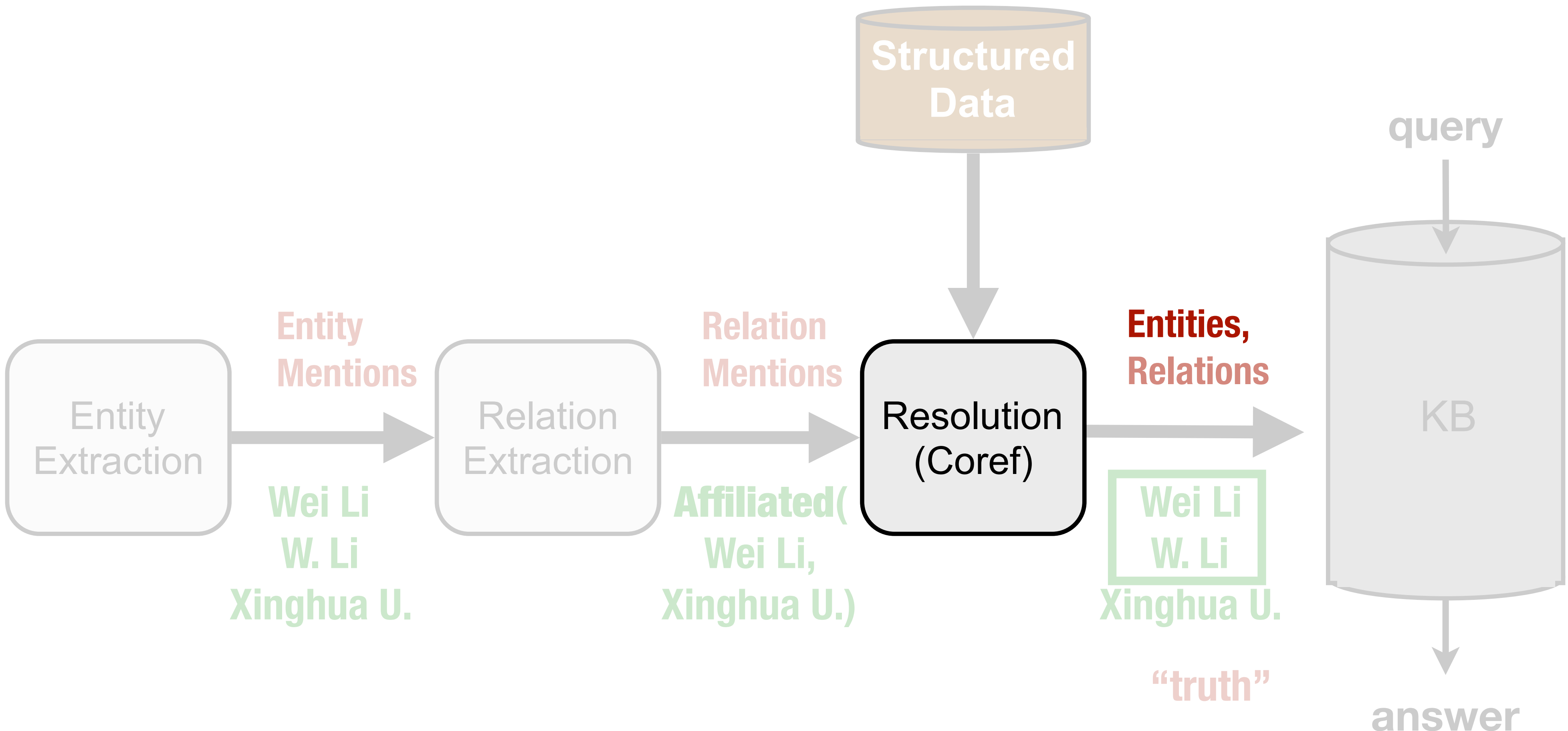
“Epistemological KnowledgeBase”



Smart Parallelism *[ACL 2011; NIPS 2011]*

- ✘ MapReduce, black-box
- ✔ Reason about inference & parallelism together

Entity Resolution



Author Entity Resolution

A. Banerjee, S. Chassang, E. Snowberg. *Decision Theoretic Approaches to Experiment Design and External Validity.* Handbook of Field Experiments. 2016.

Arindam Banerjee, S. Merugu, I. S. Dhillon, J. Ghosh. *Clustering with Bregman Divergences.* JMLR. 2006.

A. Banerjee, I. S. Dhillon, J. Ghosh, S. Sra. *Clustering on the Unit Hypersphere using von Mises-Fisher Distributions.*
Journal of Machine Learning Research. 2005

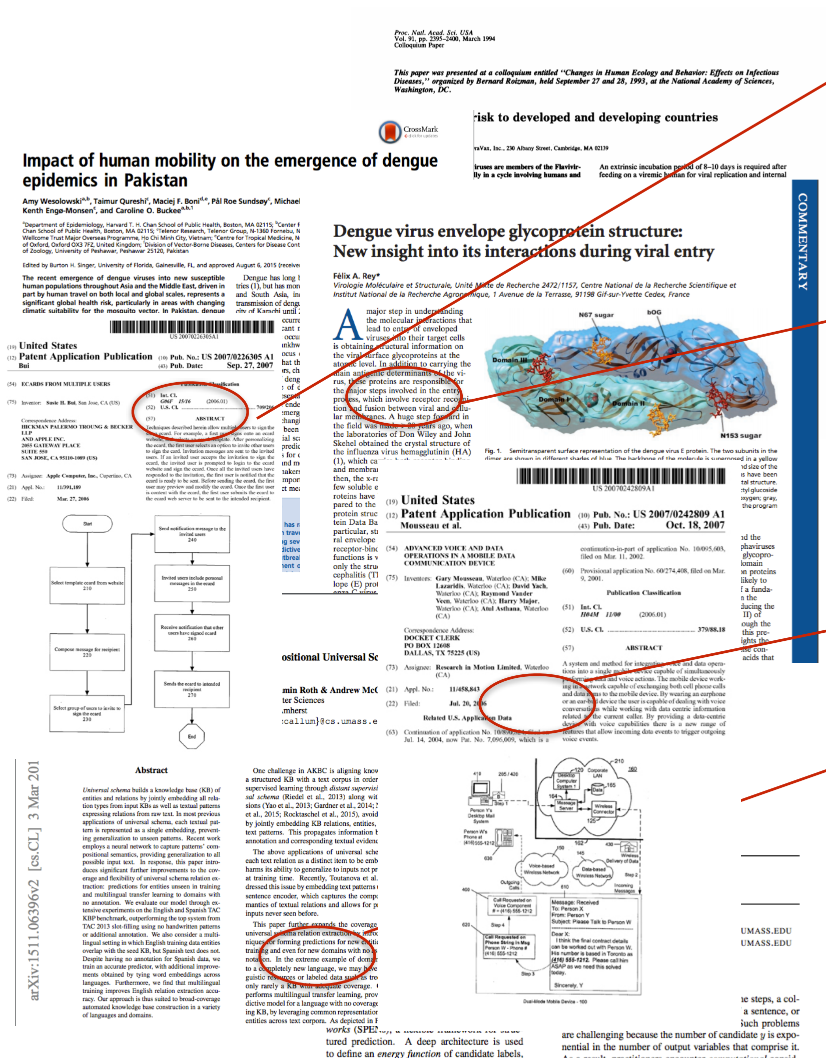
Author Entity Resolution

A. Banerjee, S. Chassang, E. Snowberg. *Decision Theoretic Approaches to Experiment Design and External Validity.* Handbook of Field Experiments. 2016.

Arindam Banerjee, S. Merugu, I. S. Dhillon, J. Ghosh. *Clustering with Bregman Divergences.* JMLR. 2006.

A. Banerjee, I. S. Dhillon, J. Ghosh, S. Sra. *Clustering on the Unit Hypersphere using von Mises-Fisher Distributions.*
Journal of Machine Learning Research. 2005

Scientific Entity Resolution



Dengue viruses are members of the **Flaviviridae**, transmitted principally in a cycle involving humans and mosquito vectors.

The virus-encoded RNA-dependent RNA polymerase (RdRp), which is required for replication of the positive-strand RNA genome, is a key enzyme of members of the virus family **Flaviviridae**

We present several modifications of the original ***recurrent neural network*** language model

Unlike Toutanova et al. (2015), we also consider **RNNs**, specifically Long-Short Term Memory Networks (LSTMs) (Hochreiter and Schmidhuber, 1997)

Entity Resolution as Clustering

Given **mentions** $M = \{m_1, m_2, \ldots, m_N\}$

Entity Resolution as Clustering

Given **mentions** $M = \{m_1, m_2, \ldots, m_N\}$

A. Banerjee, S. Chassang, E. Snowberg. *Decision Theoretic Approaches to Experiment Design and External Validity.* Handbook of Field Experiments. 2016.

Entity Resolution as Clustering

Given **mentions** $M = \{m_1, m_2, \ldots, m_N\}$

Entity Resolution as Clustering

Given **mentions** $M = \{m_1, m_2, \ldots, m_N\}$

Partition M into **entities** $E = \{e_1, e_2, \ldots, e_k\}$
where k unknown in advance

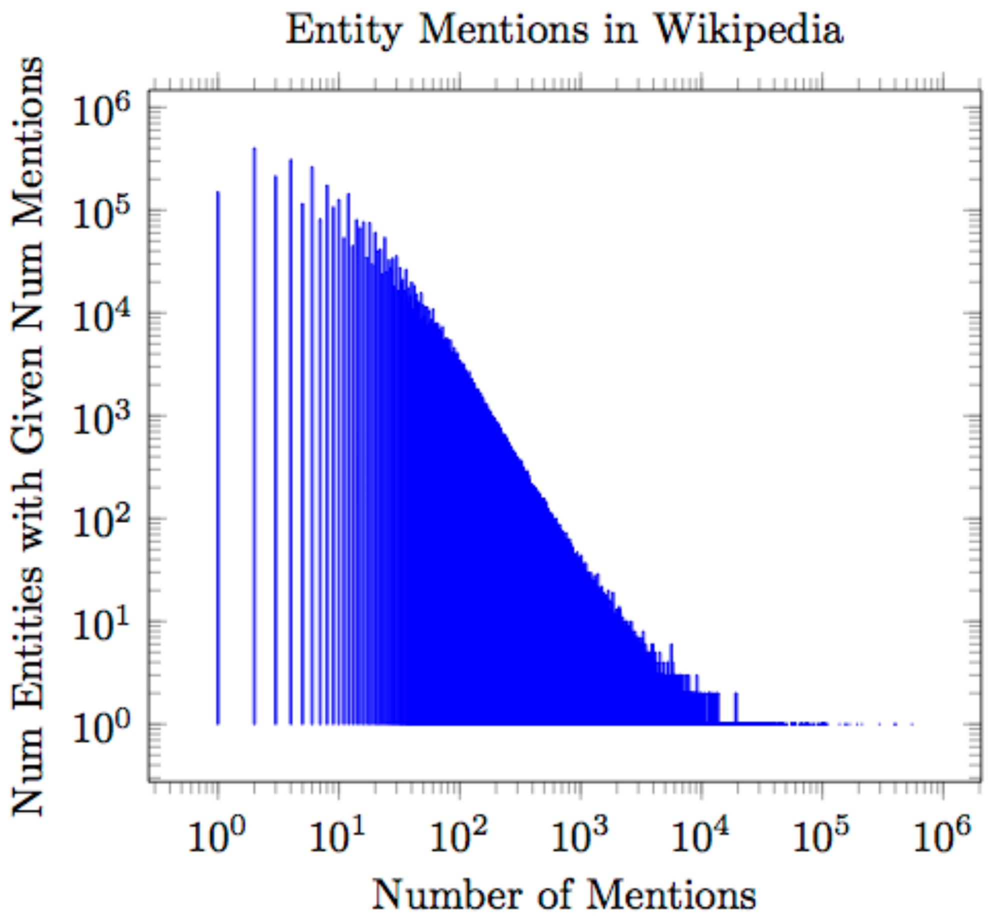
Entity Resolution as Clustering

Given **mentions** $M = \{m_1, m_2, \ldots, m_N\}$

Partition M into **entities** $E = \{e_1, e_2, \ldots, e_k\}$
where k unknown in advance

Entity Resolution Challenge

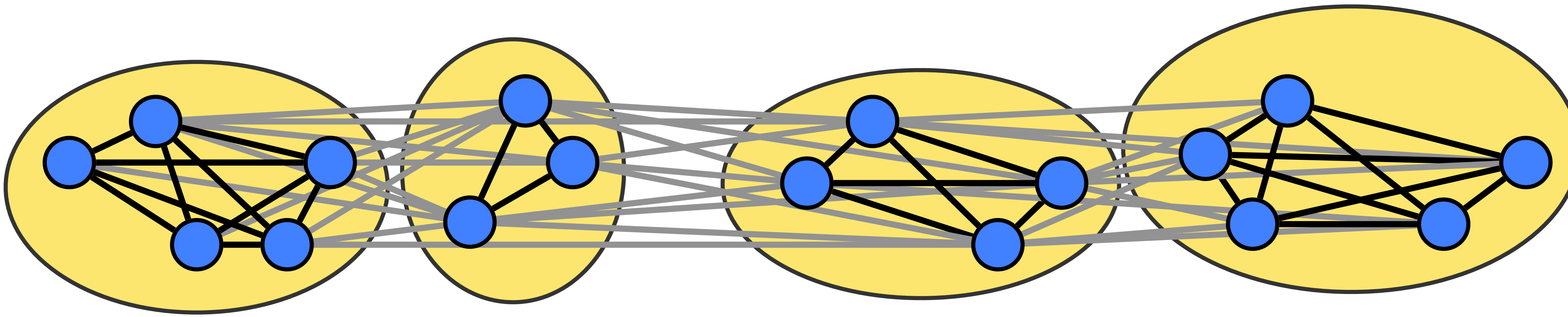
Power law of entity size



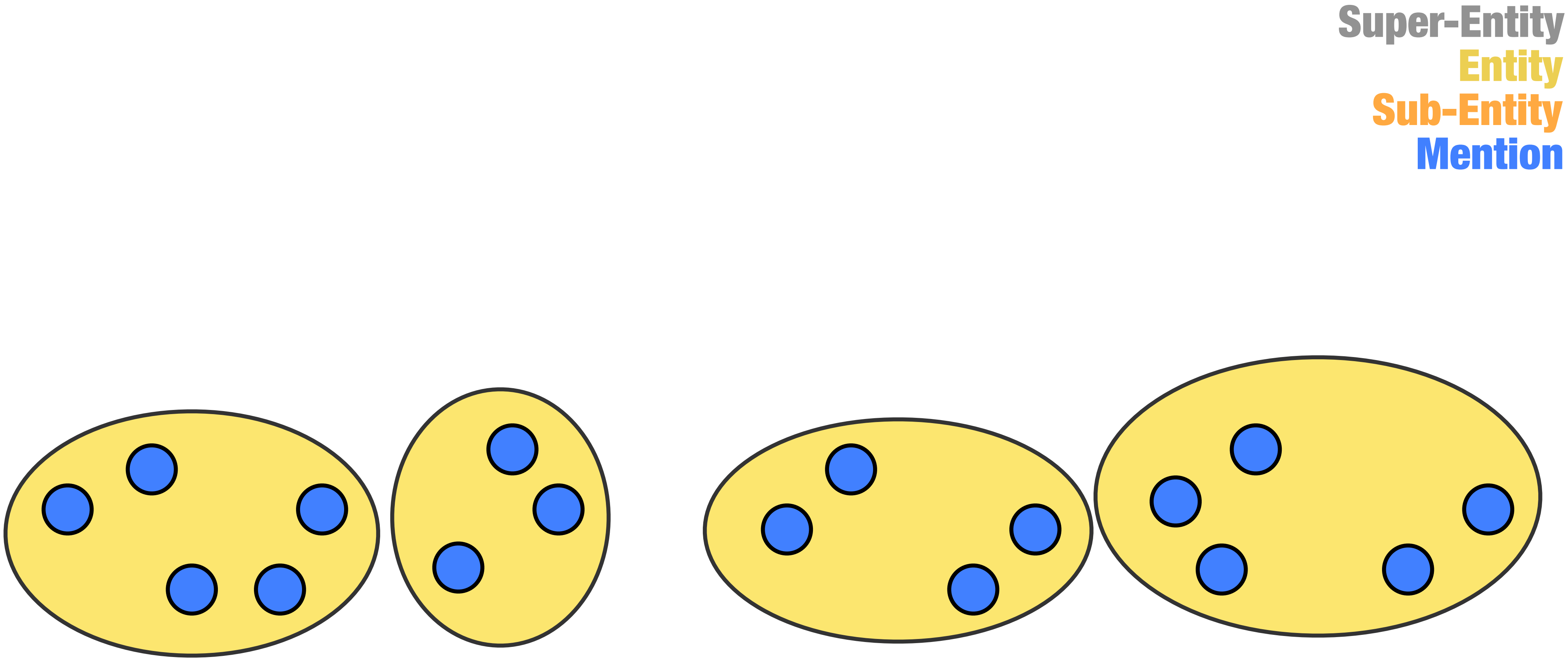
Large number of mentions (100Ks or 10Ms)
Large number of entities (many singleton clusters)

Pair-based Coref

Super-Entity
Entity
Sub-Entity
Mention

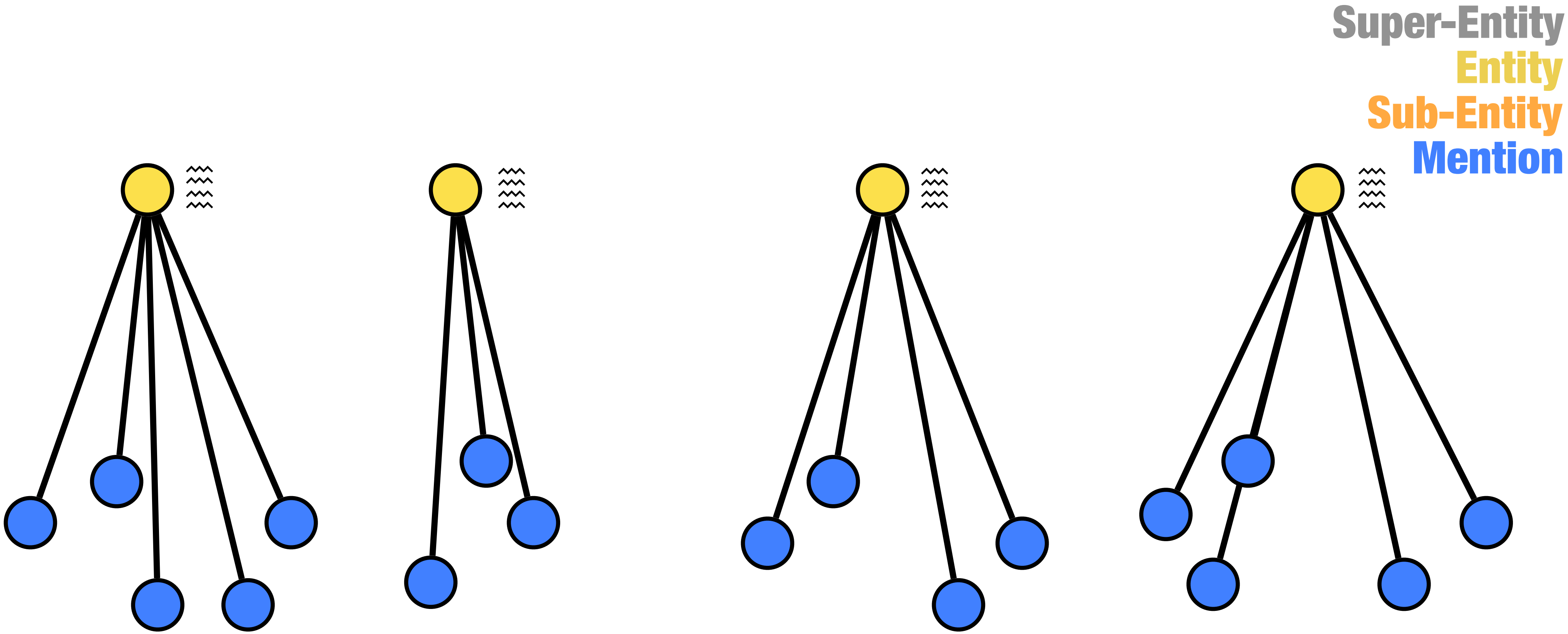


Pair-based Coref



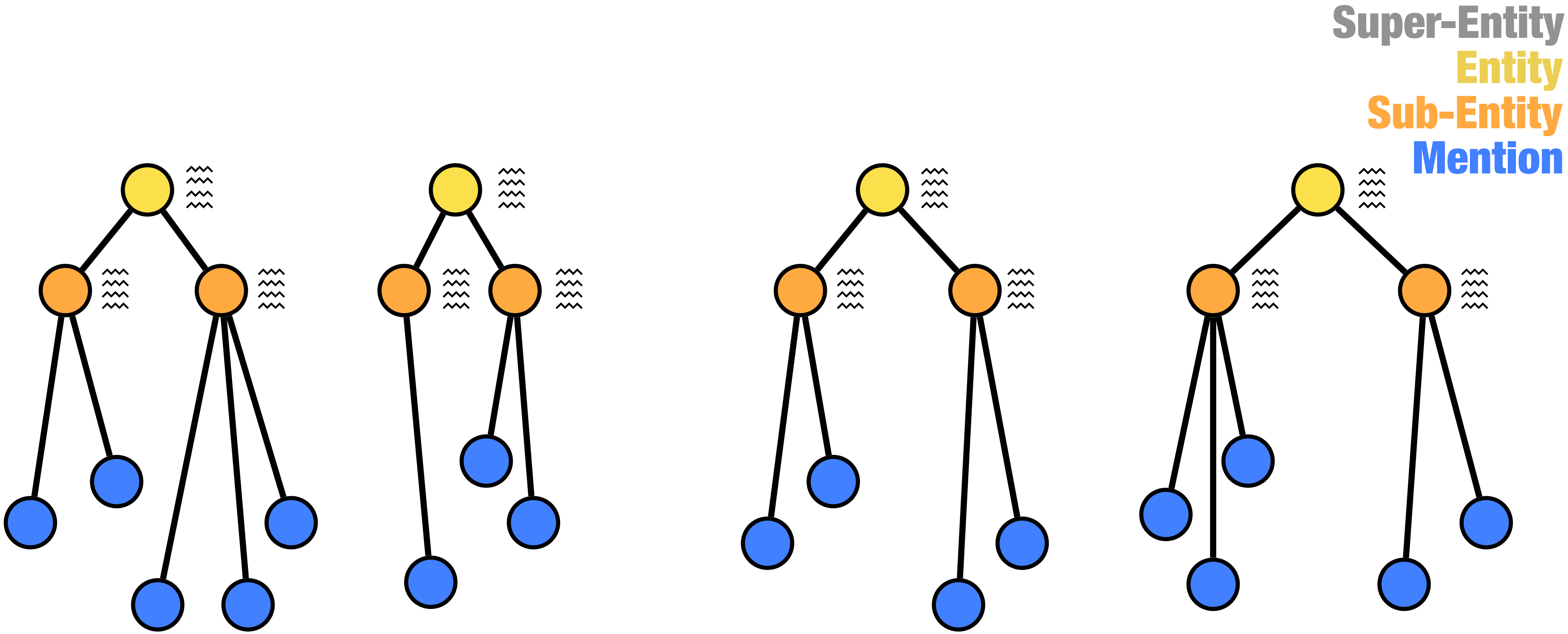
[Wick, Singh, McCallum, ACL, 2012]

