New Development in Knowledge Acquisition, Inference, and Applications

The CCF Advanced Disciplines Lectures #65

2015.12.26

Ni Lao

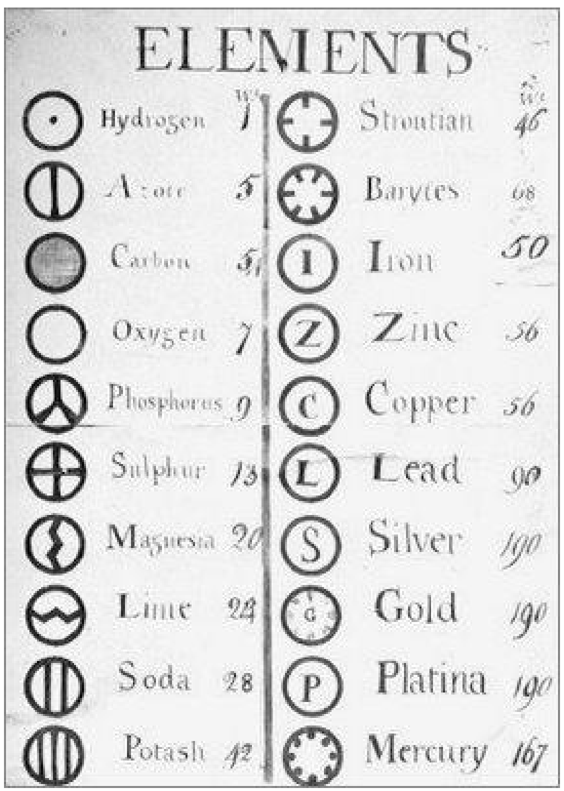
Everything presented here is publicly available.
The opinions stated here are my own, not those of Google.

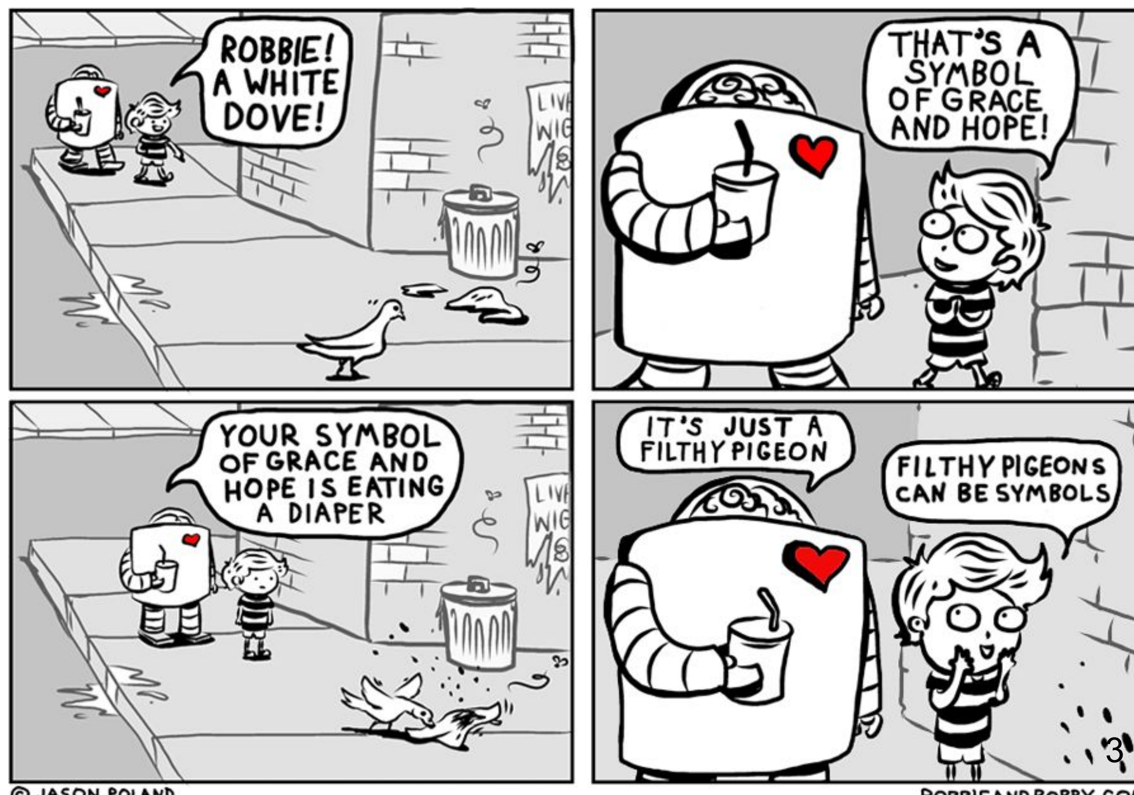
Outline

- KB and AI
  - Symbolism
  - Where does knowledge come from
- KB in action
  - Recommendation [Lao & Cohen 2010]
  - Natural language processing [Lao+ 2015][Zheng+ 2013][Nakashole & Mitchell 2015]
  - Question answering [Liang+ 2013+]
- KB Inference
  - KB completion [Lao+ 2011]
  - Path ranking algorithm [Lao& Cohen, 2010]
  - Efficient path finding [Lao& Cohen, 2011]
  - Longer paths, path with constant [Lao+, 2015]
  - First order logic [Wang+, 2015]
- KB construction & Vector space models
  - Relation extraction [Mintz+ 2009][Lao+ 2012][Dong+ 2014]
  - Open domain information extraction [Fader+ 2011][Fader+ 2013]
  - Vector space models [yao+ 2012][Guu+ 2015]
  - The Web as a KB [Pasupat & Liang 2015]
- Current trends in AI research
  - Modeless
  - Add memory
  - Unsupervised
  - Holistic
  - New applications

Symbolism

- The use of symbols to signify ideas and qualities by giving them symbolic meaning that are different from their literal sense





Why is it important?

- a caveman with no notion of symbolic logic should still do simple reasoning and learn to bicycle

$\cancel{50\ Kili < 500\ Kili}$

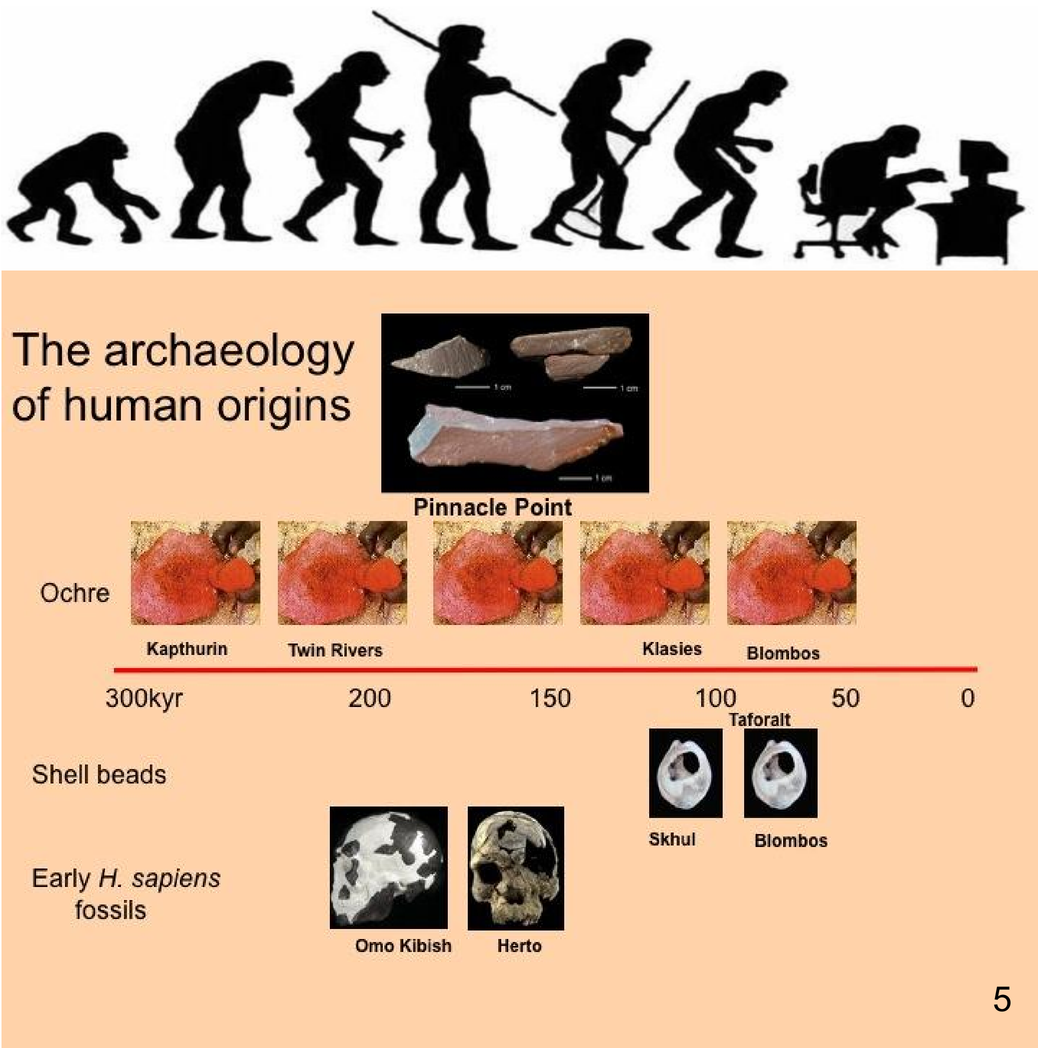
$\cancel{5\ Kili < 50\ Kili}$

$\cancel{F = ma}$

Symbolic Explosion

- "**The Human Revolution**" is a term used by specialists in human origins; it refers to the spectacular and relatively sudden emergence of language, consciousness and culture in our species.

- By 50,000 years ago, an efflorescence of human art, song, dance and ritual – were rippling across the globe



Encephalization quotient

$$E = CS^r$$

- "E" is the weight of the brain,
- "C" is the cephalization factor
- "S" is body weight and
- "r" is the exponential constant

- EQ is the ratio of "C" over the expected "C" of an animal of given its weight "S"

- Symbolism arose as a response to increasing levels of reproductive stress experienced by females during the rapid phase of encephalization

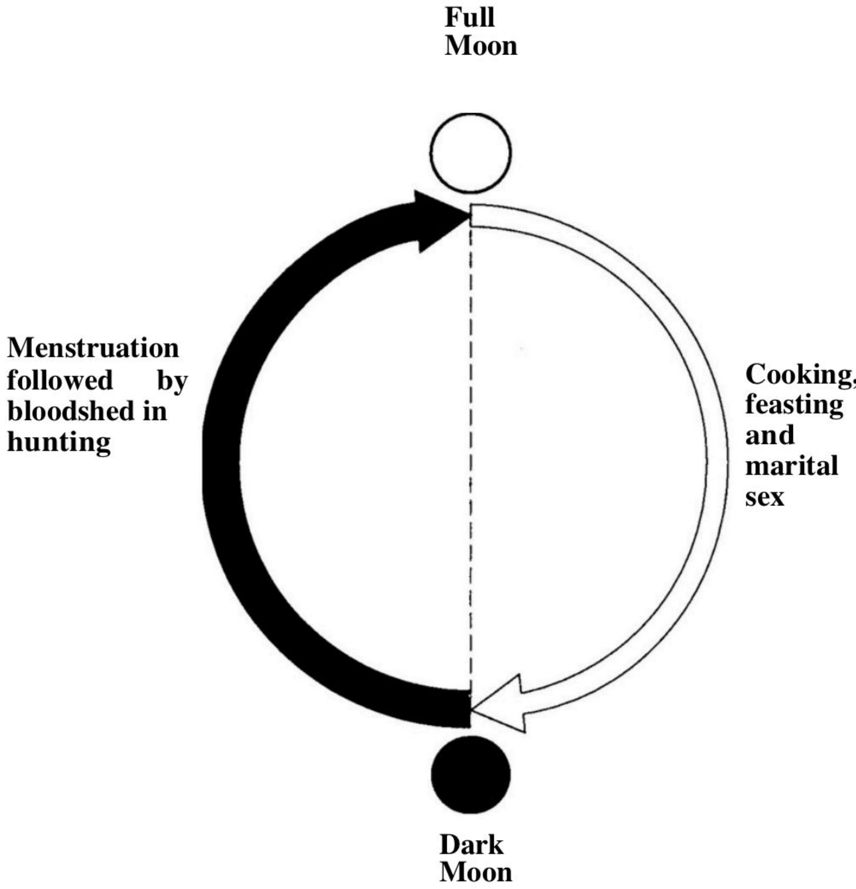
Species	EQ	Species	EQ
Human	7.44	Dog	1.17
Dolphin	5.31	Cat	1
Chimp	2.49	Horse	0.86
Raven	2.49	Sheep	0.81
Monkey	2.09	Mouse	0.5
Elephant	1.87	Rat	0.4
Whale	1.76	Rabbit	0.4