Entity-based Coref



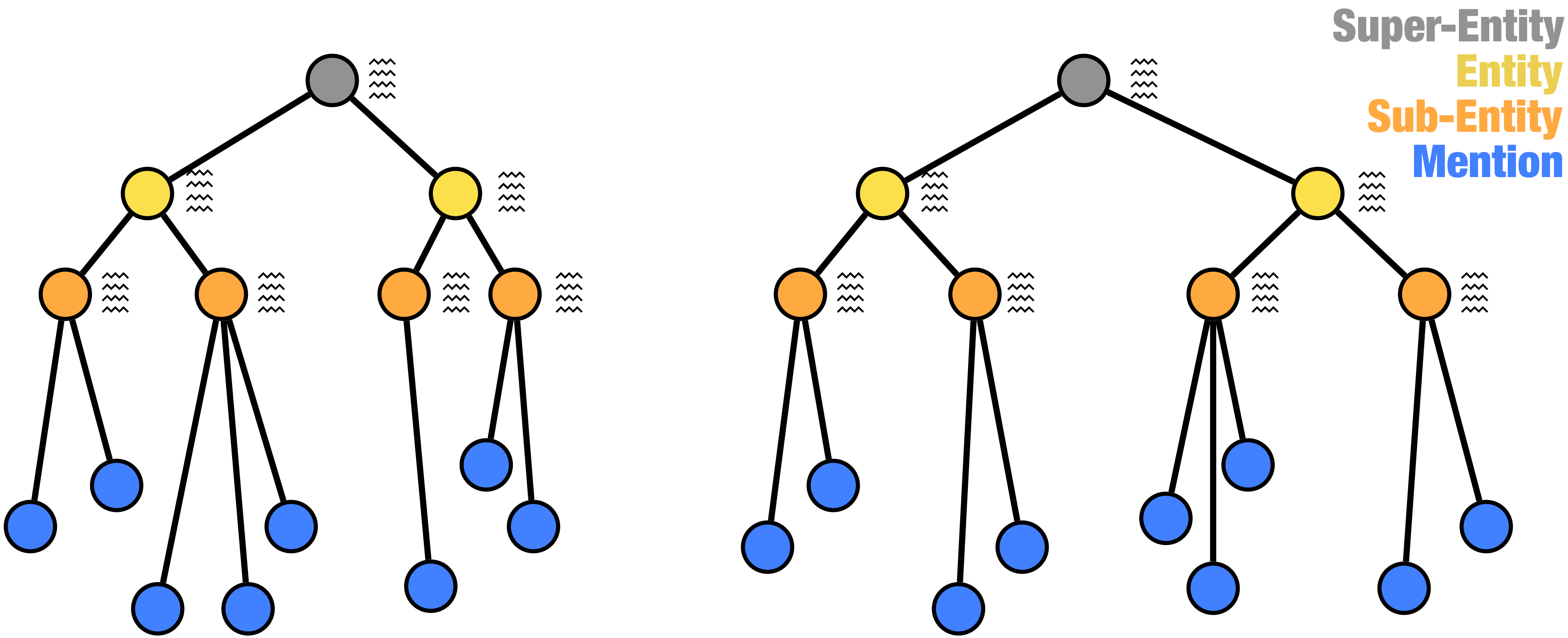
[Wick, Singh, McCallum, ACL, 2012]

Entity-based Coref



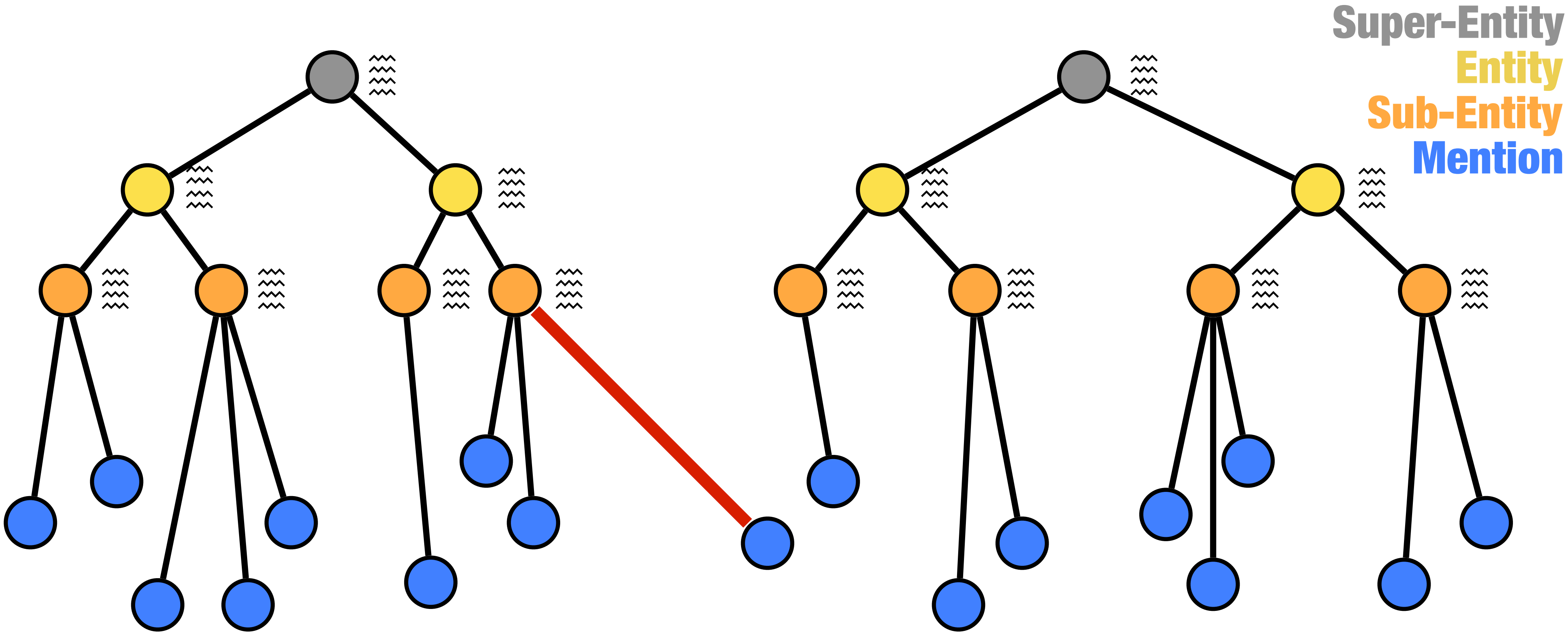
[Wick, Singh, McCallum, ACL, 2012]

Entity-based Coref



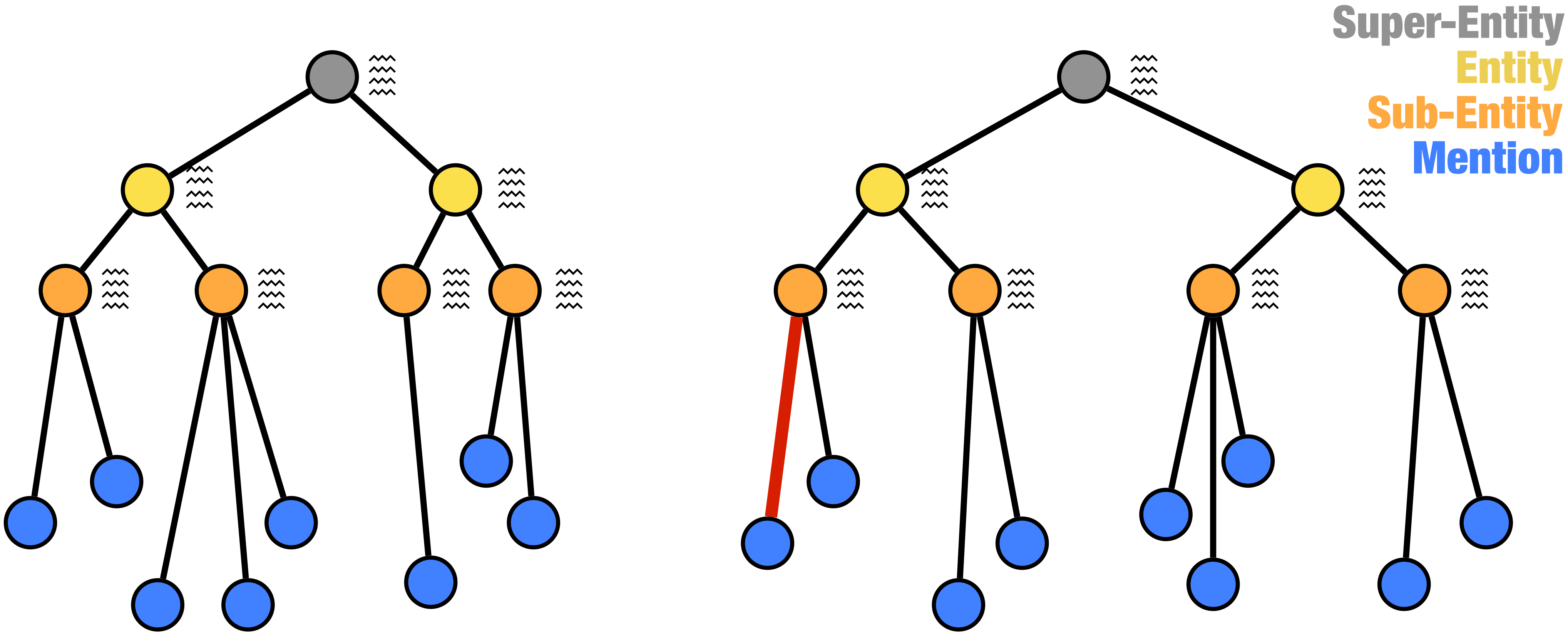
[Wick, Singh, McCallum, ACL, 2012]

Entity-based Coref



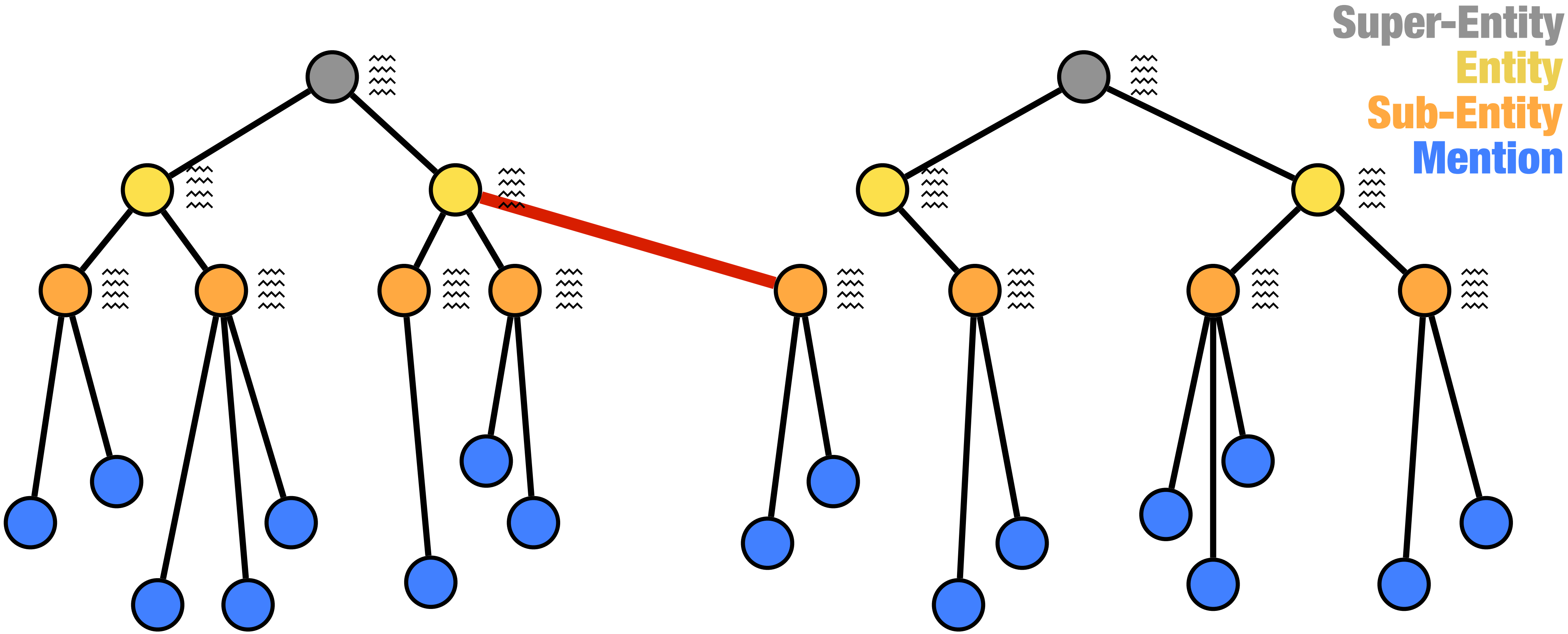
[Wick, Singh, McCallum, ACL, 2012]

Entity-based Coref



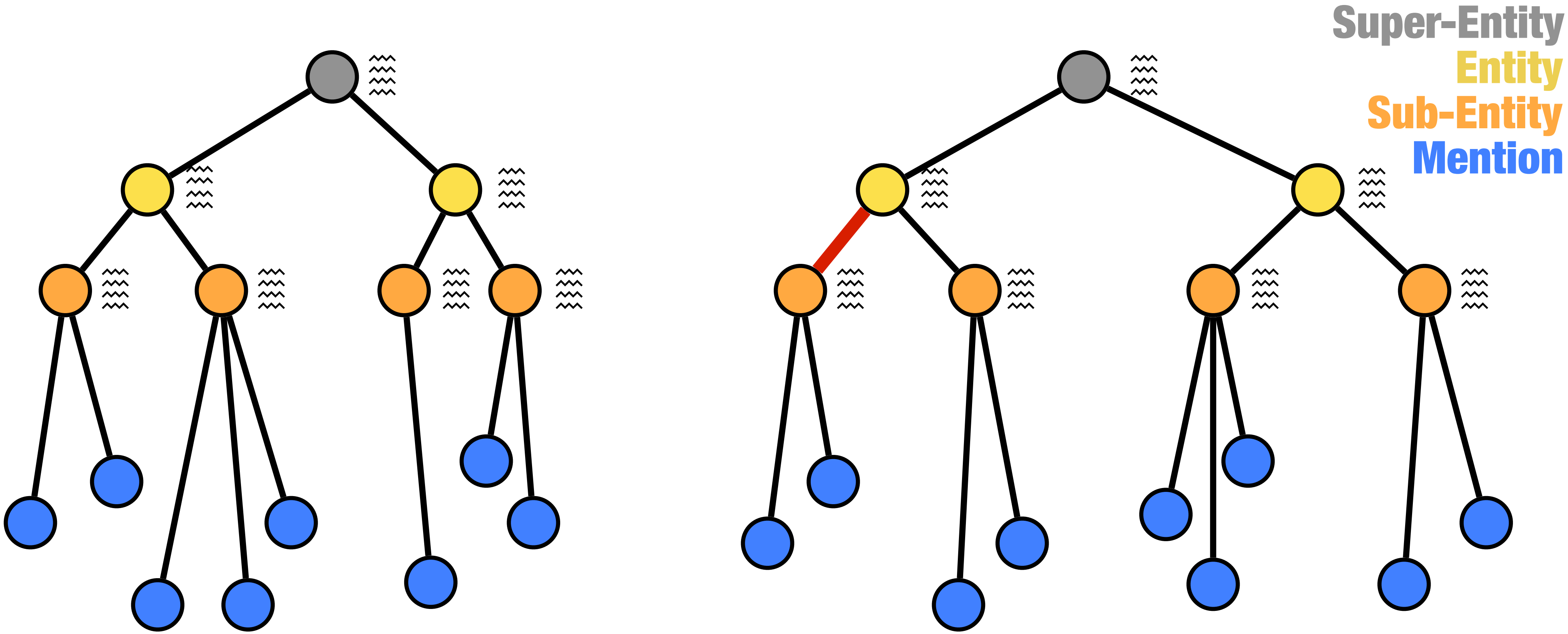
[Wick, Singh, McCallum, ACL, 2012]

Entity-based Coref



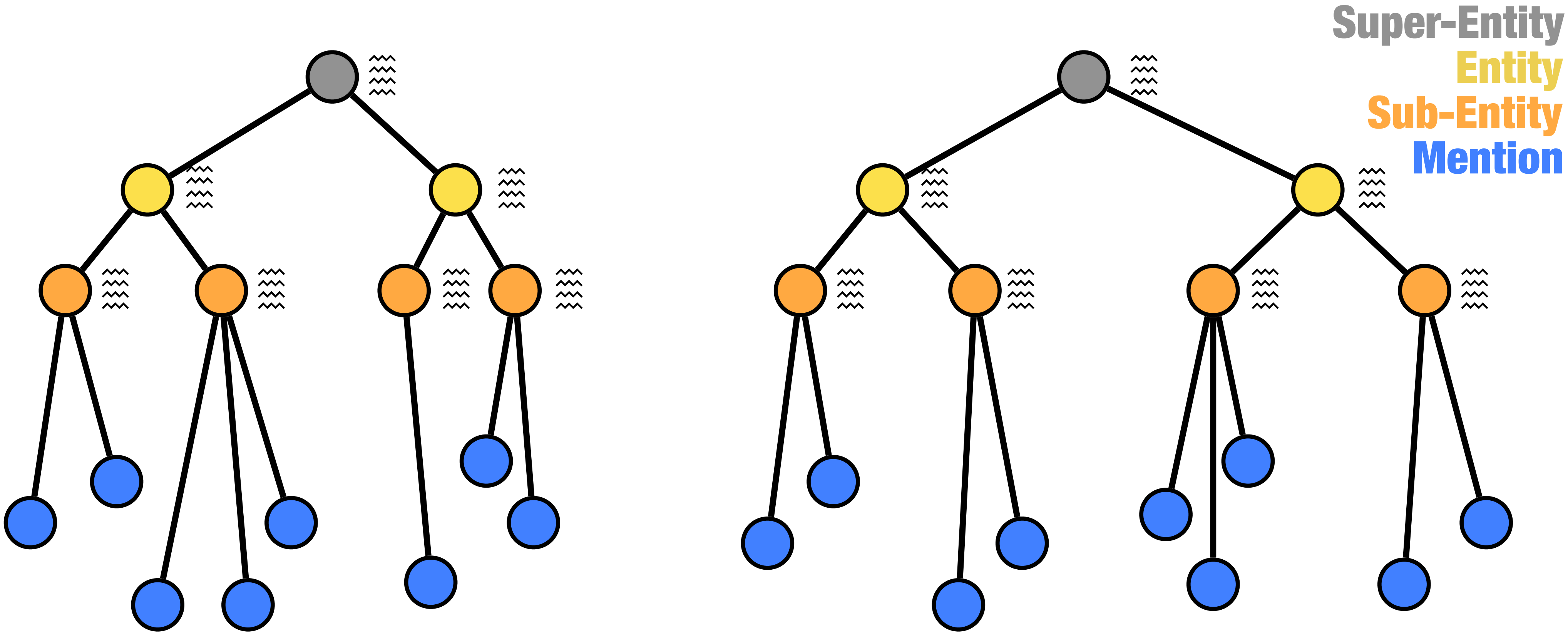
[Wick, Singh, McCallum, ACL, 2012]

Entity-based Coref



[Wick, Singh, McCallum, ACL, 2012]