The first symbolism

- Because pronounced menstrual bleeding was valuable for extracting effort from males, even non-cycling females 'cheated' by painting up with red pigments to signal 'imminent fertility'

- It is a signal belonging to an individual, capable of extracting energy from males on a one-to-one basis, has become collectivized among acoalition of females, and amplified, broadcasting information which males cannot afford to ignore

- This 'collective deception' constituted symbolism



The role of symbols in society

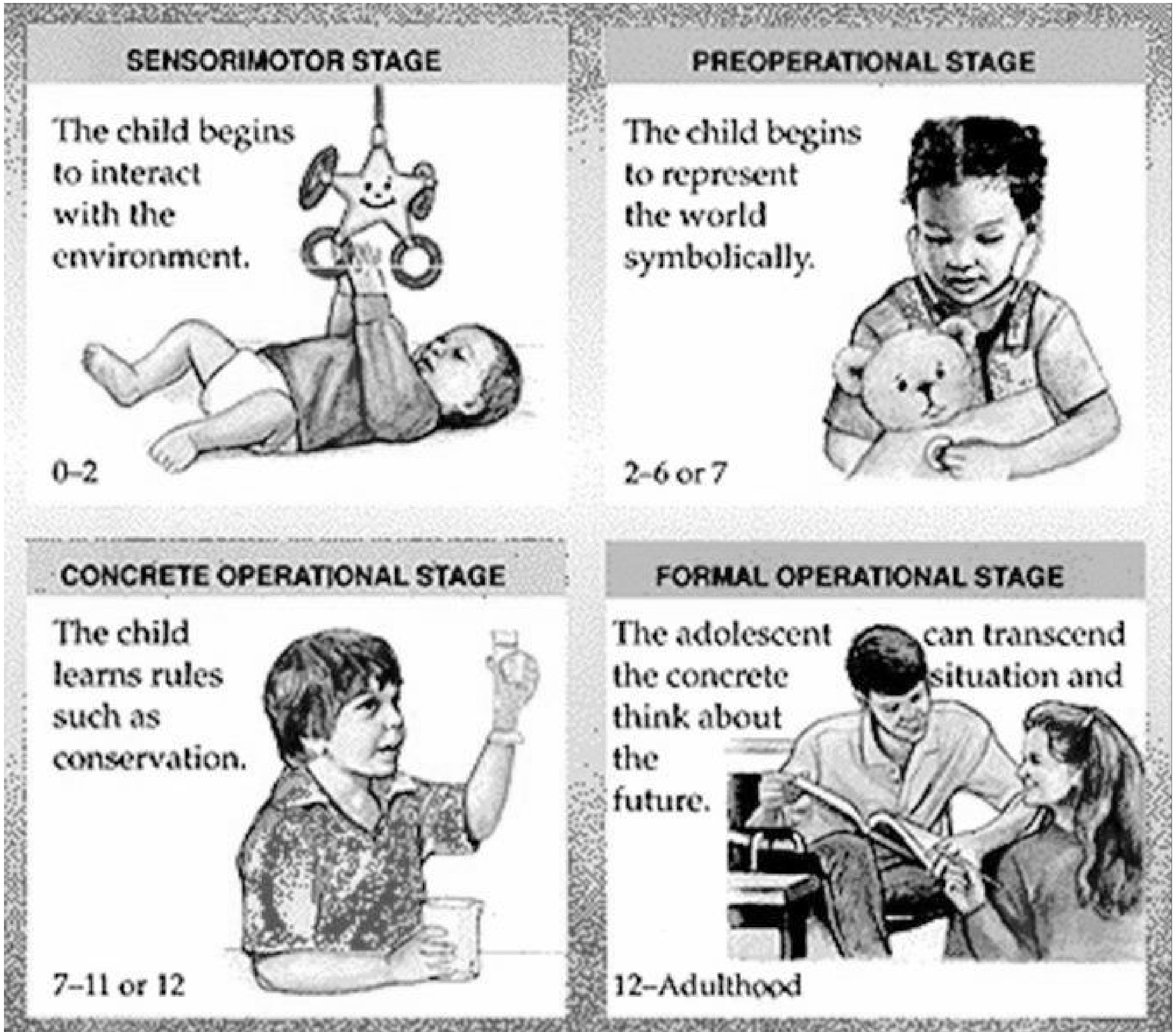
- To act as a medium for the transmission of culture
- To secure the preservation of a group or individual
- To promote social harmony and discord
- To prevent those social sentiments and ideas which are the basis of organized group life from becoming vague and lifeless distractions.

- "Far more powerful than religion, far more powerful than money, or even land or violence, are symbols. Symbols are stories. Symbols are pictures, items, or ideas that represent something else. Human beings attach such meaning and importance to symbols that they can inspire hope, stand in for gods, or convince someone that he or she is dying. These symbols are everywhere around you." — Lia Habel, Dearly, Departed