Entity-based Coref



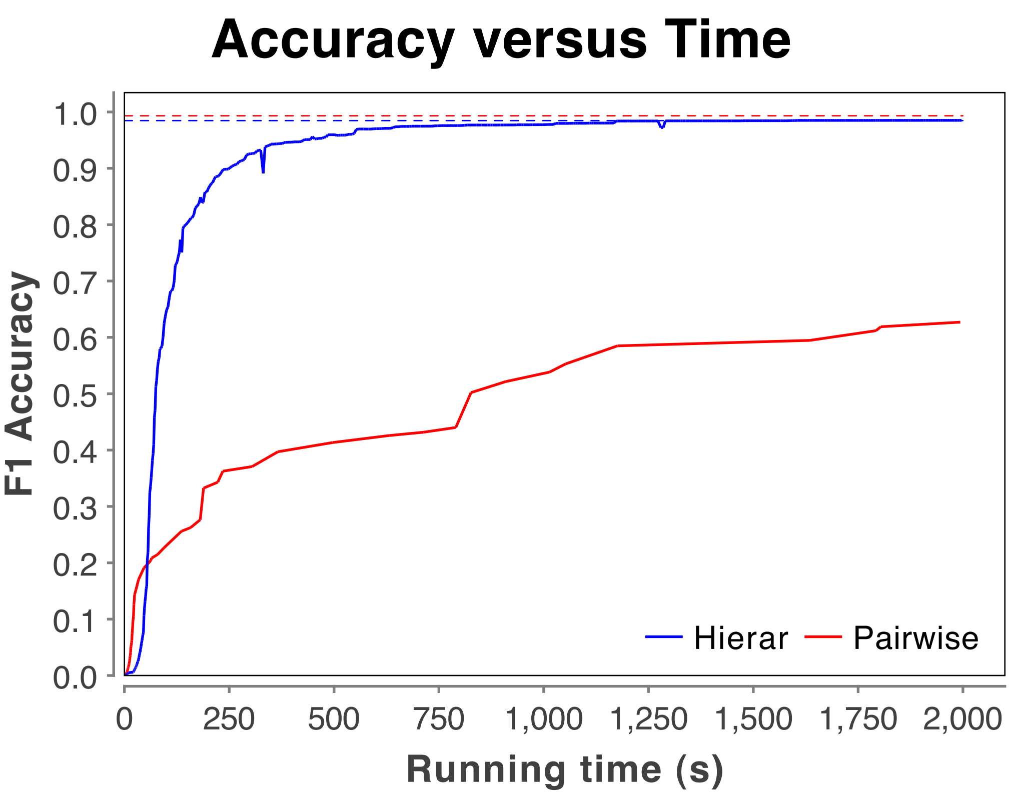
★ More efficient. Fewer factors; avoid N^2.
★ Joint inference on all attributes of entity. Pair-wise couldn't
★ 100k mentions "e coli" hidden under one sub-entity.
★ Better supports inference about crowd-sourced edits

[Wick, Singh, McCallum, ACL, 2012]

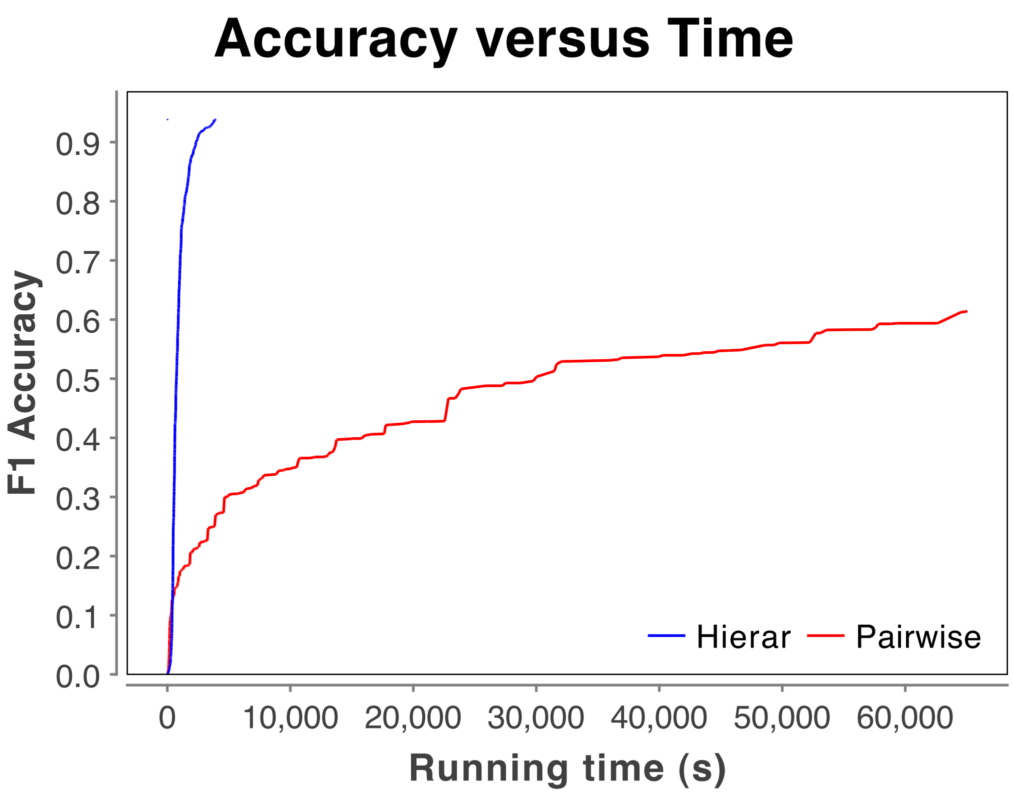
Hierarchical vs Pairwise Evaluation

Author Coreference (single threaded)

145k mentions



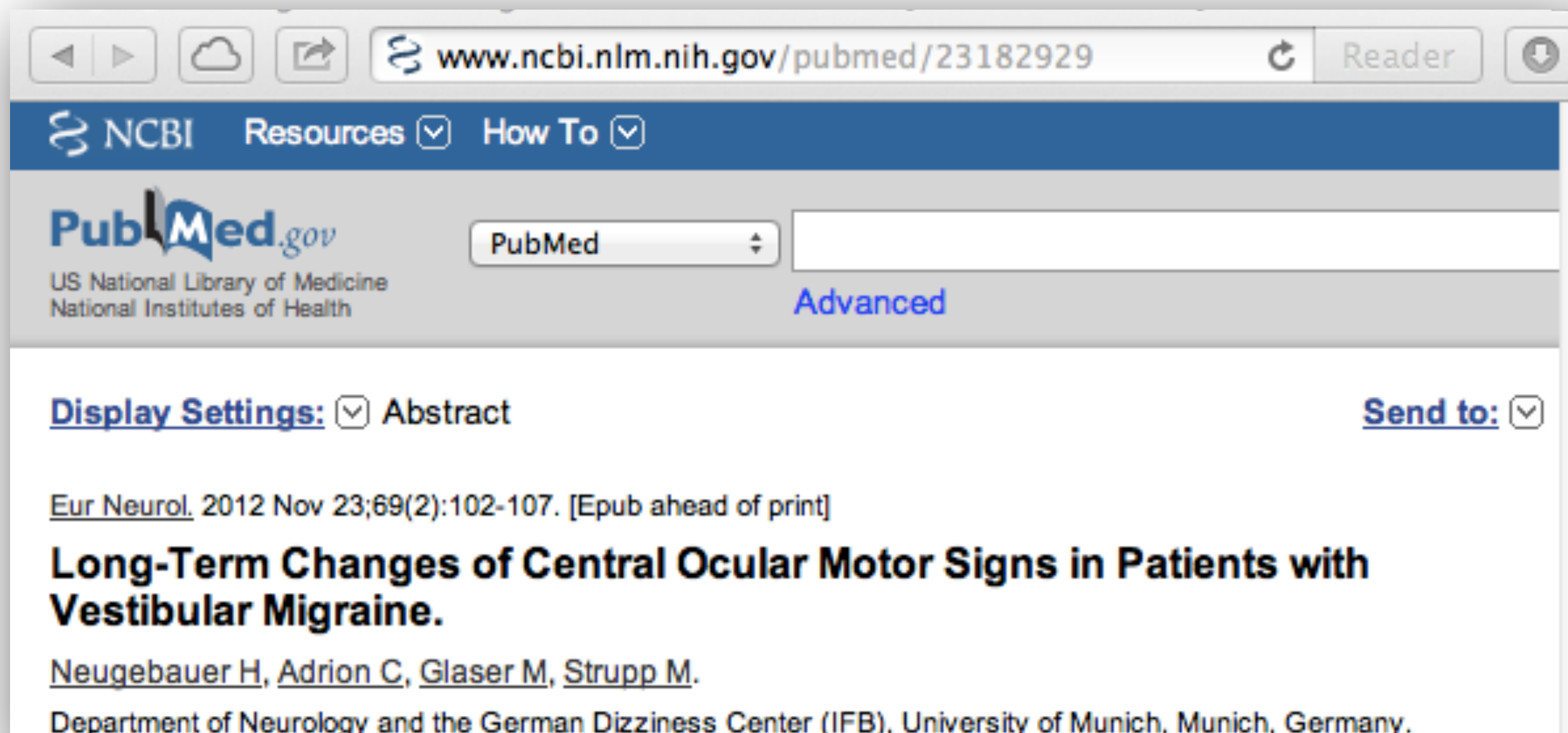
1.3m mentions



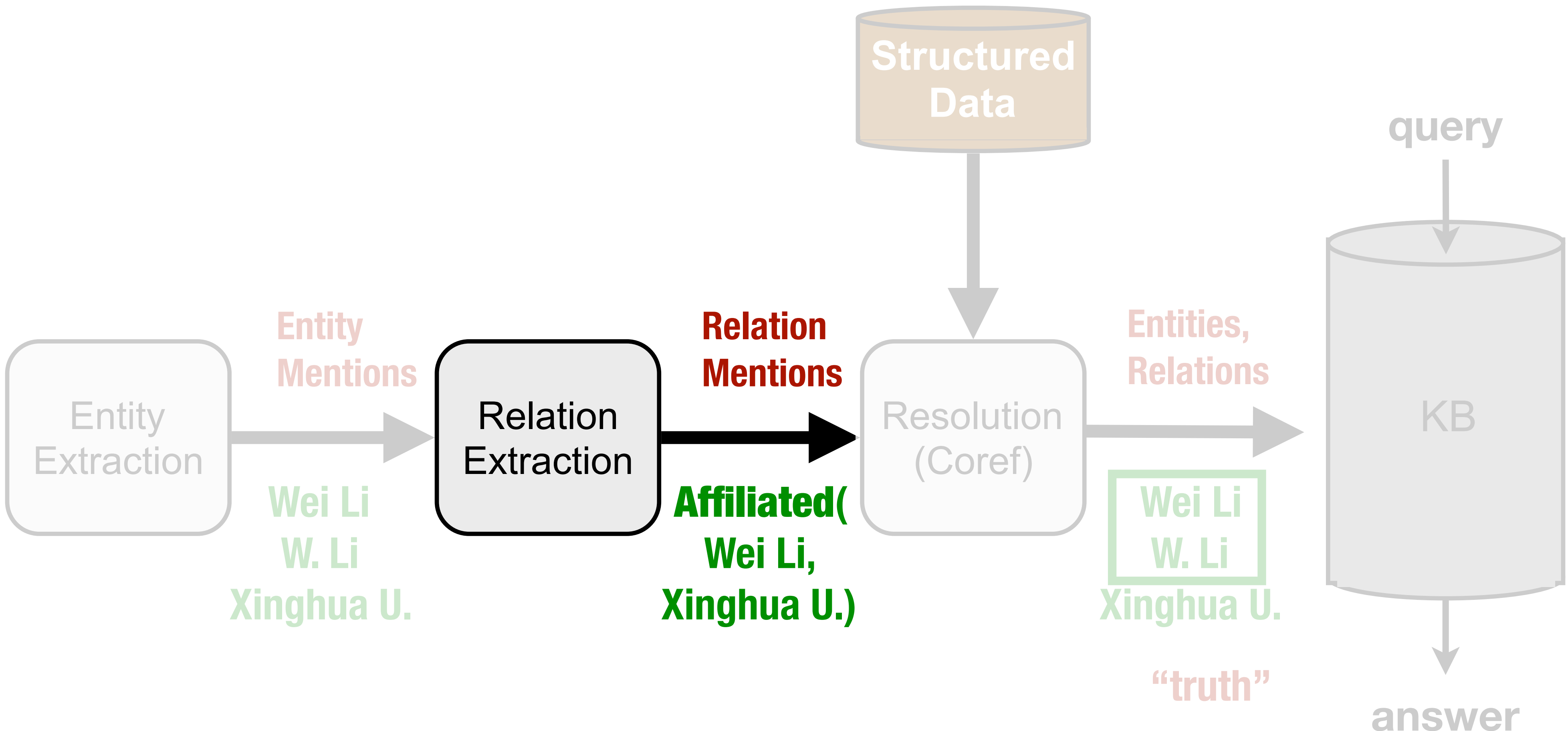
PubMed + Web of Science

- **200 million author mentions = ~400GB**
- **Inference speed**
  - ~100k samples per second
  - ~48 hours of inference time

3 machines
48 cores



Relation Extraction



January 15, 2000

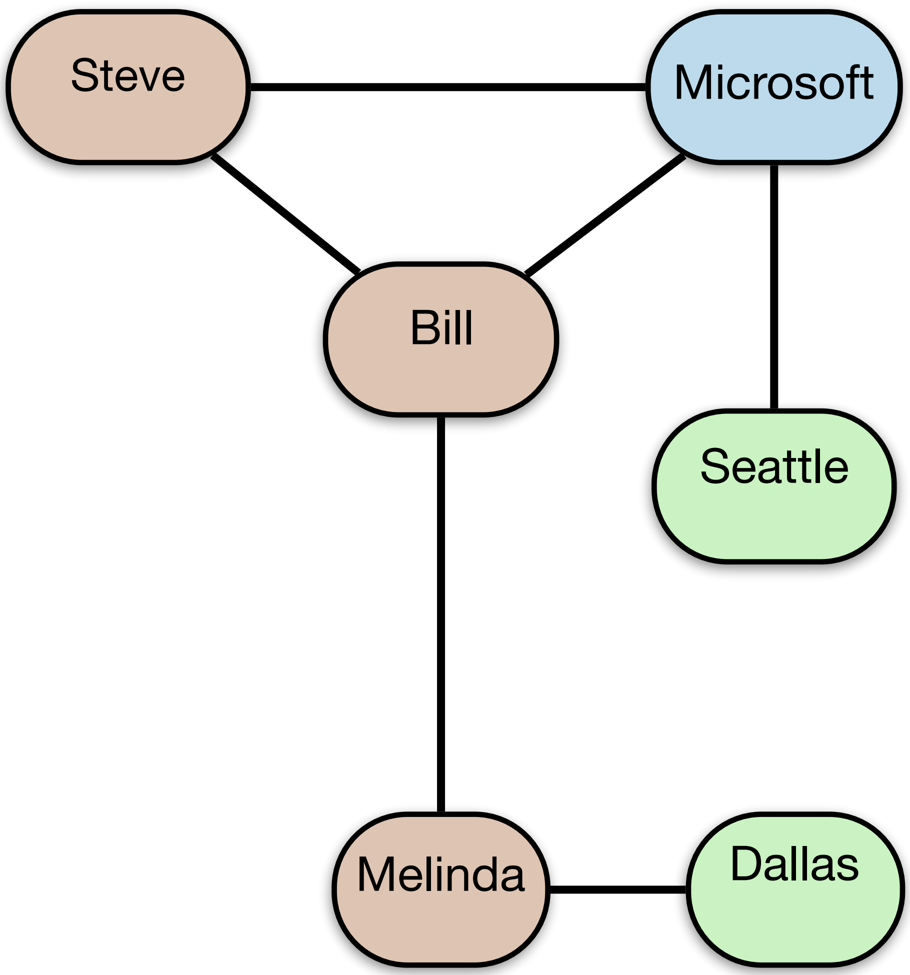
Tech pioneer Bill Gates stepped down today as chief executive officer of Microsoft, the Seattle-headquartered software giant. He will remain Chairman of the company, which rose to prominence after beating Digital Research Inc for the contract to provide an operating system for PCs. His long-time friend, Steve Balmer, will take over as CEO of Microsoft. Gates will now focus on the charitable foundation he runs with his wife Melinda French Gates. Bill and Melinda were married in a ceremony in Hawaii, rather than her hometown of Dallas. Steve Balmer was best man.

- Text → Mentions → Coref → Relations

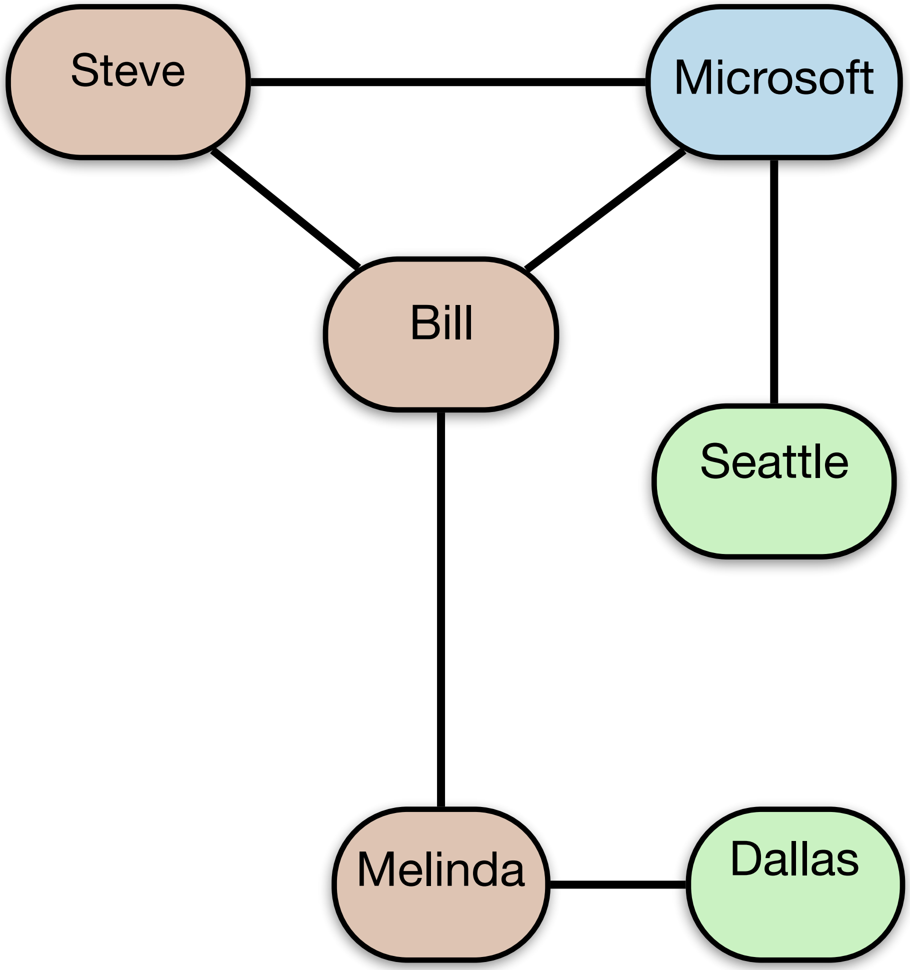
January 15, 2000

Tech pioneer Bill Gates stepped down today as chief executive officer of Microsoft, the Seattle-headquartered software giant. He will remain Chairman of the company, which rose to prominence after beating Digital Research Inc for the contract to provide an operating system for PCs. His long-time friend, Steve Balmer, will take over as CEO of Microsoft. Gates will now focus on the charitable foundation he runs with his wife Melinda French Gates. Bill and Melinda were married in a ceremony in Hawaii, rather than her hometown of Dallas. Steve Balmer was best man.

- Text → Mentions → Coref → Relations



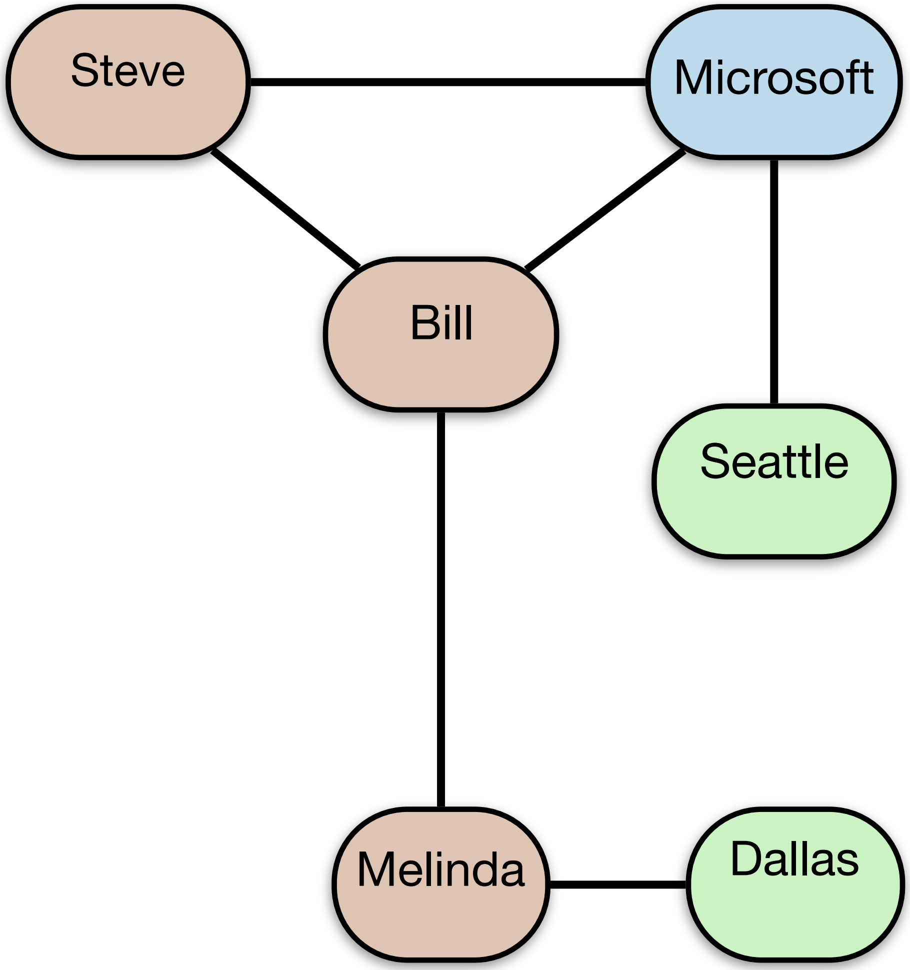
- Text → Mentions → Coref → Relations
- Schema:
  - Entity Types