Theory of cognitive development

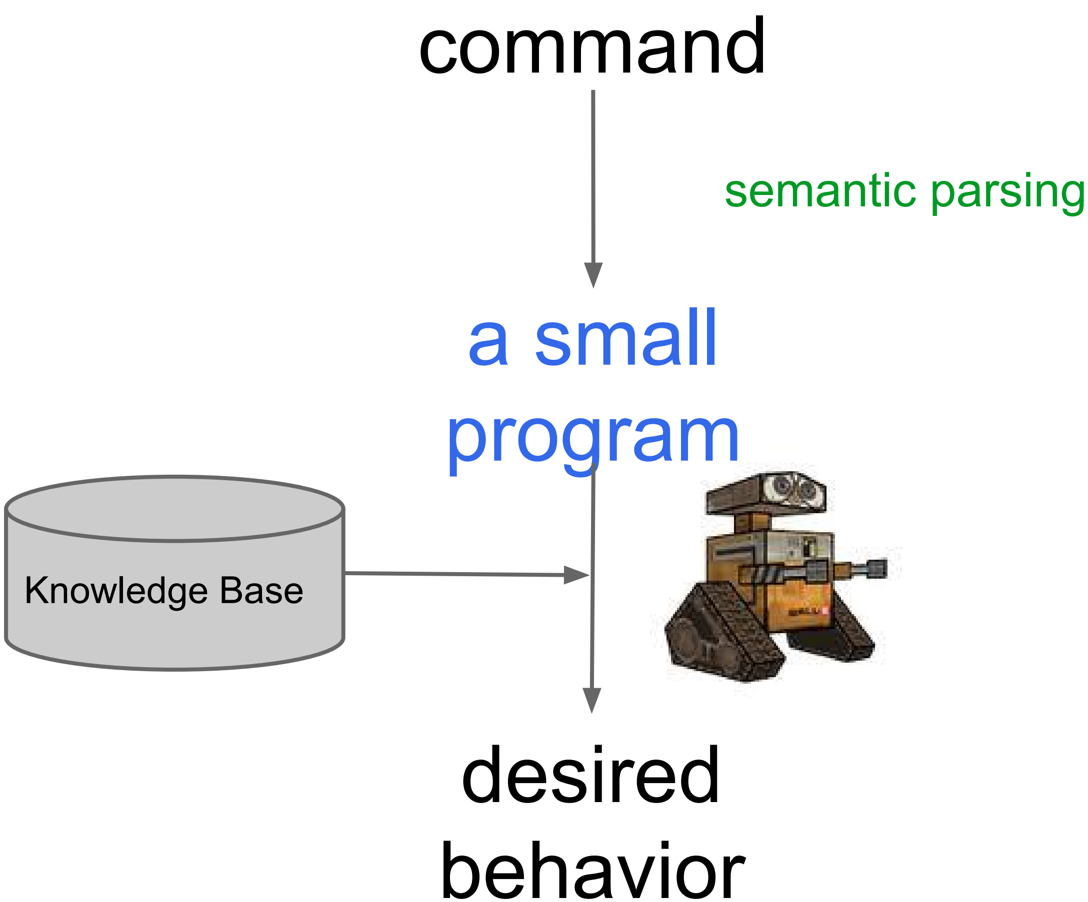
"It is with children that we have the best chance of studying the development of logical knowledge, mathematical knowledge, physical knowledge, and so forth." -- Jean Piaget

- Piaget identified several important milestones in the mental development of children



[Berant+ 2013]
[Liang 2013]

The role of symbols in HCI



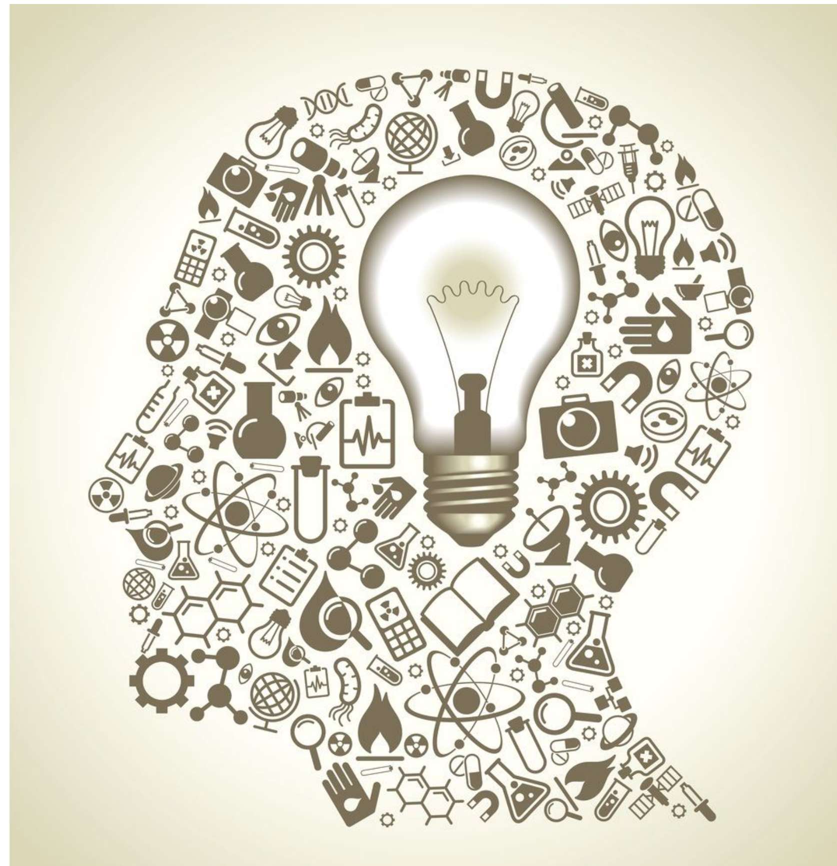
Summary

- Symbolism is very useful for regulating the reasoning process, for storing and communicating ideas, and for human computer interaction

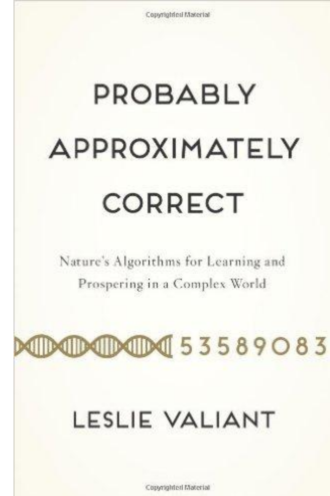
- We will come back to this later for the relationship between connectionism and symbolism

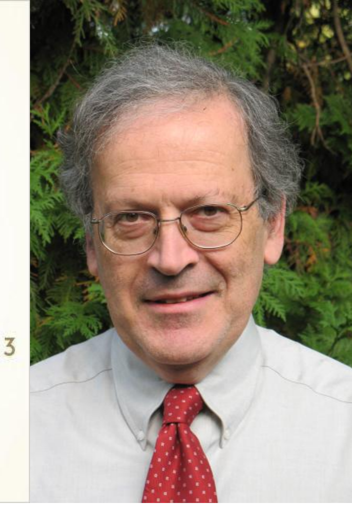
Where does knowledge come from?

- The human brain contains roughly 100 billion neurons each capable of making around 1,000 connections
- Where do we get these 100 TB parameters?
- How many lines of code do I need to write if I want to achieve AI?