- Text → Mentions → Coref → Relations
- Schema:
  - Entity Types

entity types

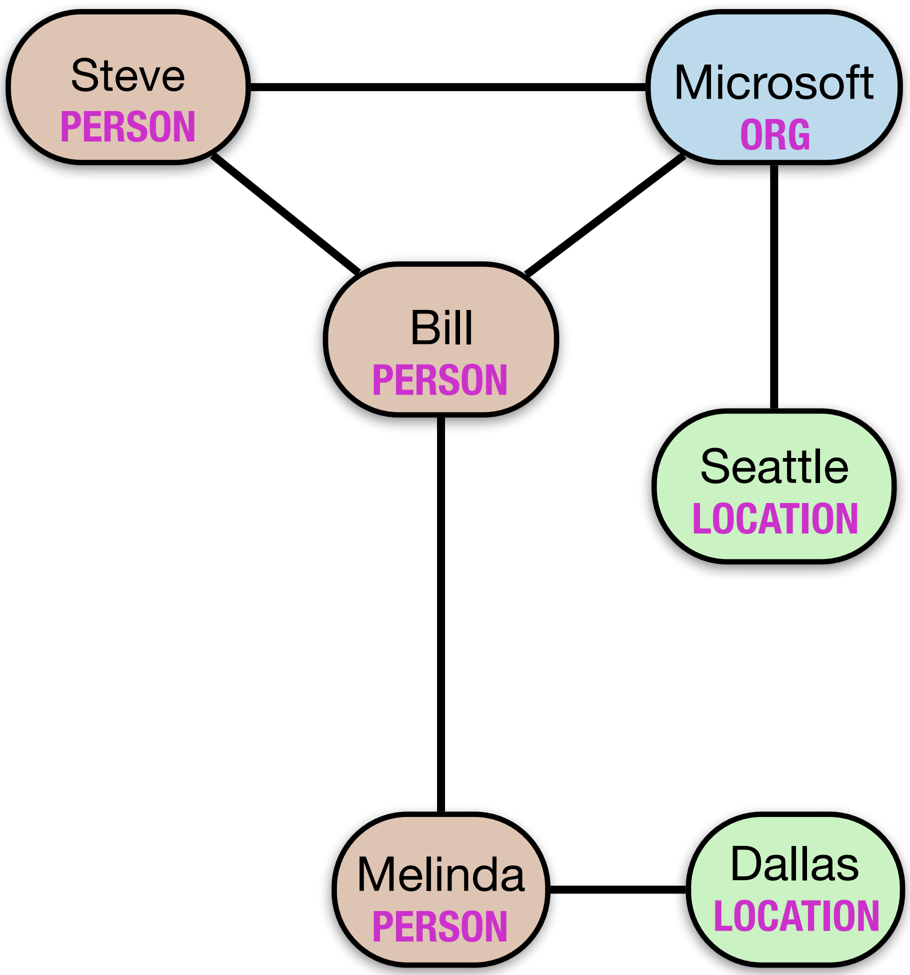
PERSON
LOCATION
ORG



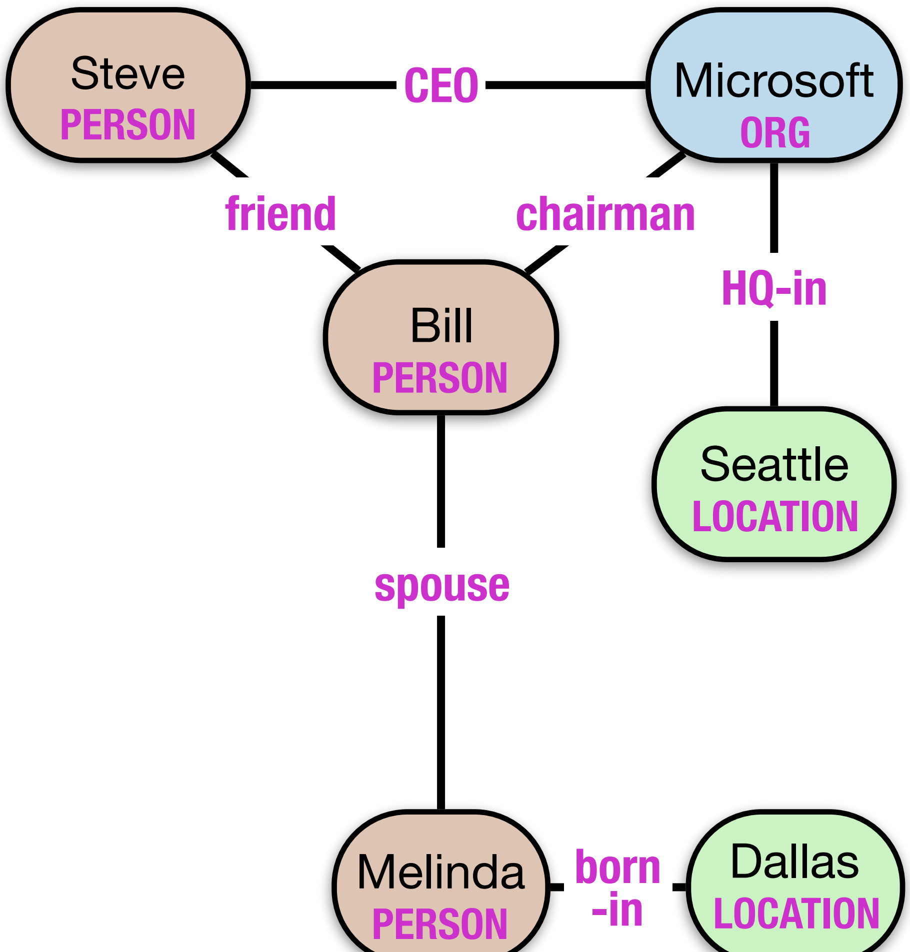
- Text → Mentions → Coref → Relations
- Schema:
  - Entity Types

entity types

PERSON
LOCATION
ORG



- Text → Mentions → Coref → Relations
- Schema:
  - Entity Types
  - Relation Types



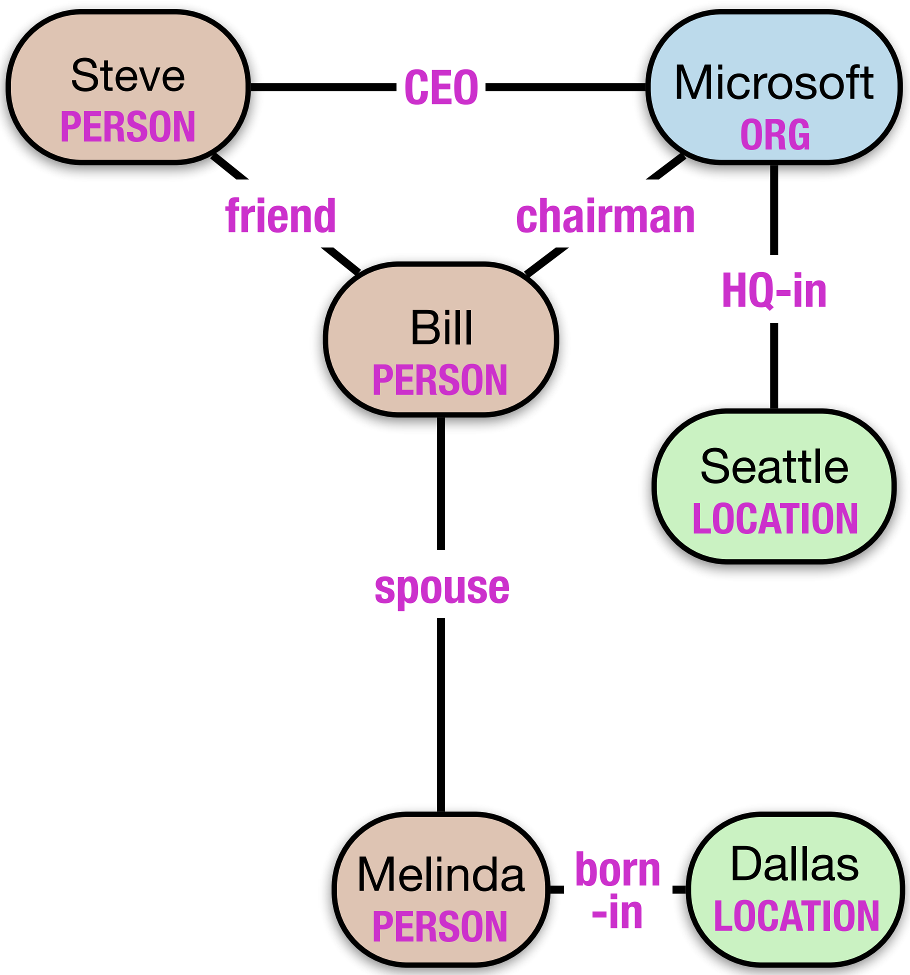
entity types

PERSON
LOCATION
ORG

relation types

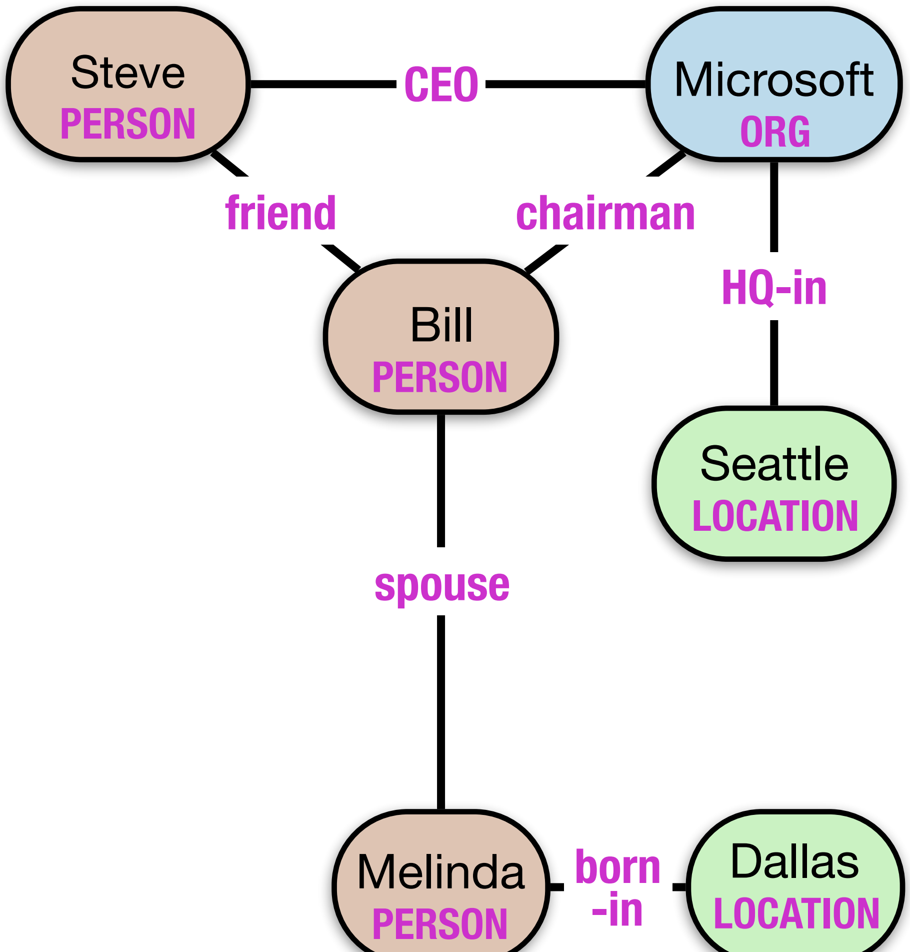
CEO
friend
chairman
HQ-in
spouse
born-in

- Text → Mentions → Coref → Relations
- Schema:
  - Entity Types
  - Relation Types



entity types

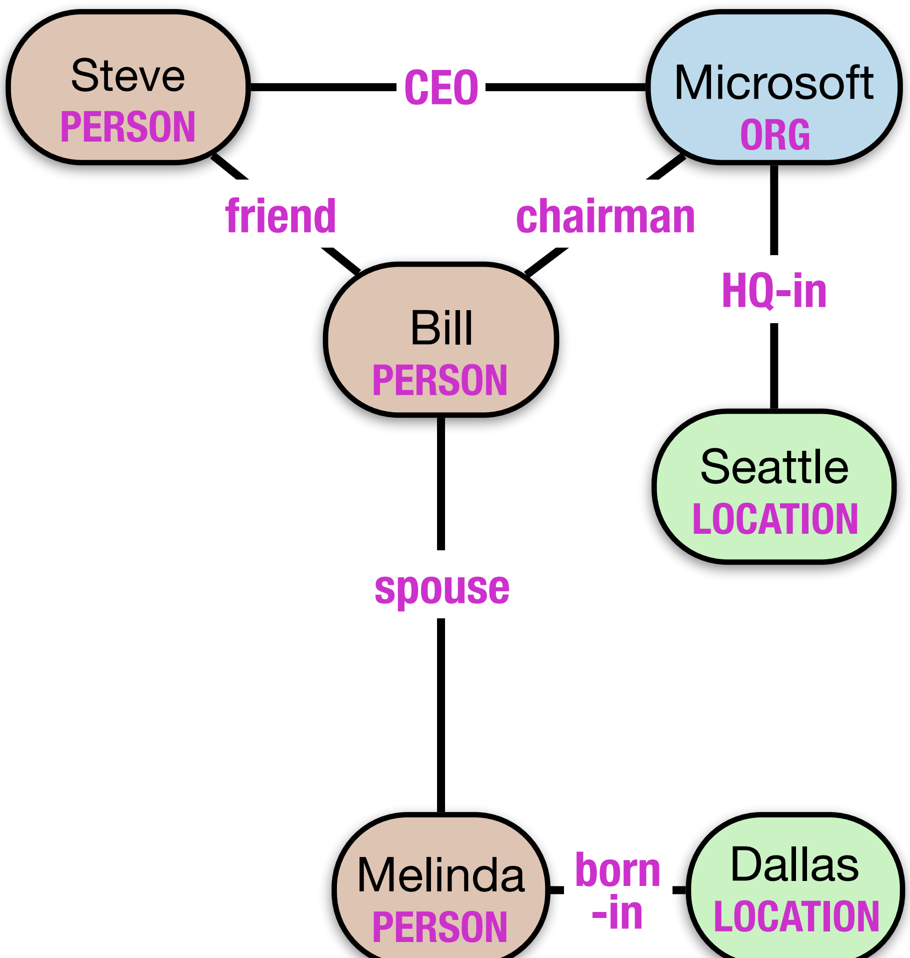
- Text → Mentions → Coref → Relations
- Schema:
  - Entity Types
  - Relation Types

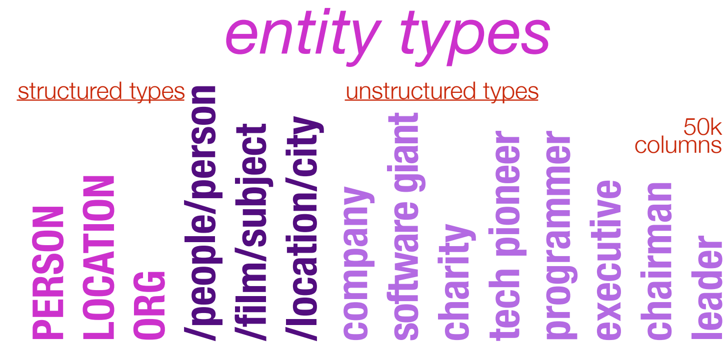


entity types

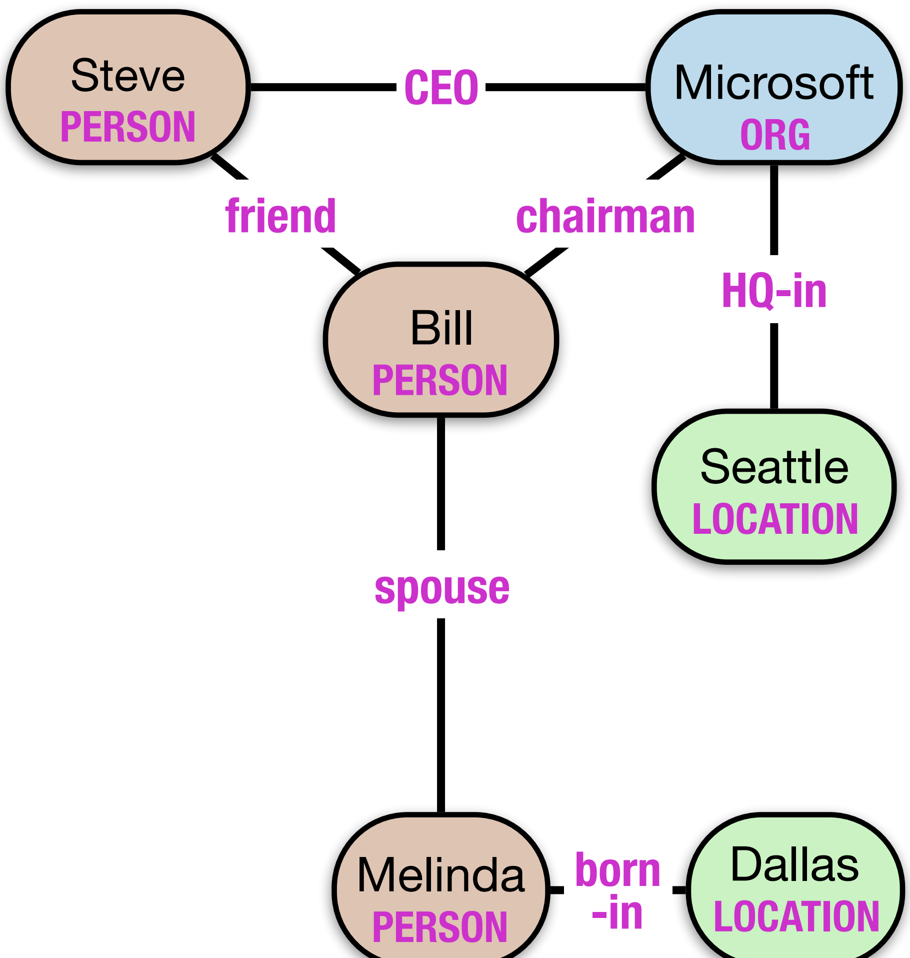
PERSON
LOCATION
ORG
/people/person
/film/subject
/location/city

- Text → Mentions → Coref → Relations
- Schema:
  - Entity Types
  - Relation Types





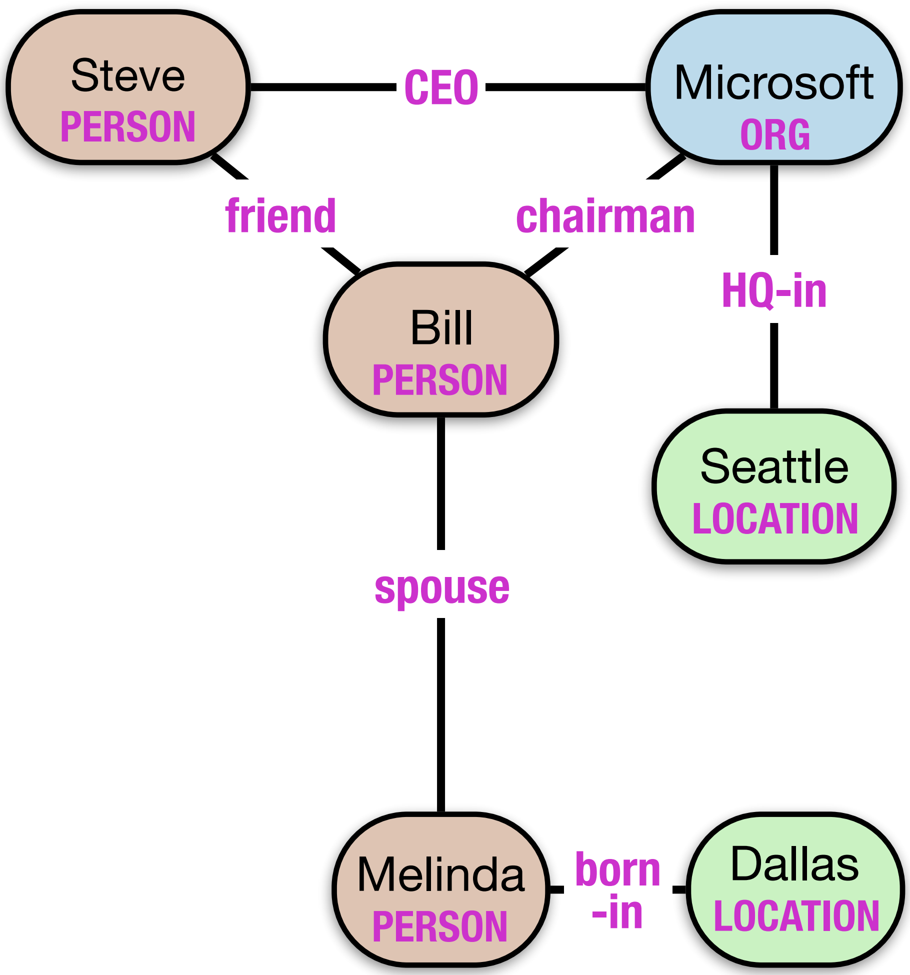
- Text → Mentions → Coref → Relations
- Universal Schema: [AKBC 2012]
  - Entity Types
  - Relation Types



[Riedel, Yao, Marlin, McCallum NAACL 2012]

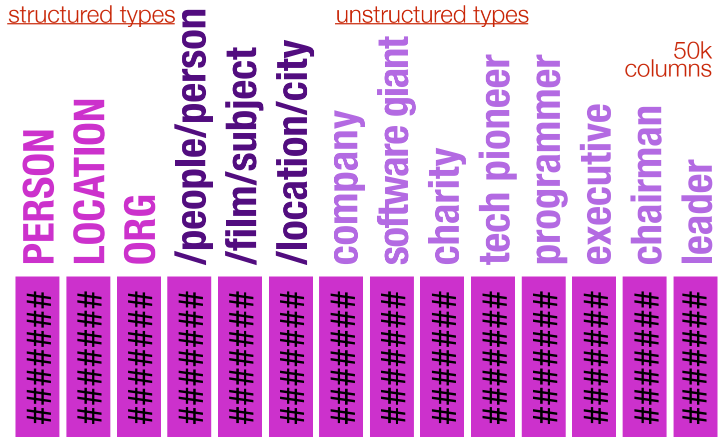
Universal Schema *entity types*

- Text → Mentions → Coref → Relations
- Universal Schema: [AKBC 2012]
  - Entity Types
  - Relation Types

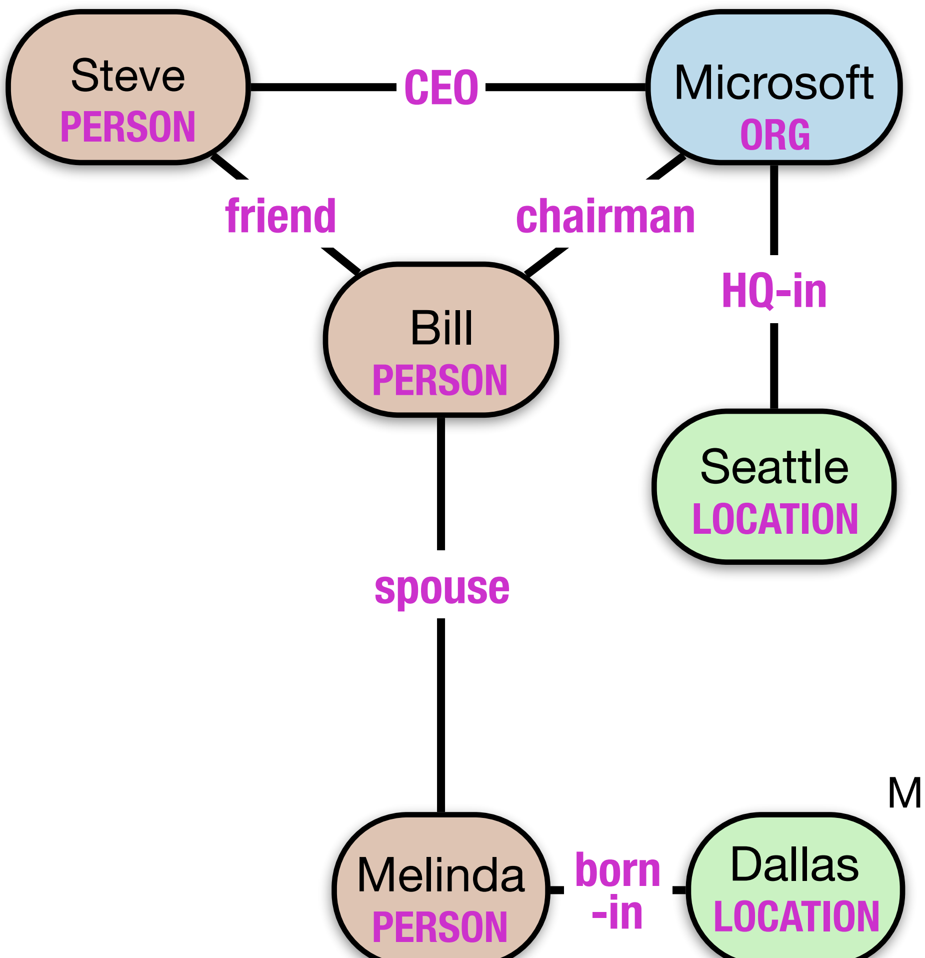


[Riedel, Yao, Marlin, McCallum NAACL 2012]

Universal Schema *entity types*

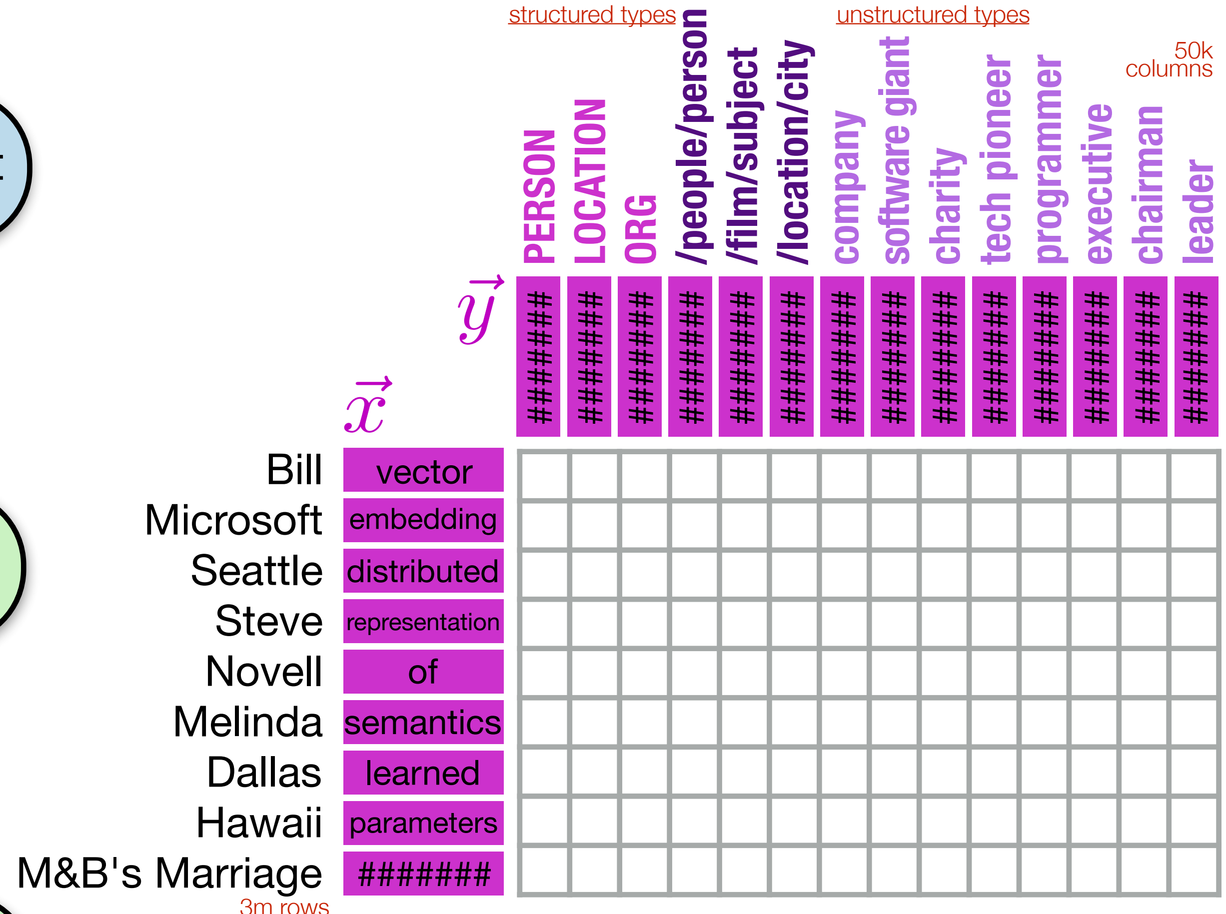


- Text → Mentions → Coref → Relations
- Universal Schema: [AKBC 2012]
  - Entity Types
  - Relation Types



[Riedel, Yao, Marlin, McCallum NAACL 2012]

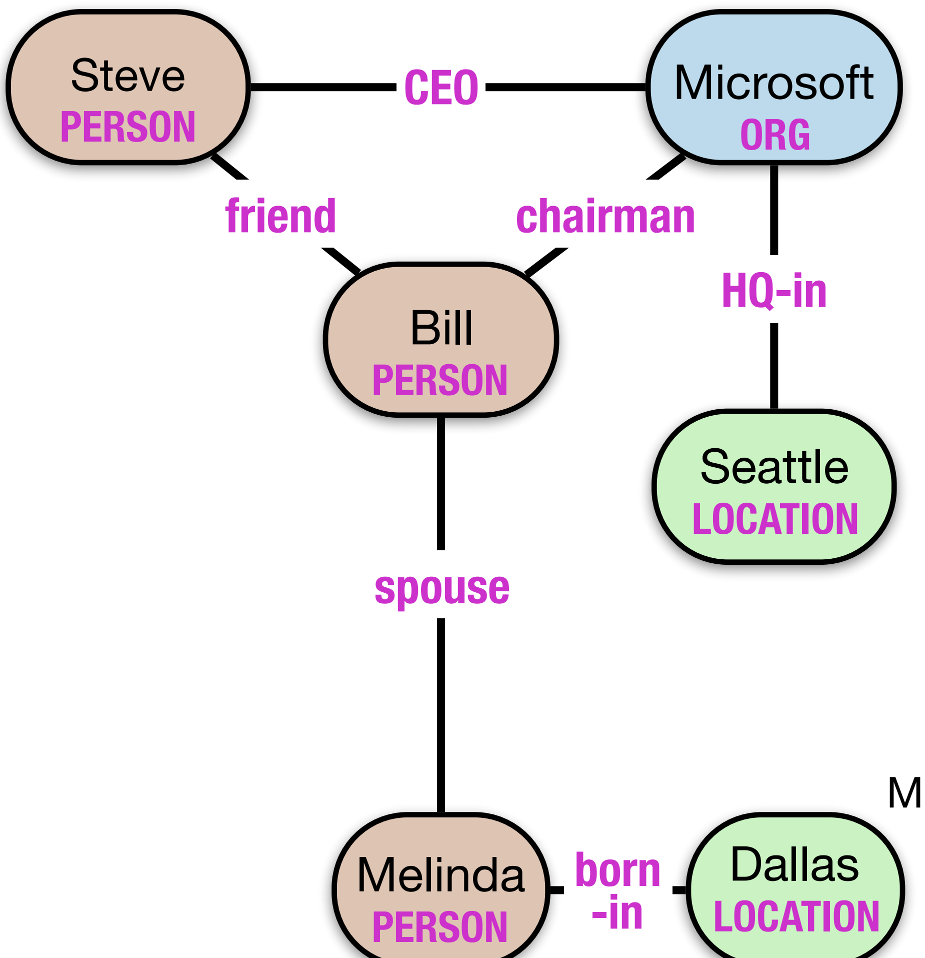
Universal Schema entity types



$$f_{e,t} = \sigma(\vec{x}_e \cdot \vec{y}_t)$$

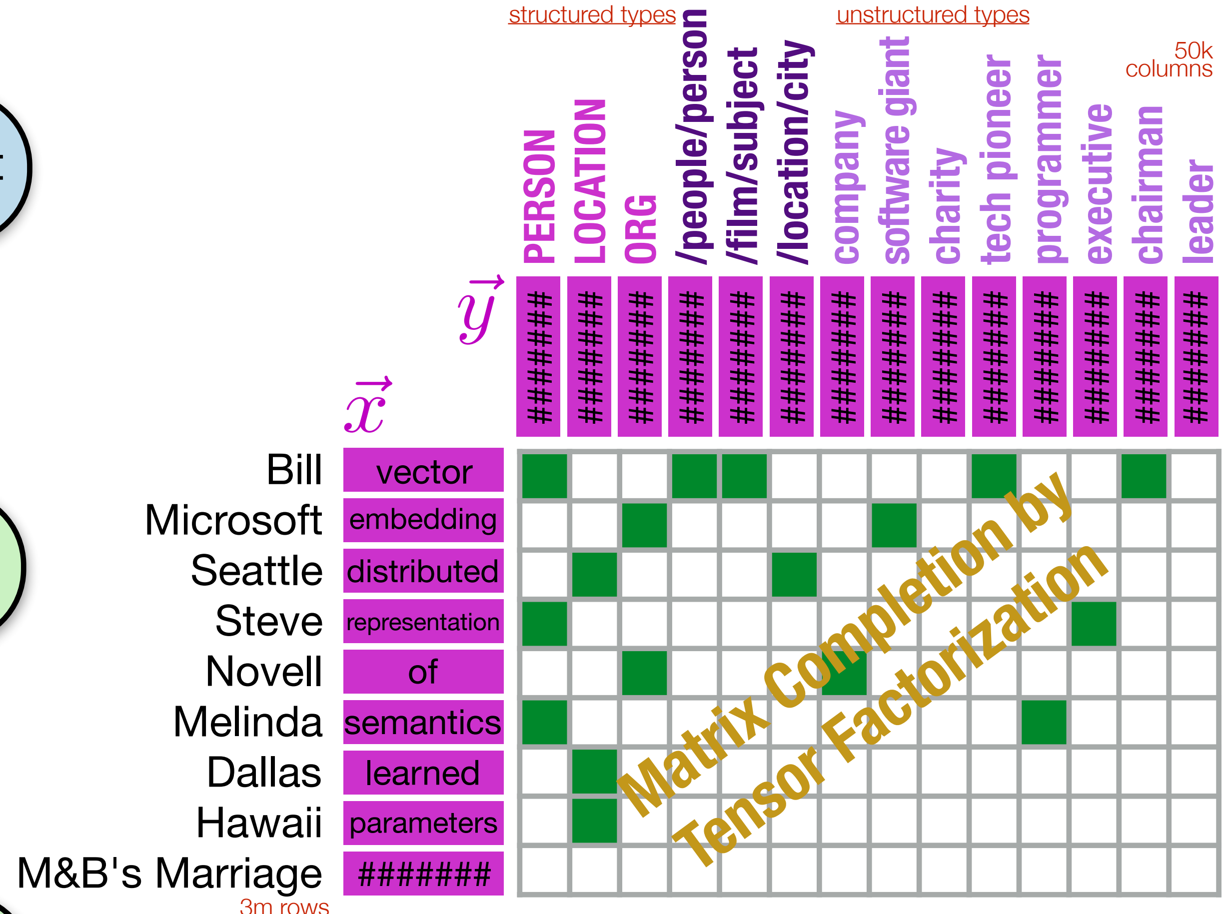
$$\sigma(\theta) = \frac{1}{1 + \exp(-\theta)}$$

- Text → Mentions → Coref → Relations
- Universal Schema: [AKBC 2012]
  - Entity Types
  - Relation Types



[Riedel, Yao, Marlin, McCallum NAACL 2012]

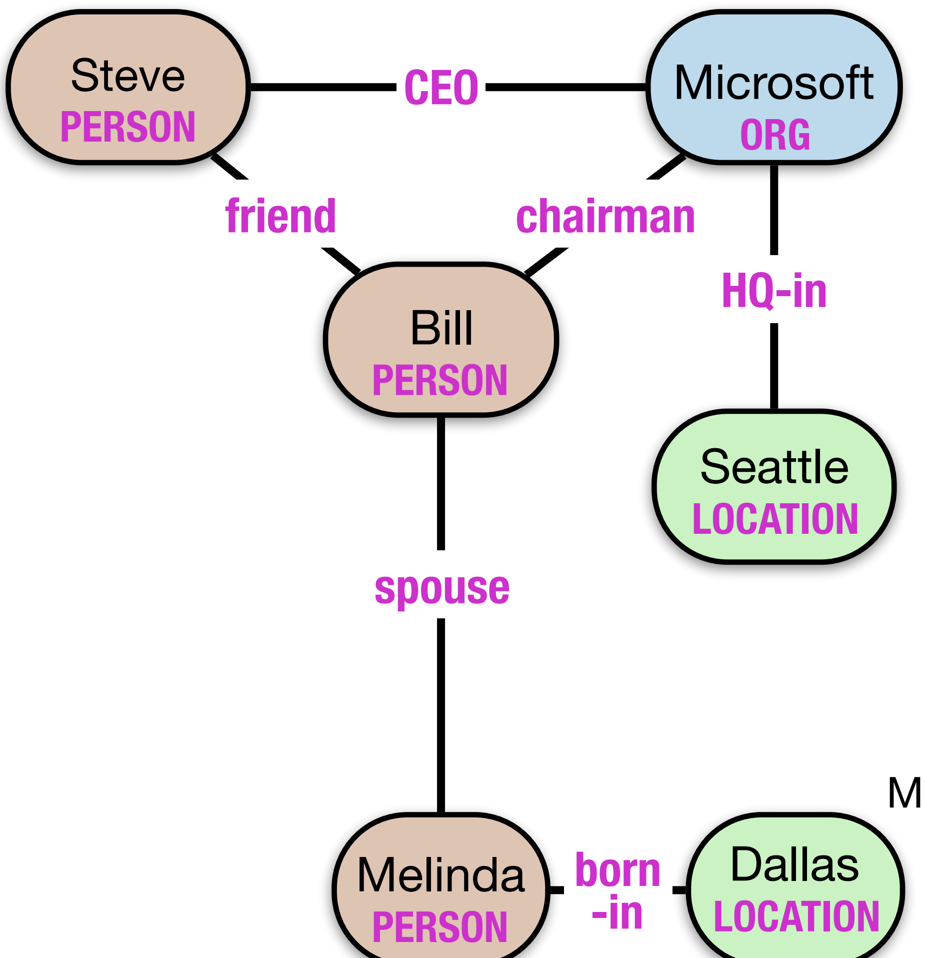
***Universal Schema** entity types*



$$f_{e,t} = \sigma(\vec{x}_e \cdot \vec{y}_t)$$

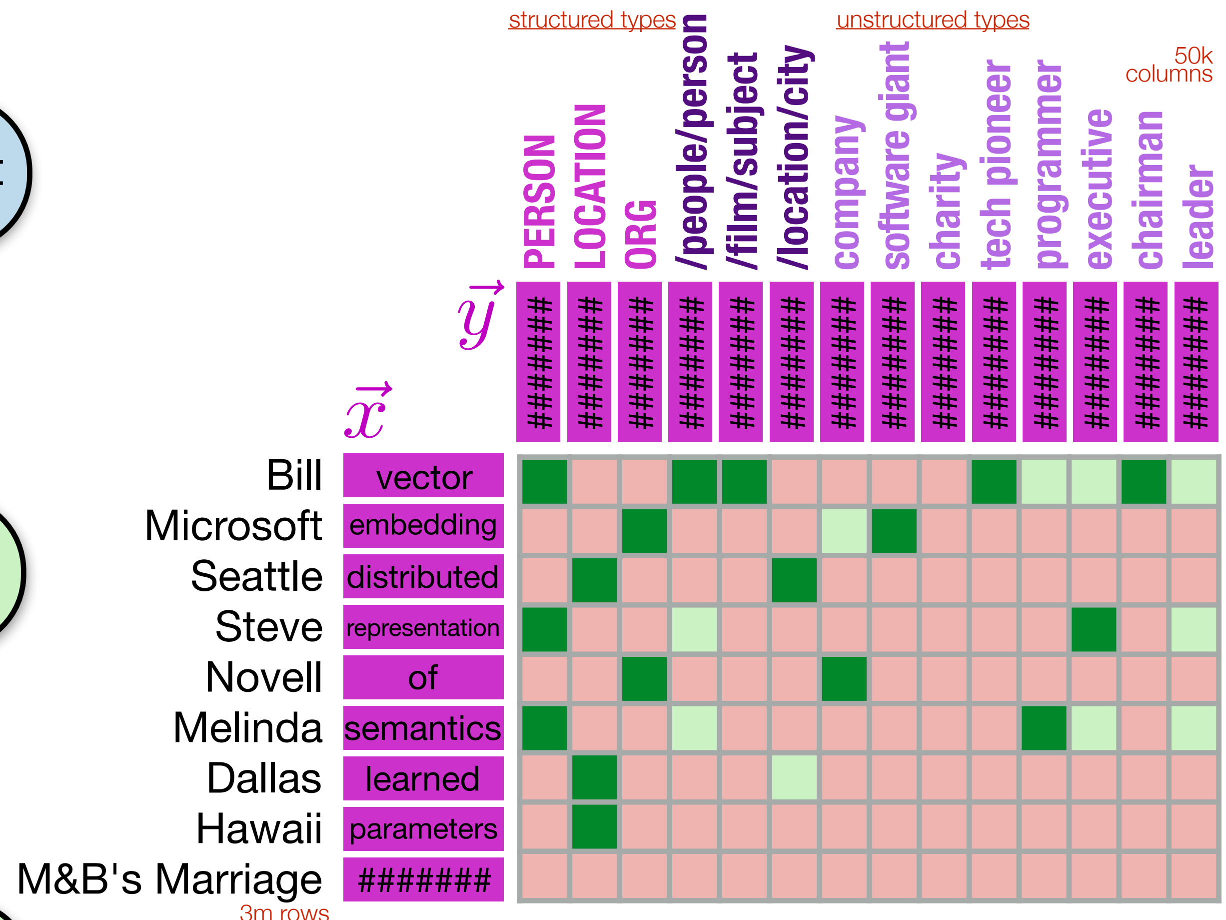
$\sigma(\theta) = \frac{1}{1+\exp(-\theta)}$

- Text → Mentions → Coref → Relations
- Universal Schema: [AKBC 2012]
  - Entity Types
  - Relation Types



[Riedel, Yao, Marlin, McCallum NAACL 2012]

Universal Schema *entity types*



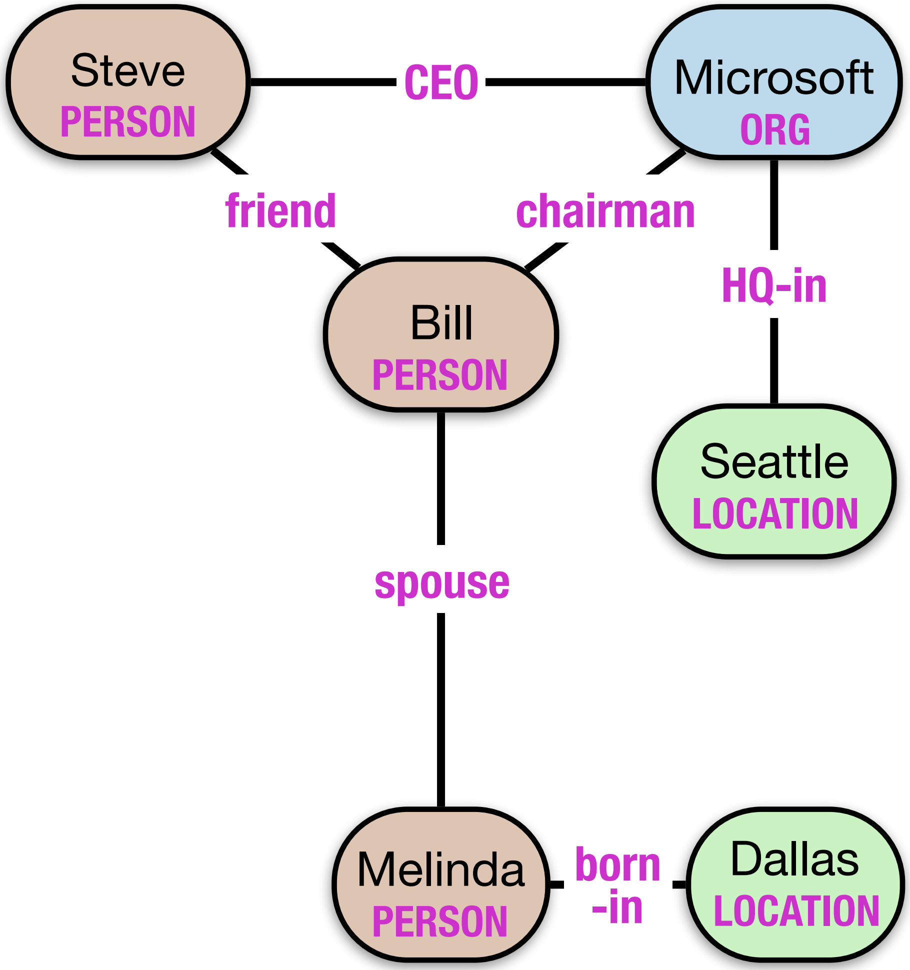
$$f_{e,t} = \sigma(\vec{x}_e \cdot \vec{y}_t)$$

$$\sigma(\theta) = \frac{1}{1 + \exp(-\theta)}$$

- Text → Mentions → Coref → Relations
- Universal Schema: [AKBC 2012]
  - Entity Types
  - Relation Types