The Mind's Eye



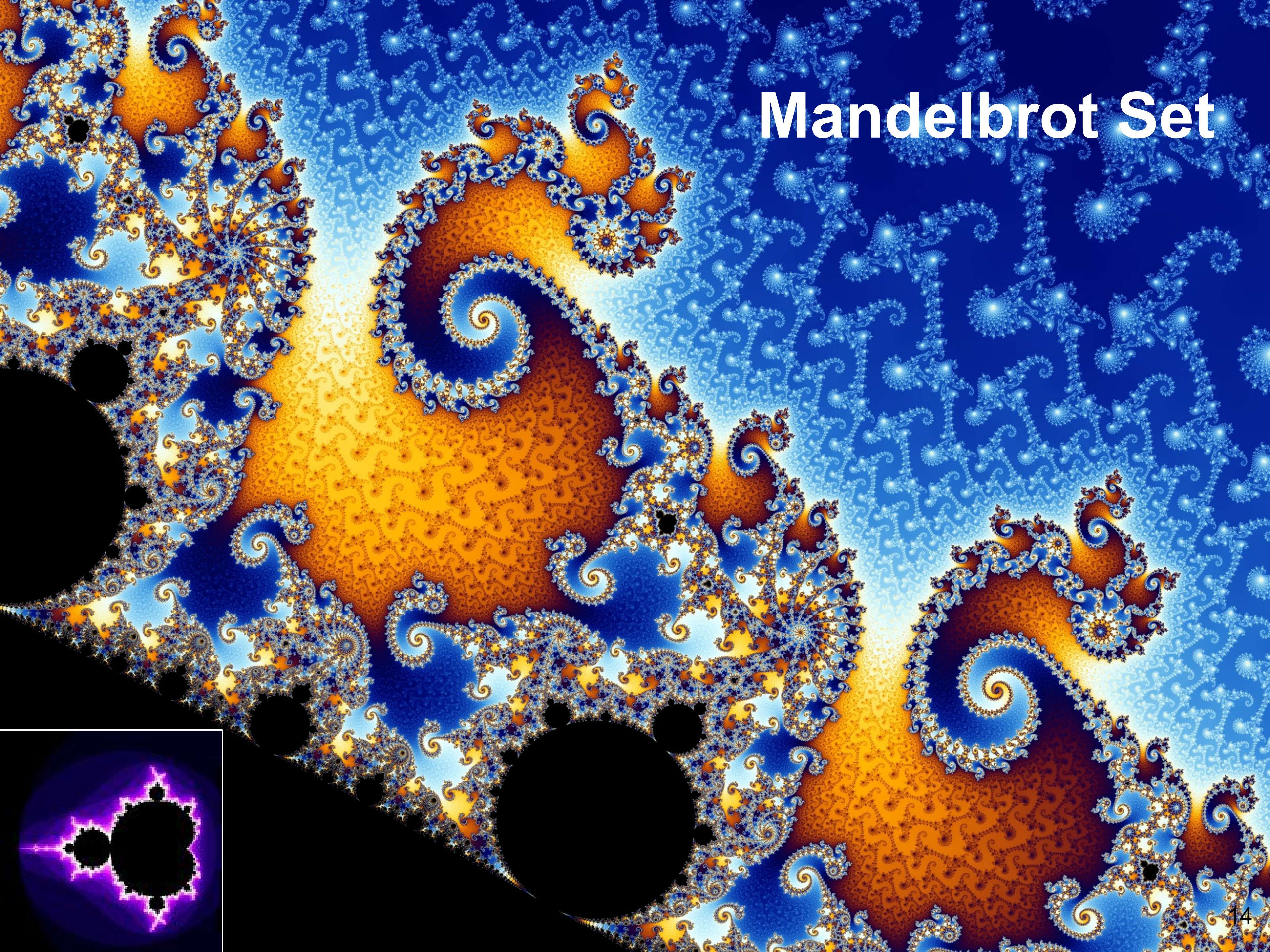


a small machine which can copy large amount of complexity from the world to the brain, while only observes a tiny part of the world at any moment