[Riedel, Yao, Marlin, McCallum NAACL 2012]

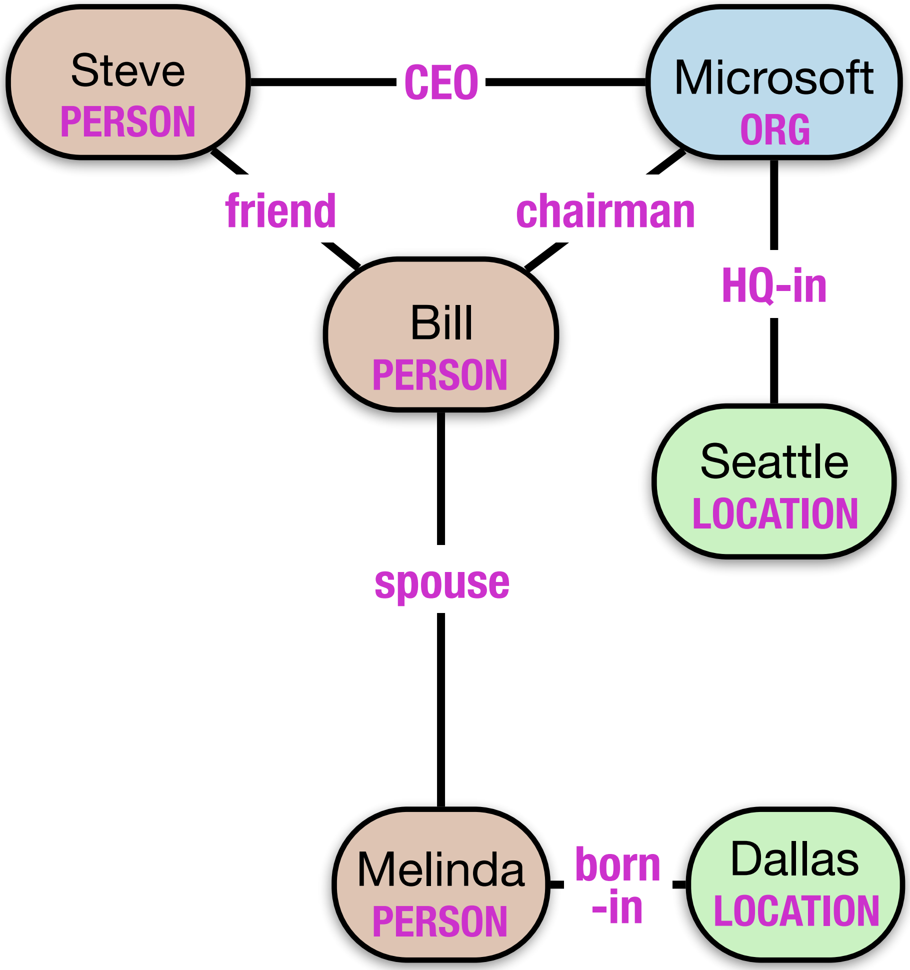
Universal Schema *entity types*



- Text → Mentions → Coref → Relations
- Universal Schema: [AKBC 2012]
  - Entity Types
  - Relation Types

[Riedel, Yao, Marlin, McCallum NAACL 2012]

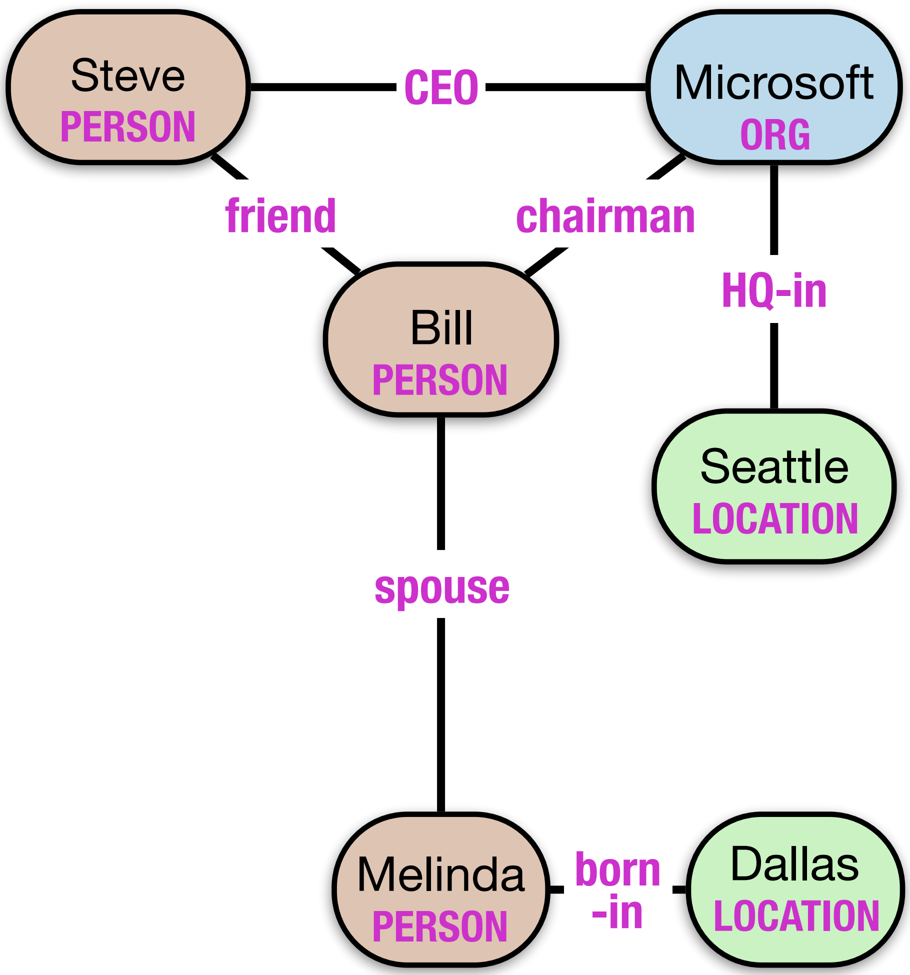
***Universal Schema** relation types*

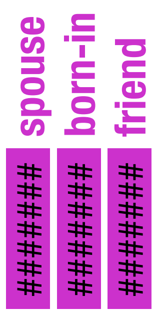


- Text → Mentions → Coref → Relations
- Universal Schema: [AKBC 2012]
  - Entity Types
  - Relation Types

[Riedel, Yao, Marlin, McCallum NAACL 2012]

<u>Universal Schema</u> *relation types*

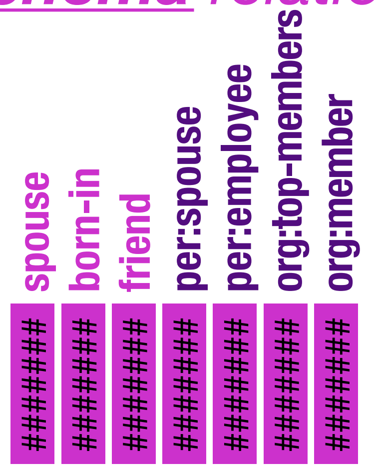


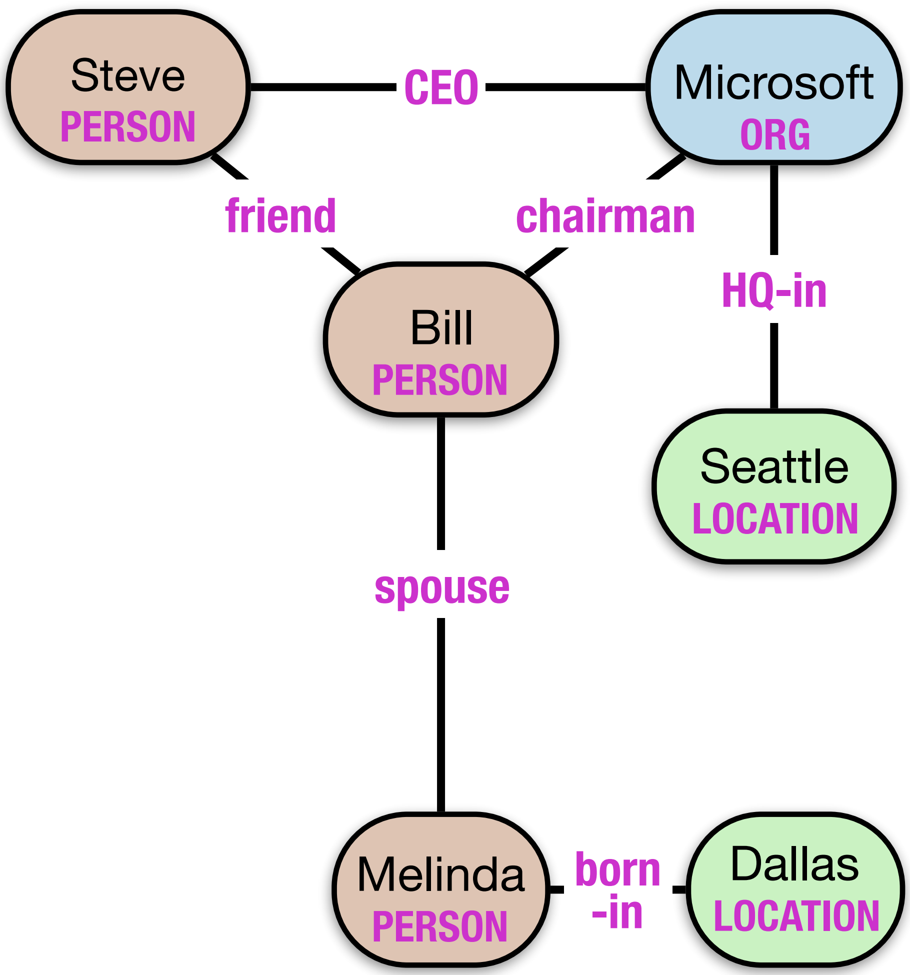


- Text → Mentions → Coref → Relations
- Universal Schema: [AKBC 2012]
  - Entity Types
  - Relation Types

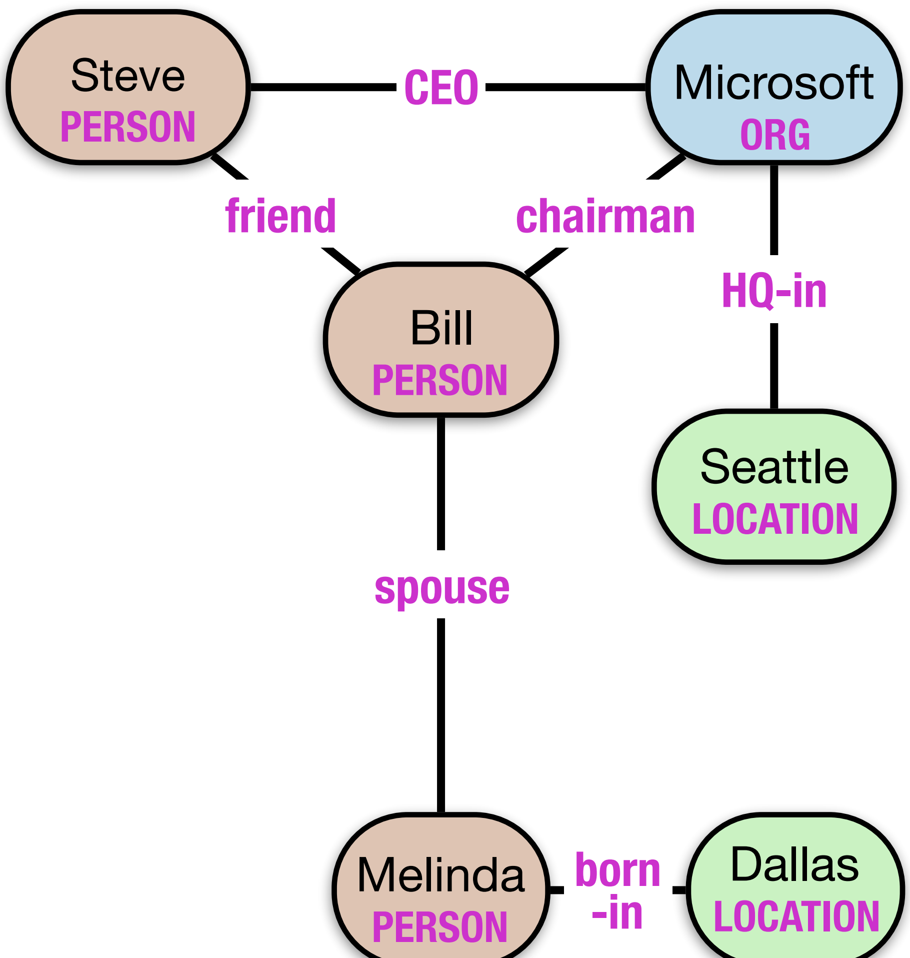
[Riedel, Yao, Marlin, McCallum NAACL 2012]

Universal Schema *relation types*





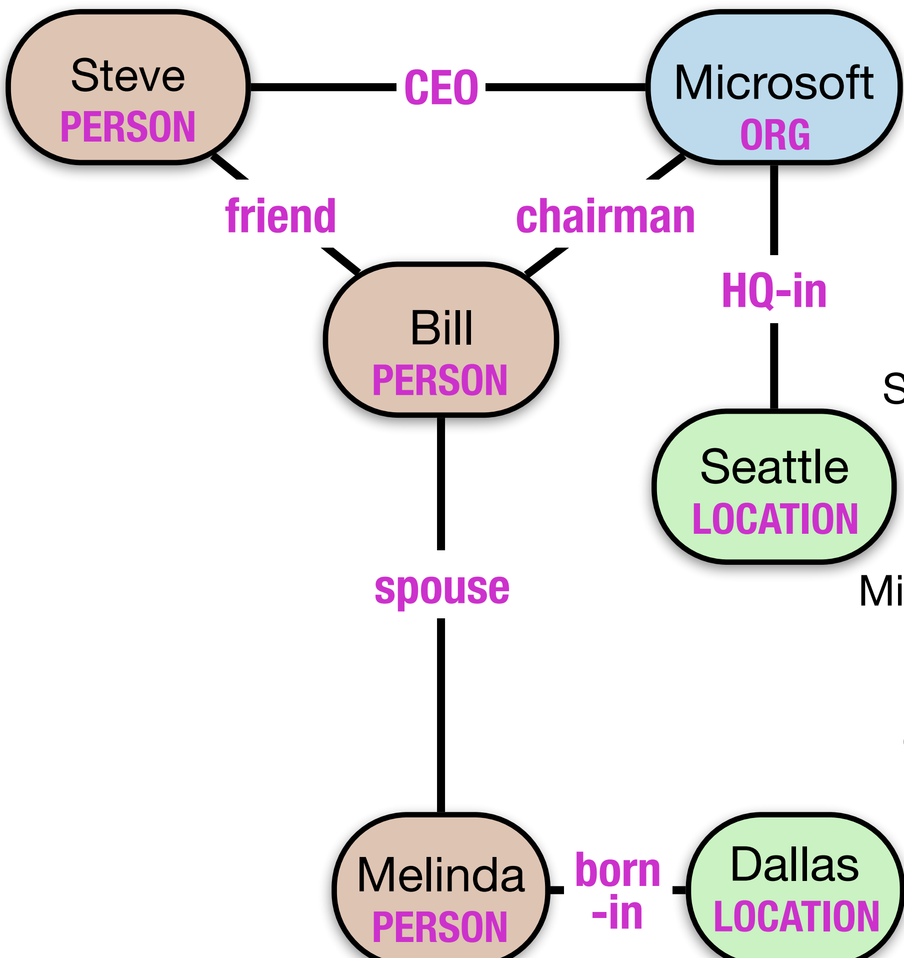
- Text → Mentions → Coref → Relations
- Universal Schema: [AKBC 2012]
  - Entity Types
  - Relation Types



[Riedel, Yao, Marlin, McCallum NAACL 2012]

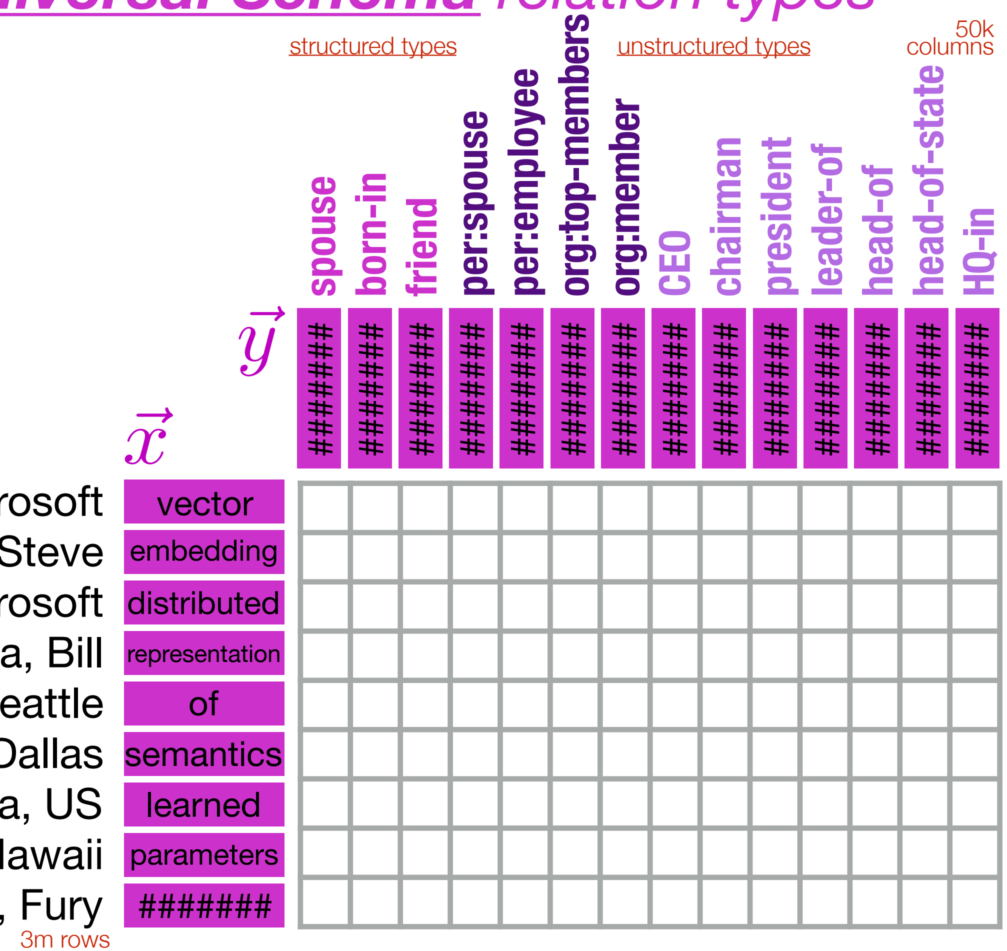
Universal Schema *relation types*

- Text → Mentions → Coref → Relations
- Universal Schema: [AKBC 2012]
  - Entity Types
  - Relation Types



[Riedel, Yao, Marlin, McCallum NAACL 2012]

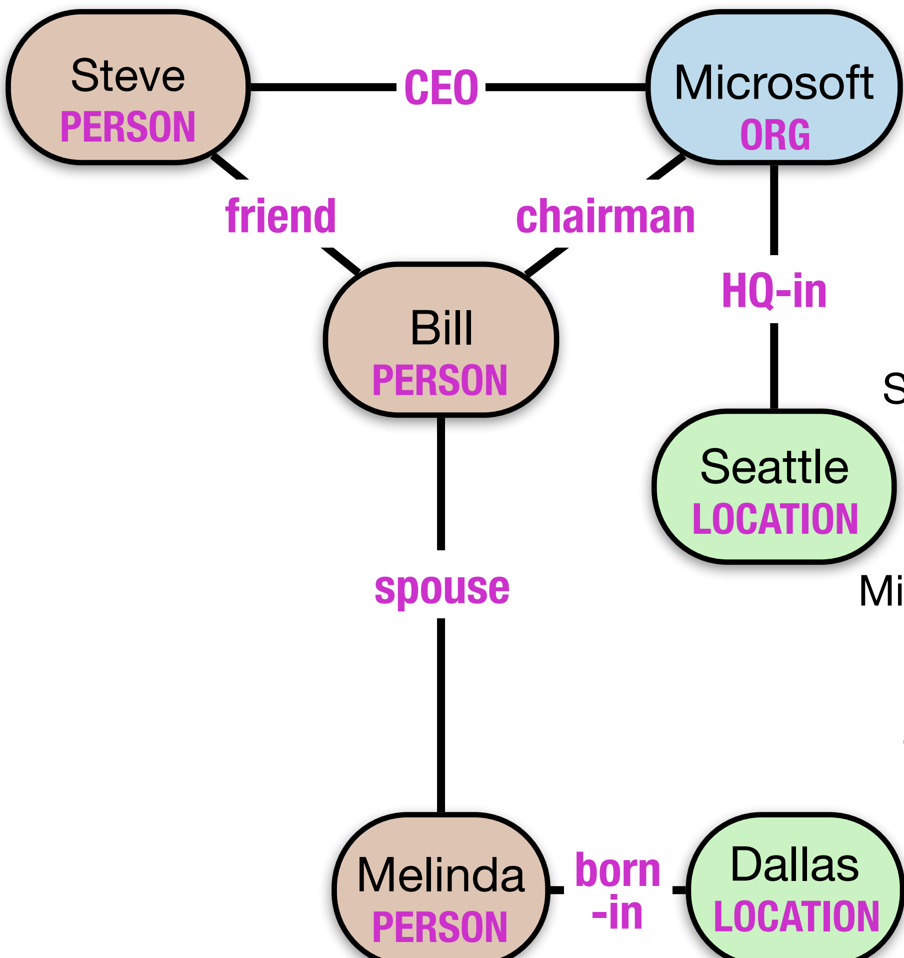
Universal Schema relation types



$$f_{e_1,e_2,r} = \sigma(\vec{x}_{e_1,e_2} \cdot \vec{y}_r)$$

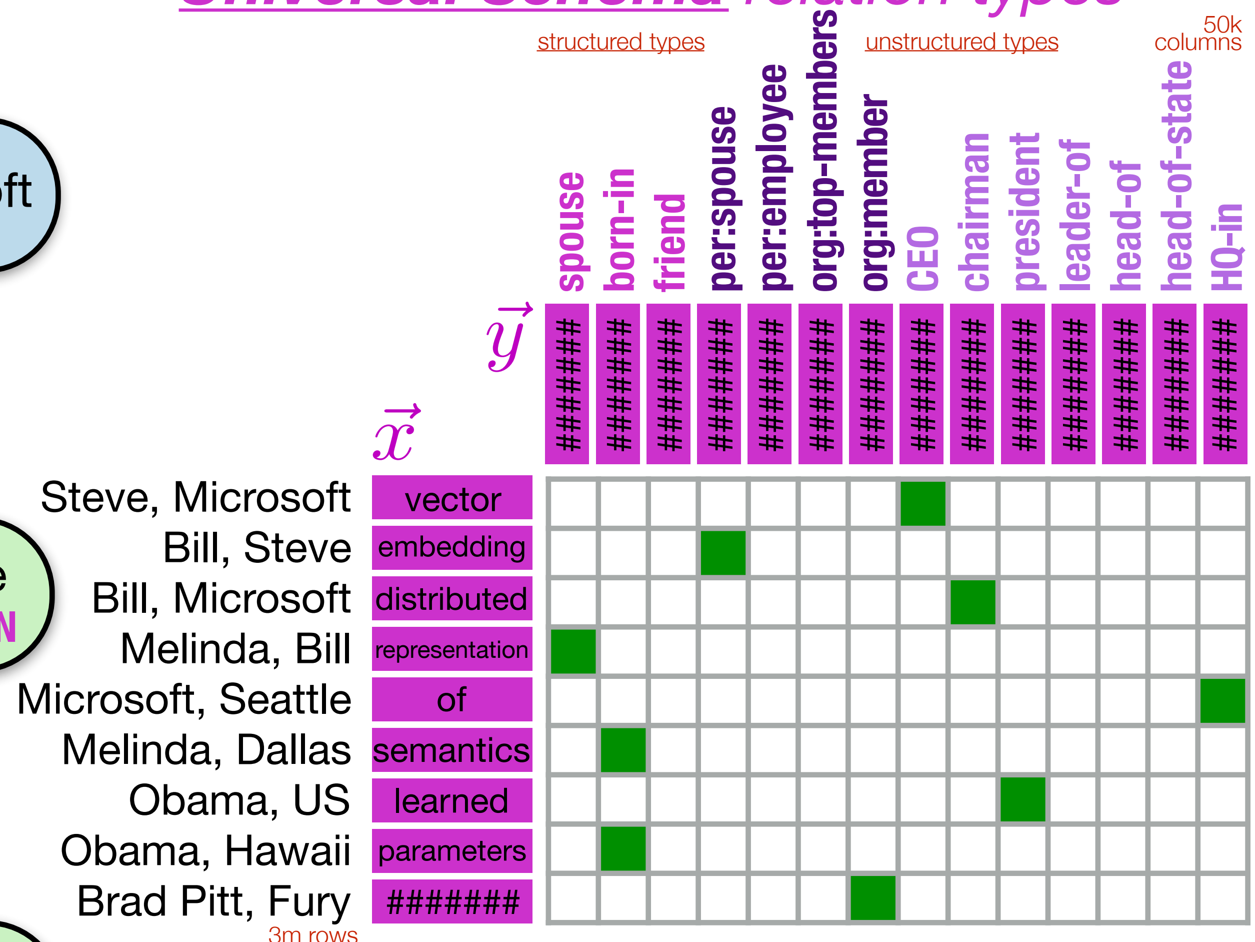
$$\sigma(\theta) = \frac{1}{1+\exp(-\theta)}$$

- Text → Mentions → Coref → Relations
- Universal Schema: [AKBC 2012]
  - Entity Types
  - Relation Types



[Riedel, Yao, Marlin, McCallum NAACL 2012]

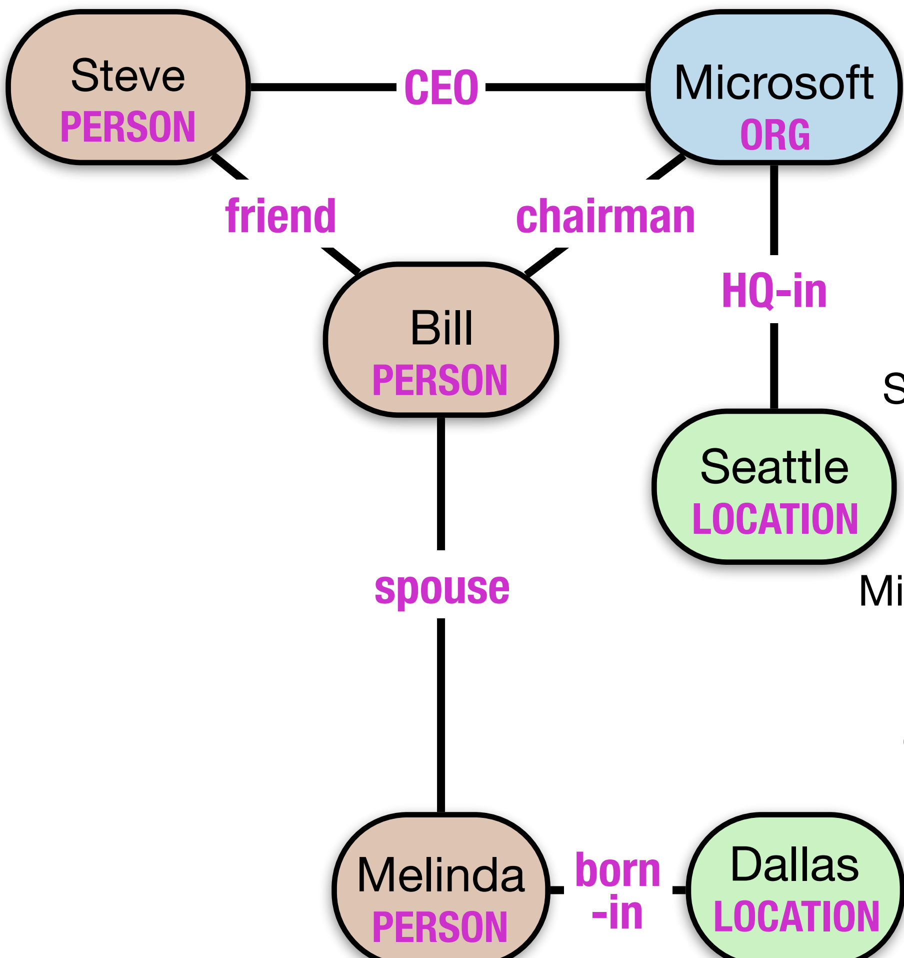
Universal Schema *relation types*



$$f_{e_1,e_2,r} = \sigma(\vec{x}_{e_1,e_2} \cdot \vec{y}_r)$$

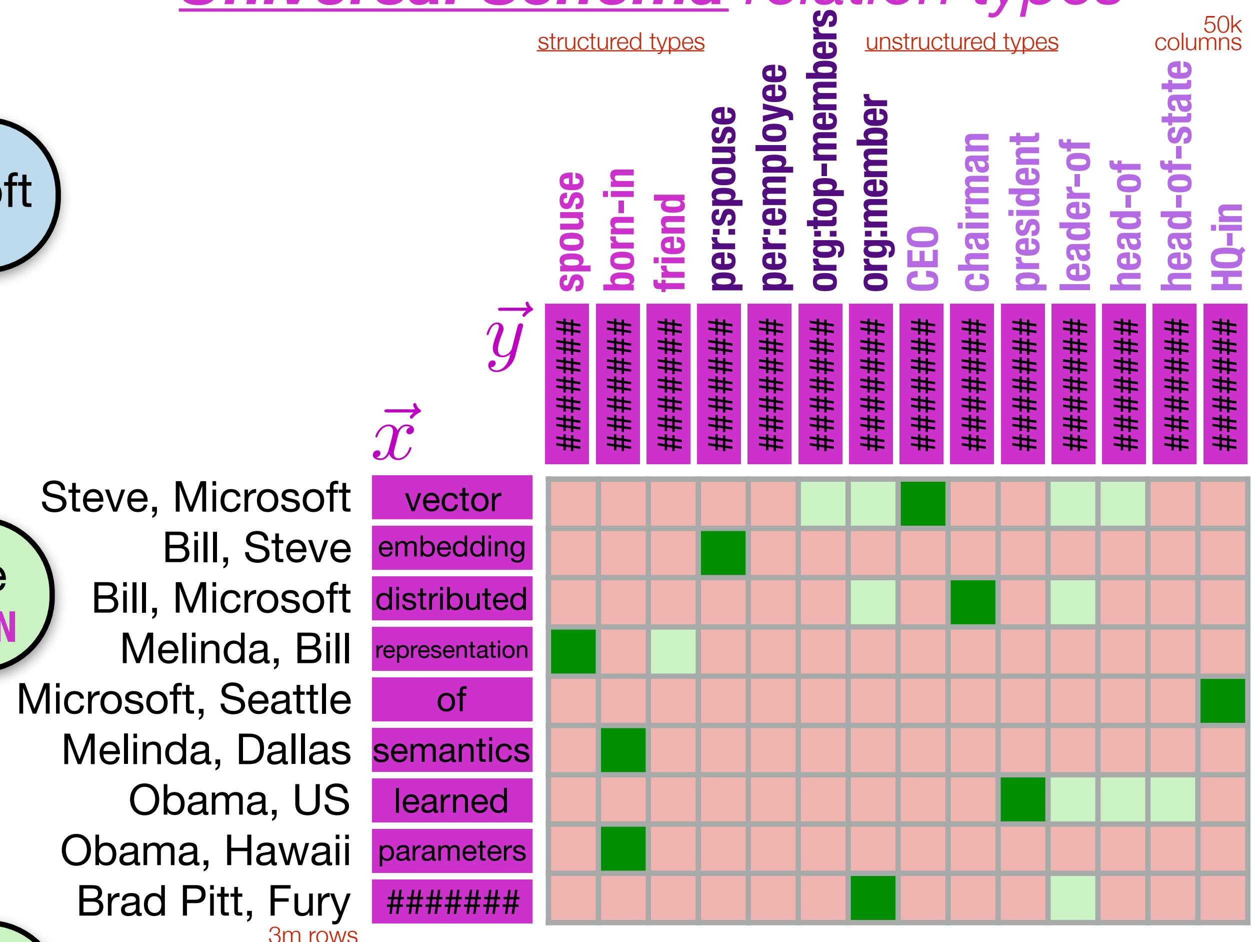
$$\sigma(\theta) = \frac{1}{1+\exp(-\theta)}$$

- Text → Mentions → Coref → Relations
- Universal Schema: [AKBC 2012]
  - Entity Types
  - Relation Types



[Riedel, Yao, Marlin, McCallum NAACL 2012]

Universal Schema *relation types*



$$f_{e_1,e_2,r} = \sigma(\vec{x}_{e_1,e_2} \cdot \vec{y}_r)$$

$\sigma(\theta) = \frac{1}{1+\exp(-\theta)}$

[Yao, Riedel, McCallum, AKBC 2012]

“Universal Schema” Relation Types

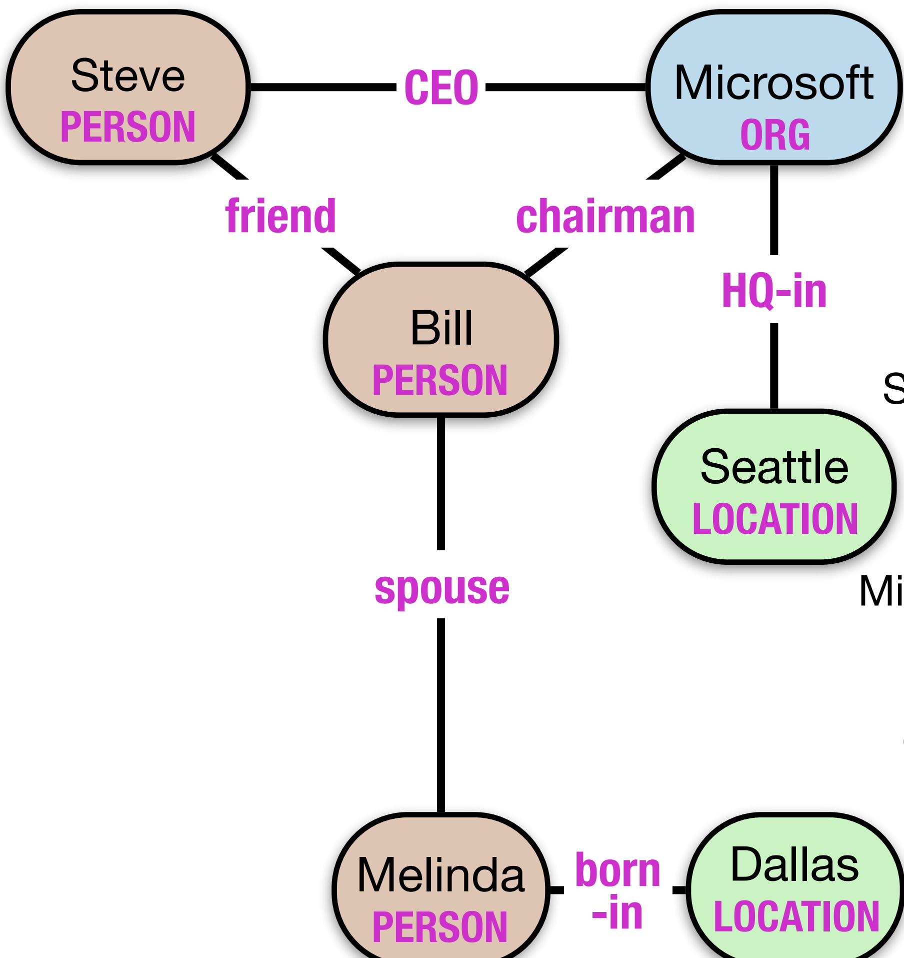
	<subj<**professor**>prep>**at**>	<subj<**historian**>prep>**at**>
Kevin Boyle Ohio State		Y
R. Freeman Harvard	Y	

Learns asymmetric entailment:

PER historian at UNIV → PER professor at UNIV

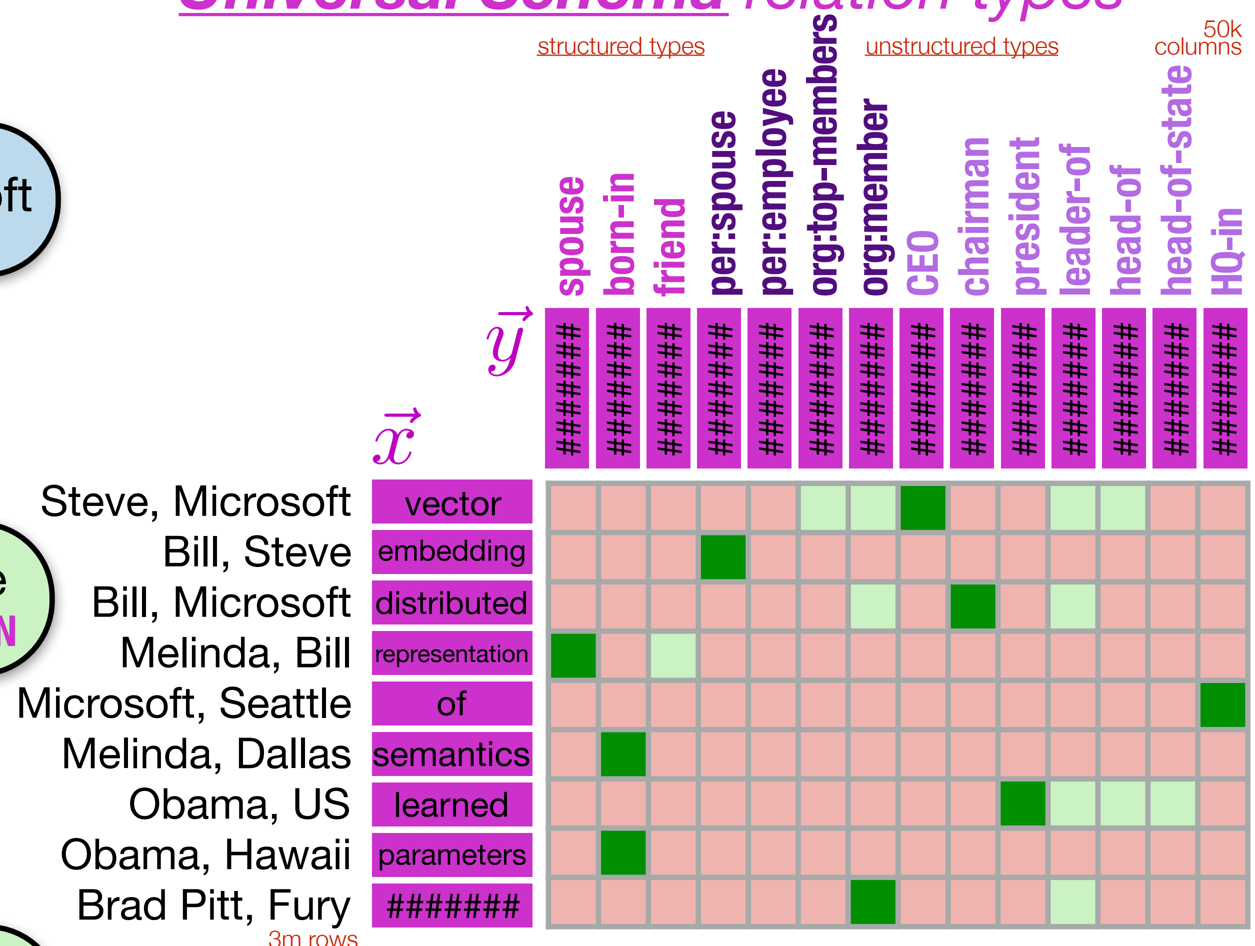
but PER professor at UNIV ↛ PER historian at UNIV

- Text → Mentions → Coref → Relations
- Universal Schema: [AKBC 2012]
  - Entity Types
  - Relation Types



[Riedel, Yao, Marlin, McCallum NAACL 2012]