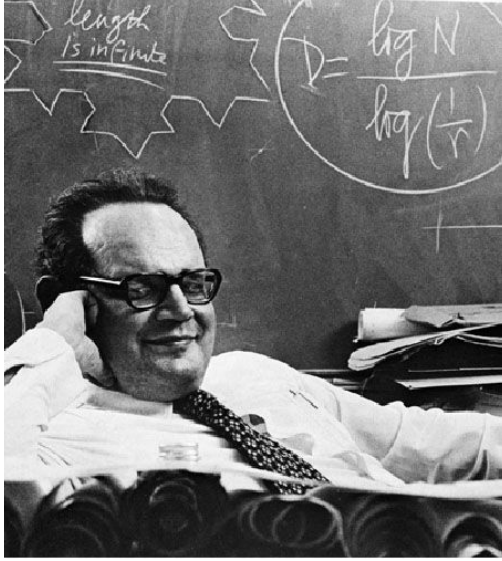
the world

a suitable representation

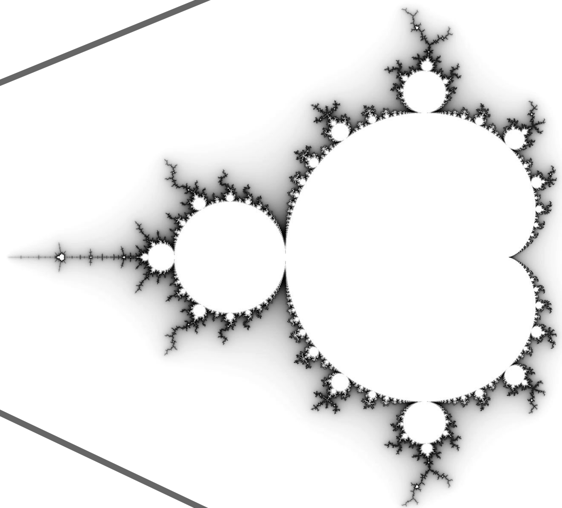
Mandelbrot Set



Mandelbrot Set



the nature
of complex
numbers



$$z_0 = 0$$

$$z_{n+1} = z_n^2 + c$$

$$c \in M \iff \lim_{n \to \infty} |z_{n+1}| \leq 2$$

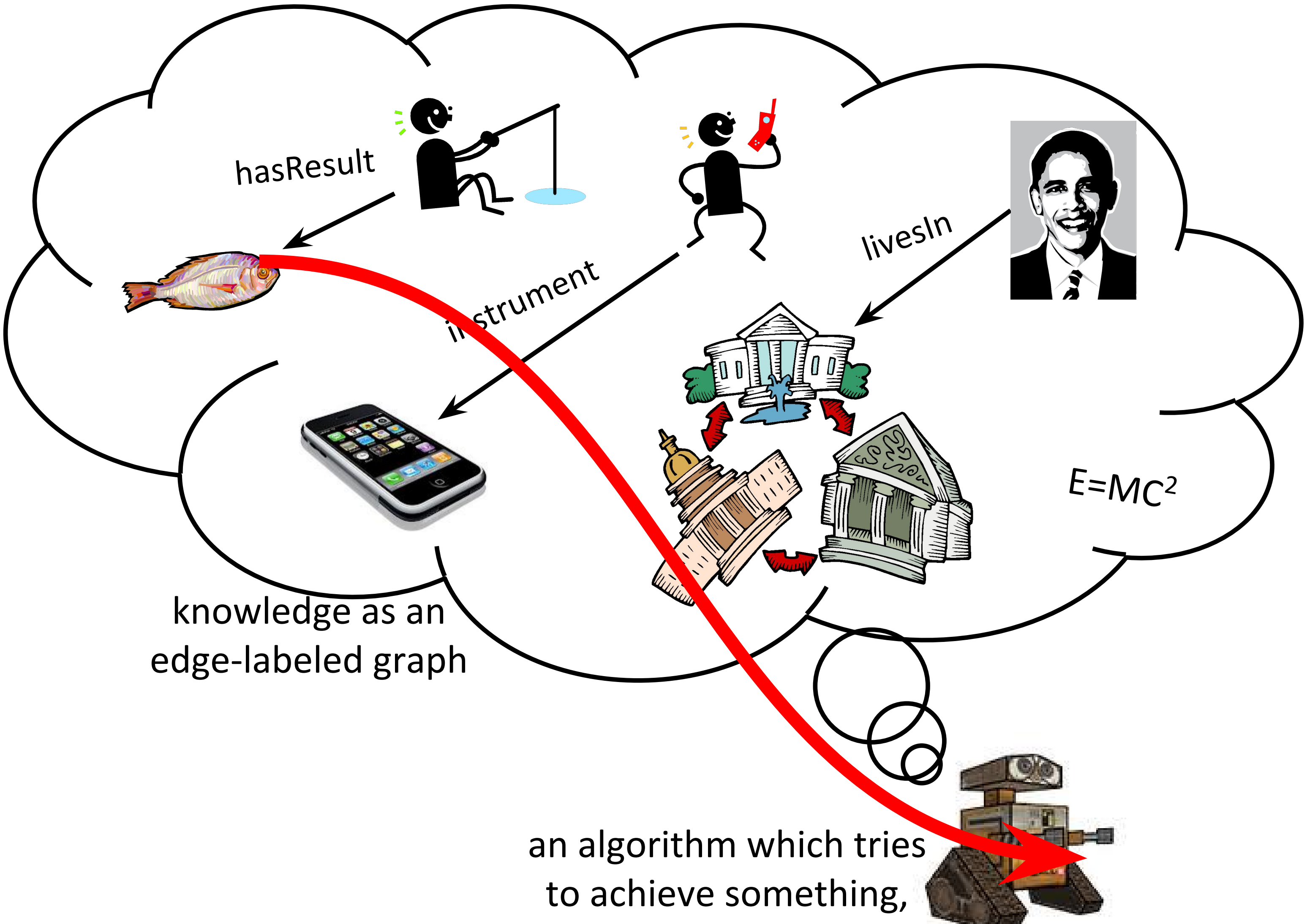
Summary

- We need to discover
  - a small machine which can copy large amount of complexity from the world to the brain
  - a suitable representation for storing worldly knowledge

Outline

- KB and AI
  - Symbolism
  - Where does knowledge come from
- KB in action
  - Recommendation [Lao & Cohen 2010]
  - Natural language processing [Lao+ 2015][Zheng+ 2013][Nakashole & Mitchell 2015]
  - Question answering [Liang+ 2013+]
- KB Inference
  - KB completion [Lao+ 2011]
  - Path ranking algorithm [Lao& Cohen, 2010]
  - Efficient path finding [Lao& Cohen, 2011]
  - Longer paths, path with constant [Lao+, 2015]
  - First order logic [Wang+, 2015]
- KB construction & Vector space models
  - Relation extraction [Mintz+ 2009][Lao+ 2012][Dong+ 2014]
  - Open domain information extraction [Fader+ 2011][Fader+ 2013]
  - Vector space models [yao+ 2012][Guu+ 2015]
  - The Web as a KB [Pasupat & Liang 2015]
- Current trends in AI research
  - Modeless
  - Add memory
  - Unsupervised
  - Holistic
  - New applications

Knowledge itself is power
--Francis Bacon



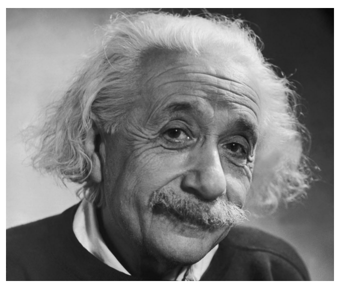
knowledge as an
edge-labeled graph

an algorithm which tries
to achieve something,
e.g. IR/IE/QA/MT

[Lao & Cohen 2010]

Reading Recommendation

a scientist



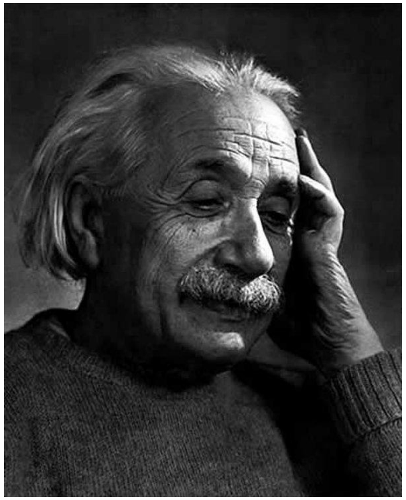
these are interesting papers

a paper stream

[Lao & Cohen 2010]