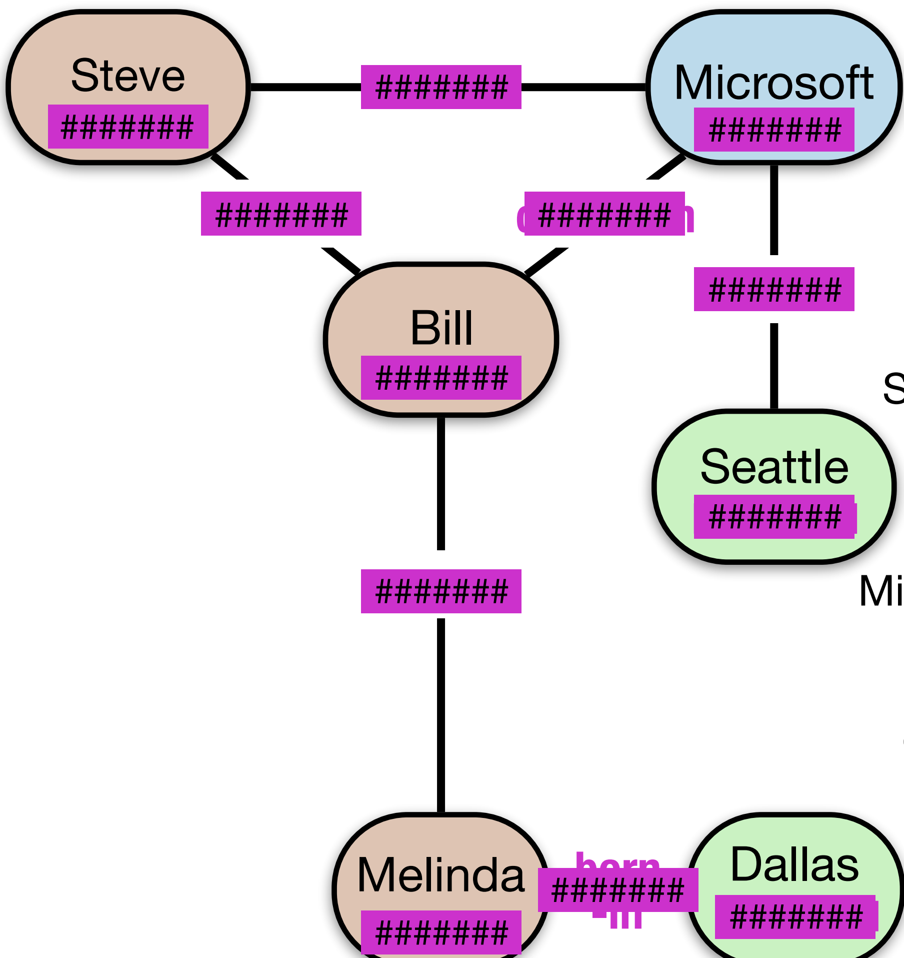
Universal Schema *relation types*



$$f_{e_1,e_2,r} = \sigma(\vec{x}_{e_1,e_2} \cdot \vec{y}_r)$$

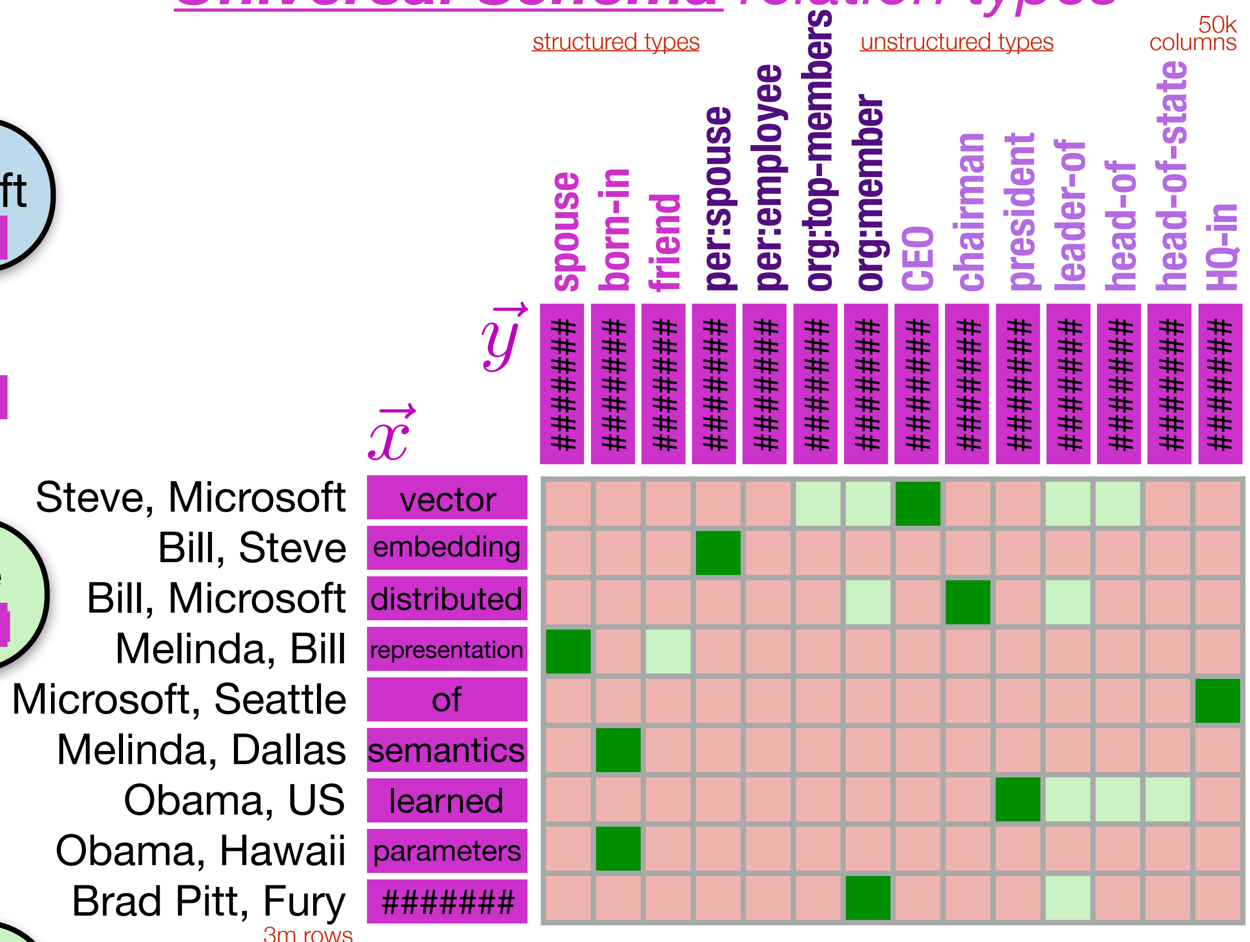
$$\sigma(\theta) = \frac{1}{1+\exp(-\theta)}$$

- Text → Mentions → Coref → Relations
- Universal Schema: [AKBC 2012]
  - Entity Types
  - Relation Types



[Riedel, Yao, Marlin, McCallum NAACL 2012]

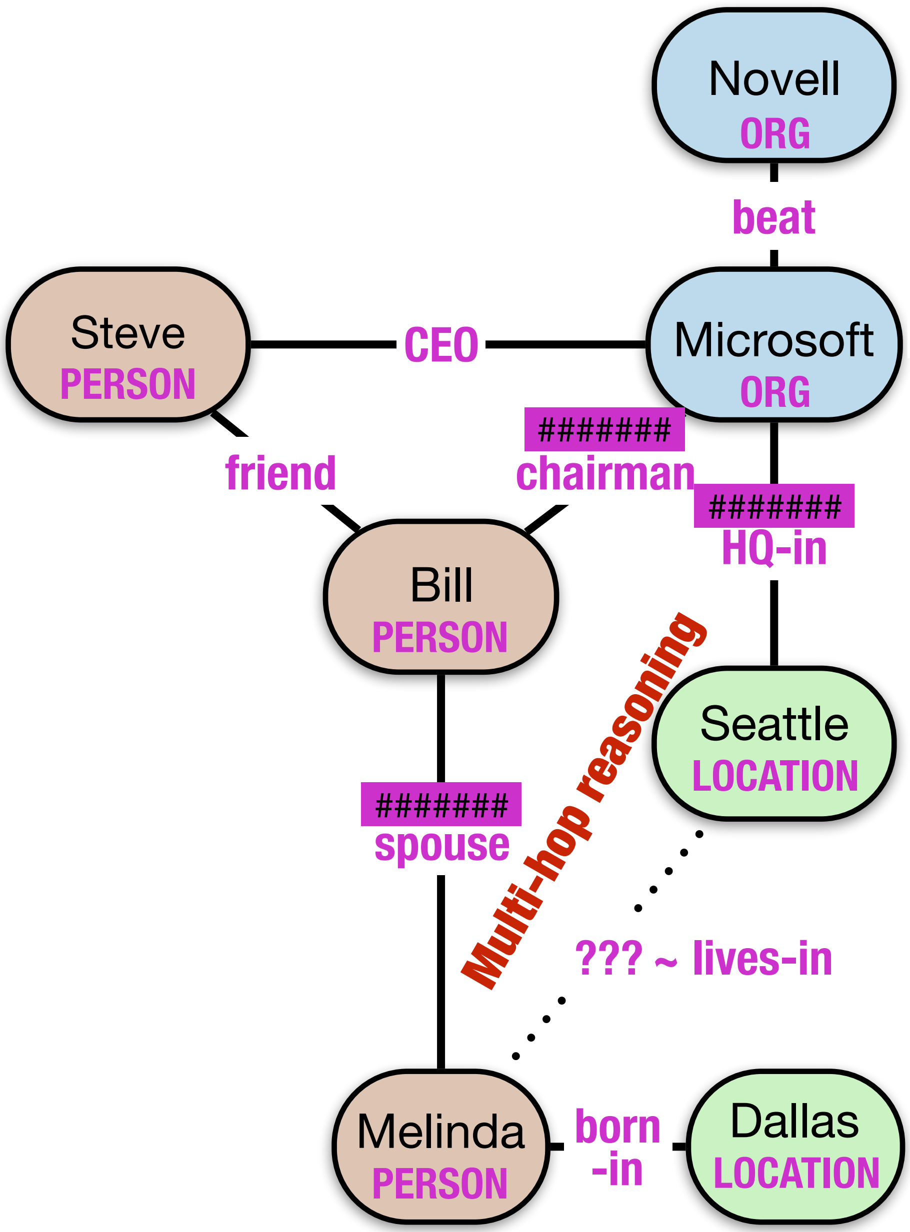
Universal Schema *relation types*



$$f_{e_1,e_2,r} = \sigma(\vec{x}_{e_1,e_2} \cdot \vec{y}_r)$$

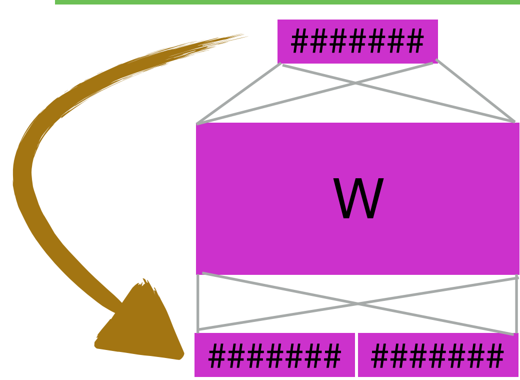
$$\sigma(\theta) = \frac{1}{1+\exp(-\theta)}$$

- Text → Mentions → Coref → Relations
- Universal Schema: [AKBC 2012]
  - Entity Types
  - Relation Types
  - Implicature of implicit info



Spouse(A,B) & Chairman(B,C) & HQ-in(C,D) → Lives-in(A,D)
Spouse(A,B) & **CEO(B,C)** & HQ-in(C,D) → Lives-in(A,D)
Spouse(A,B) & **COO(B,C)** & HQ-in(C,D) → Lives-in(A,D)
Child-of(A,B) & COO(B,C) & HQ-in(C,D) → Lives-in(A,D)

Chain of reasoning on vectors



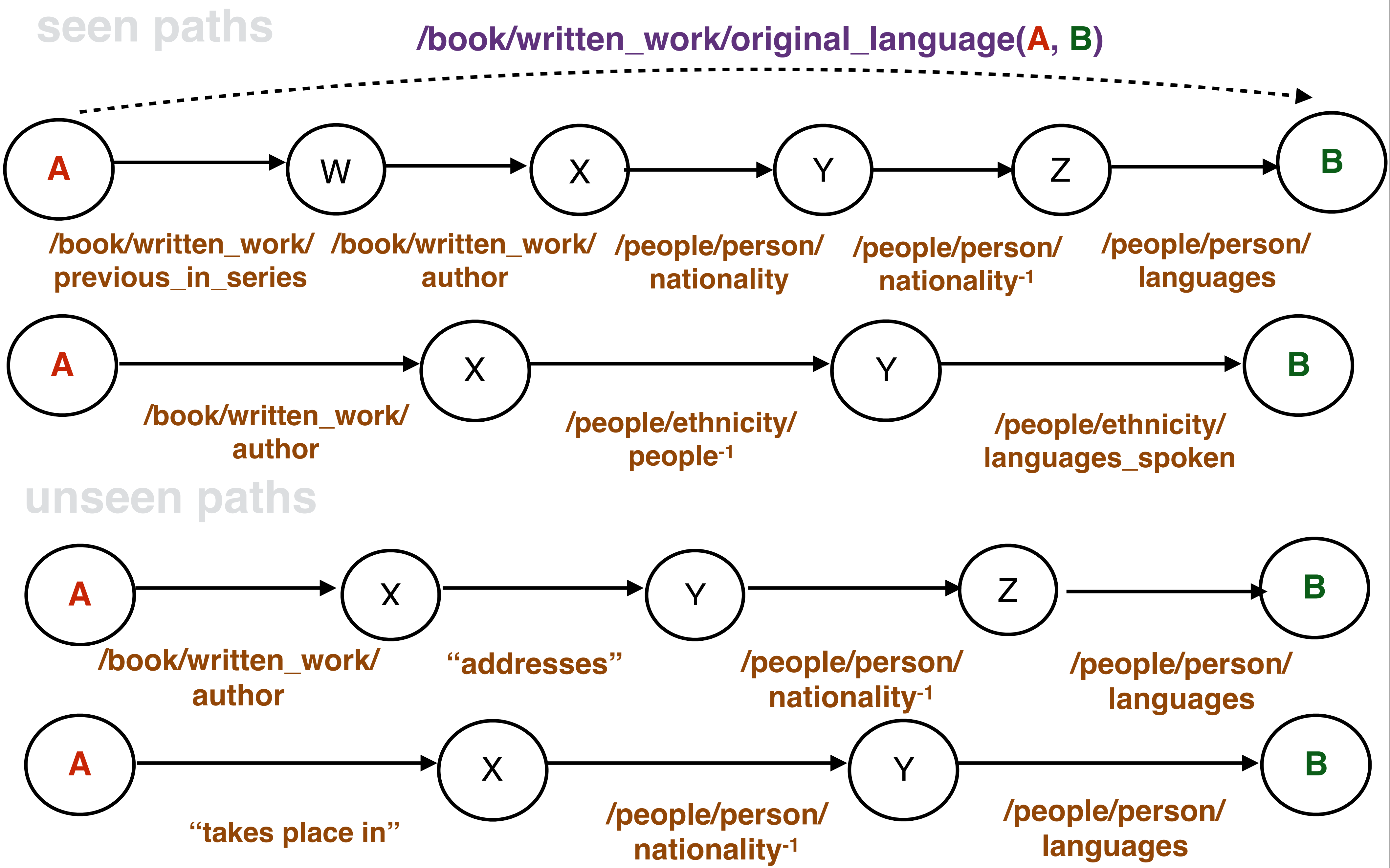
Deep Recursive Neural Network

[Neelakantan, Roth, McCallum, 2015]

Data

Entities	**18M**
Freebase triples	**40M**
ClueWeb triples	**12M**
Relation types	**25,994**

Predictive Paths



Applications & Collaborations

- OpenReview.net
- MIT Material Science
- US Patent Office
- Meta.com

OpenReview.net

- Backend API
- ICLR 2017
  UAI 2017
  ...
- ArXiv overlay
- lightweight "reviewing entities"
- Experimentation & social science on peer review culture

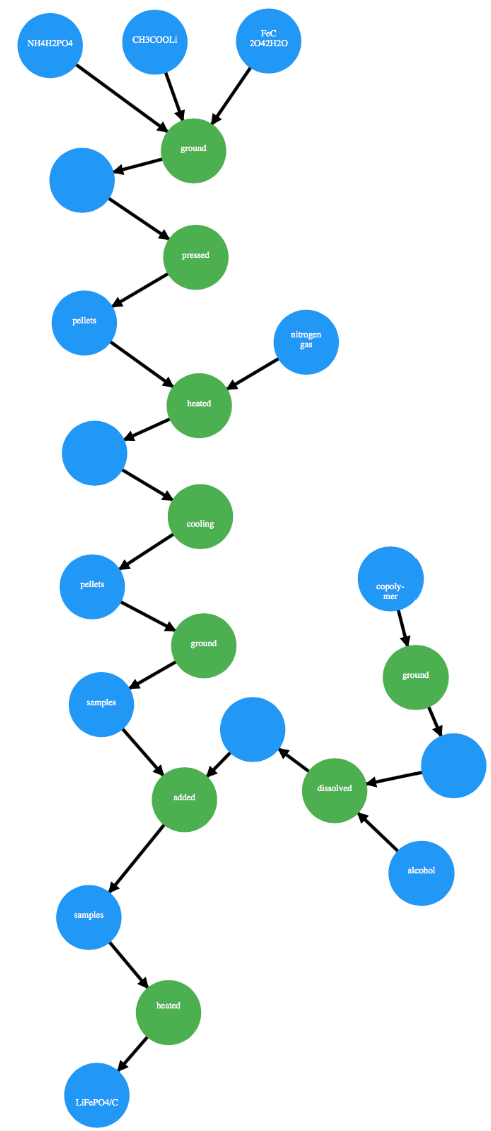
MIT Material Science

Recipe paragraphs from 300k papers

Extracted recipe structure

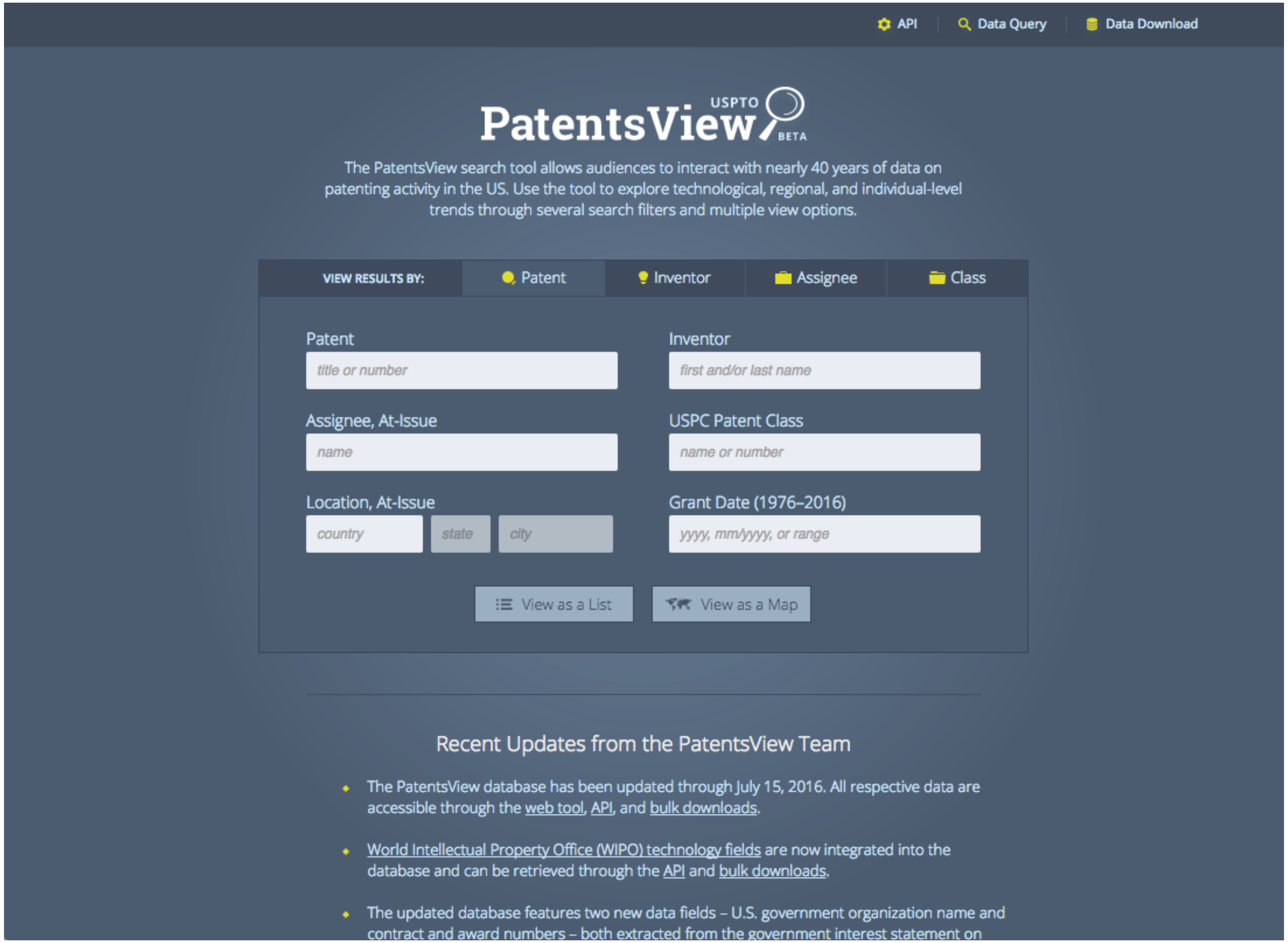
2.1. Synthesis procedure

$LiFePO_4$ was synthesized from a stoichiometric mixture of reagent grade $NH_4H_2PO_4$ (Alfa-Aesar), CH_3COOLi (Aldrich), and $FeC_2O_4{\cdot}2H_2O$ (Aldrich) by a conventional solid-state reaction method. These materials were ground for 20 min, then pressed into pellets and heated at 623 K in a quartz-tube furnace with flowing nitrogen gas for 6 h. After slowly cooling to room temperature, pellets were ground again for 20 min and up to 6 wt.% copolymer (guluronic acid) was added to the samples. The guluronic acid powder was ground and dissolved in the alcohol solution. These samples were heated to 973 K at a heating rate of approximately 3 K min^{-1} and held at that temperature for 10 h in order to derive the $LiFePO_4$/C composite materials. After solid-state reaction, the total carbon content of $LiFePO_4$/C powder was measured by EA. These carbons were obtained from the synthesized precursors and guluronic acid.



New recipe ideas

USPTO PatentsView
Inventor Disambiguation Competition



UMass: 1st place. Deploying at USPTO.

Metaα

Mission - Organize and Deliver All of the World's Scientific and Technical Information.

Founded in 2010 • Team of 25+ • Venture Backed

Toronto (HQ) • San Francisco • Montreal

MIT Technology Review · Bloomberg · TheScientist · OUTSELL · VentureBeat · WIRED · TECHVIBES · YAHOO! · TechCrunch

Large Commercial STEM Text-Mining Collection

37	38K	28M	+14M
Major STM Publishers	Serial Titles (Books & Journals)	Closed Access Full-Text Articles	Open Access Full-Text Articles

Metaα's Scientific Knowledge Graph:

3.5B recommendations	**1B** paper-concept matches	**422M** citations	**26M** papers	**16M** genetic elements
20M concepts	**14.5M** researchers	**2M** antibodies	**407K** drugs	**427K** institutes
234K bacteria	**96K** diseases	**85K** products	**36K** journals	**4.6B** Knowledge Graph connections

Summary

- Building and leveraging knowledge bases for science
- **Representation**
  - Knowledge graph: entities (nodes) & relations (edges):
  - ~~symbols~~ → *universal schema* vector embeddings (on nodes & edges)
  - Reasoning by RNN paths through network.
  - Next: efficient search for scientific reasoning by RL through this graph
- **Applications**
  - OpenReview.net (+ KB of all researchers, expertise, career path)
  - MIT Material Science
  - USPTO Patent Inventor Disambiguation
  - Meta.com