Reading Recommendation

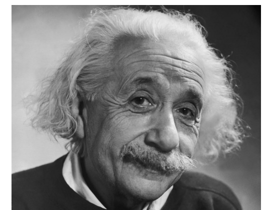
try to study biology

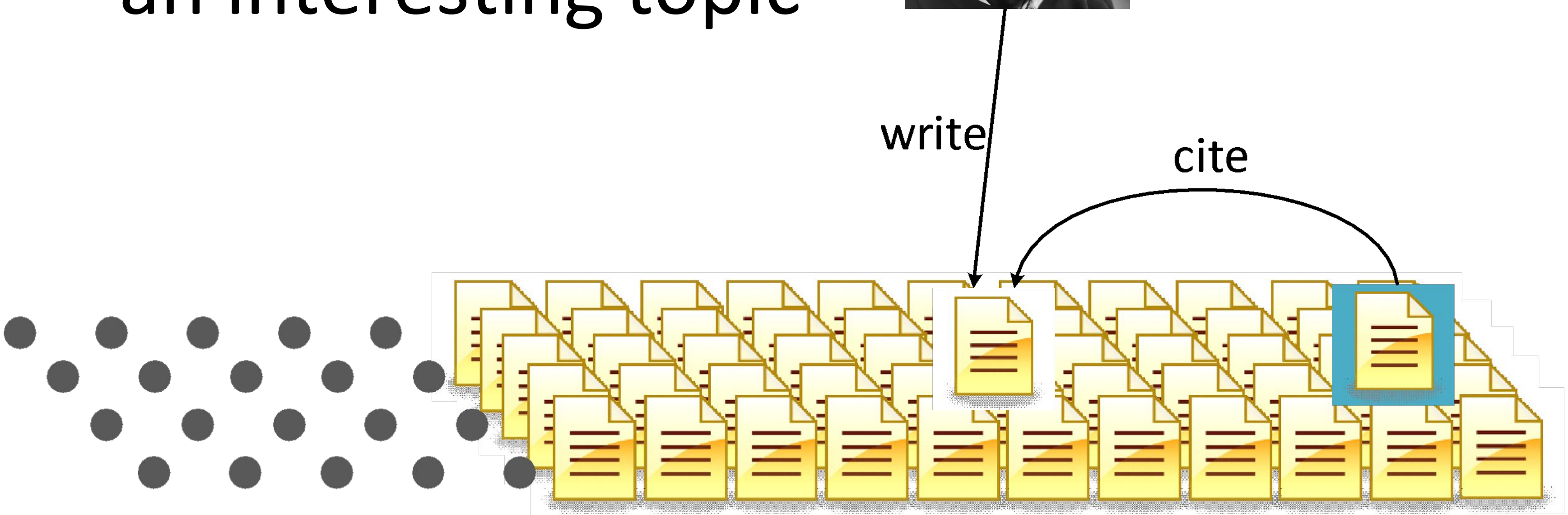


a paper river

Reading Recommendation

new development of
an interesting topic



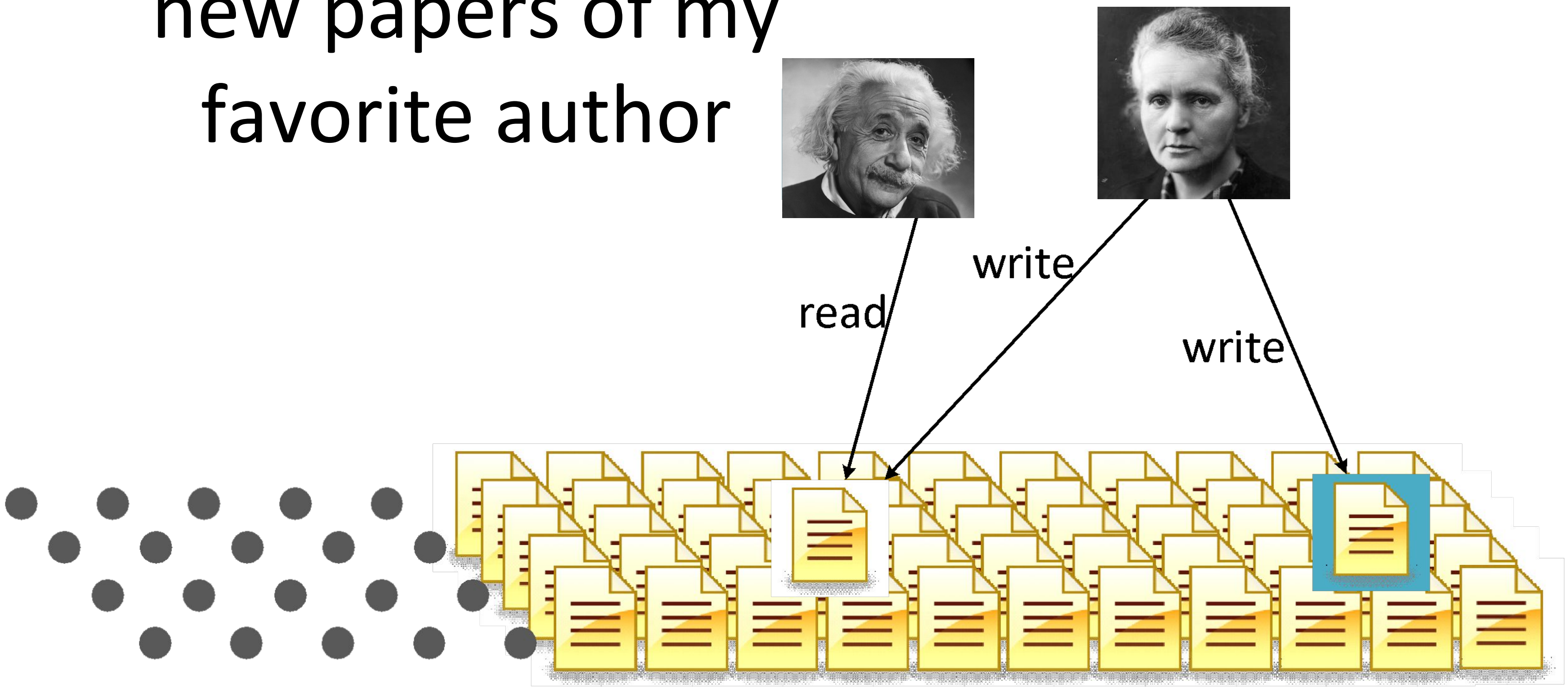


a paper river

[Lao & Cohen 2010]

Reading Recommendation

new papers of my favorite author

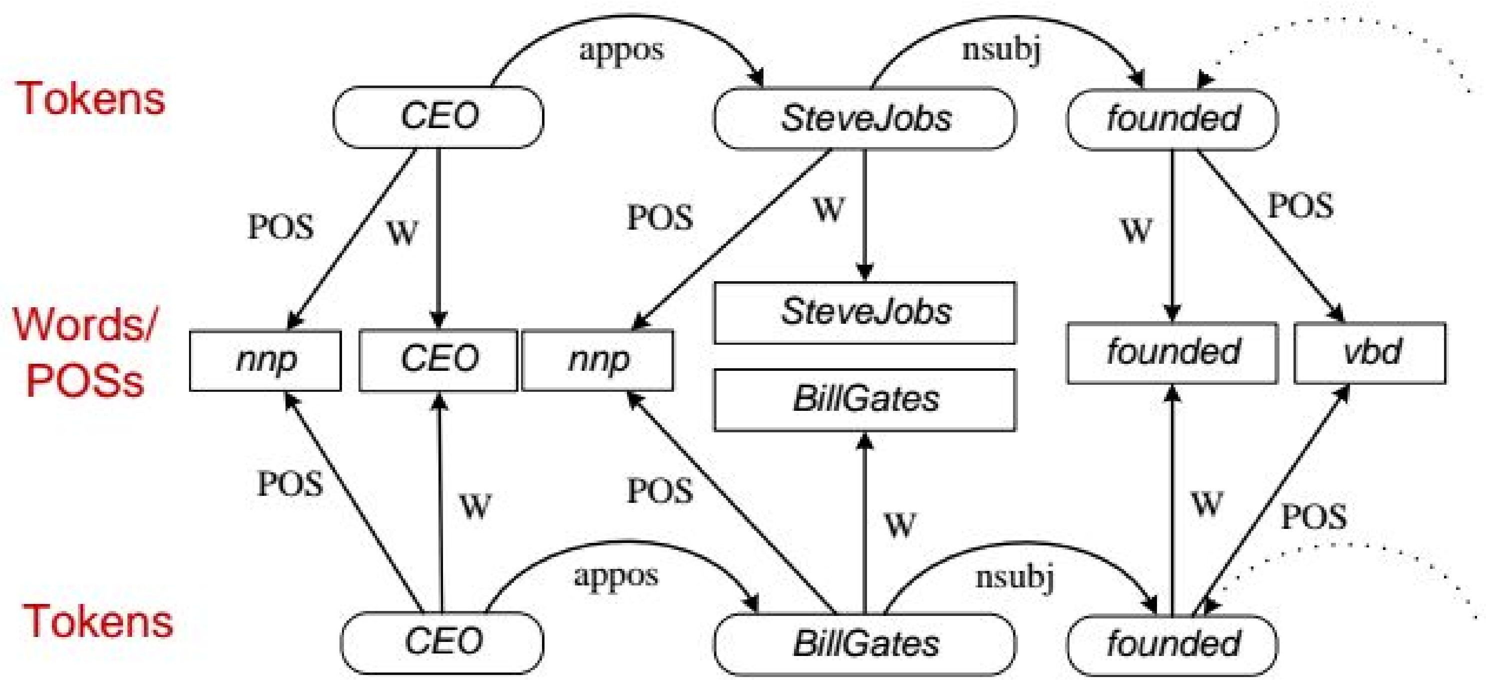


a paper river

[Lao+ 2015]

Entity Extraction with Parsed Text

- Find target nodes that are related to the query nodes over the relation similar-to, or, coordinate-term. e.g.,
  - {steve jobs, larry page} ==> {bill gates, ...}



[Zheng+ 2013]

an AI-complete task

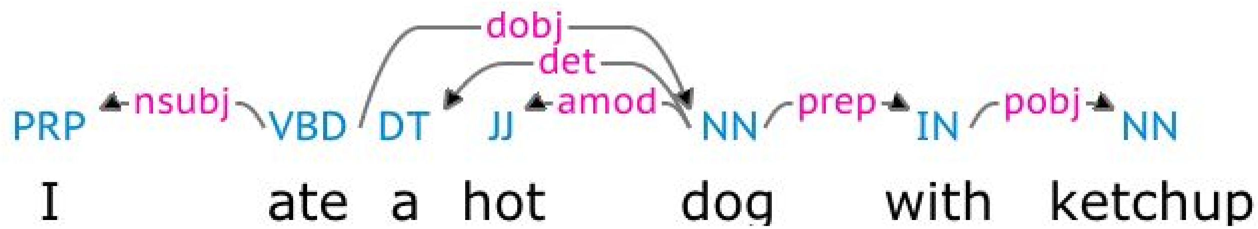
Coreference Resolution

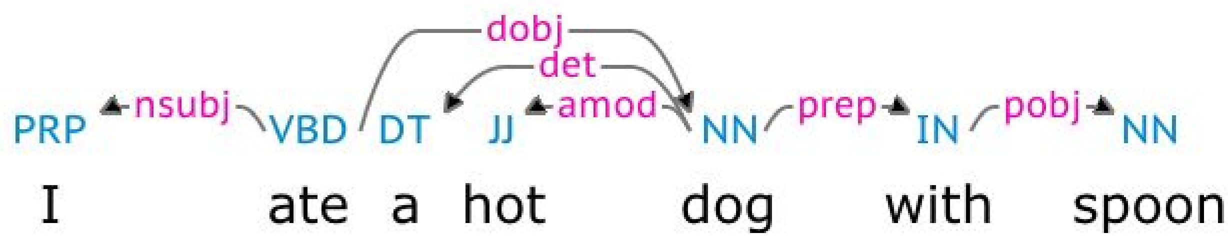
The Chicago suburb of Arlington Heights is the first stop for George W. Bush today. The Texas governor stops in Gore's home state of Tennessee this afternoon ...

- Task:
  - Identify mentions that refer to the same entity.
  - Useful in relation extraction, question answering, machine translation, etc
- coreferencing (m1, m2) requires knowledge that George W. Bush (m1) is the governor of Texas (m2)

Prepositional Phrase Attachment

One of the following parses is wrong





from http://demo.ark.cs.cmu.edu/parse

Prepositional Phrase Attachment

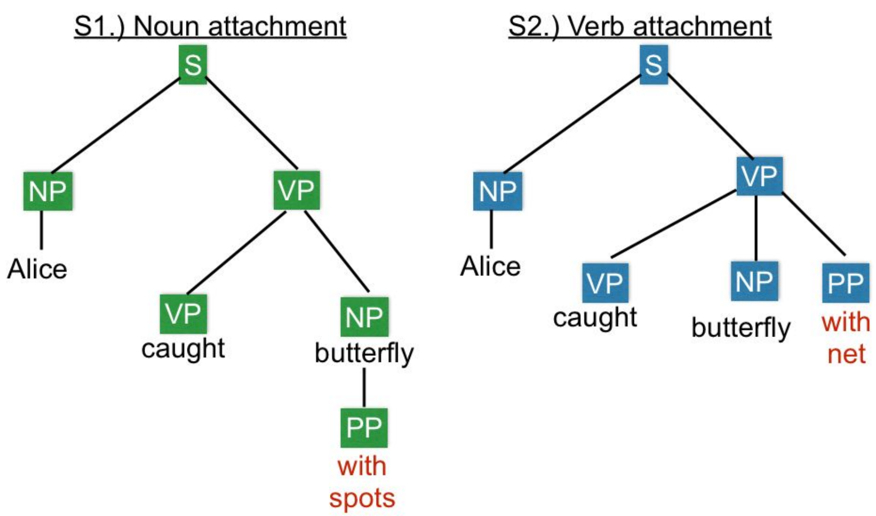


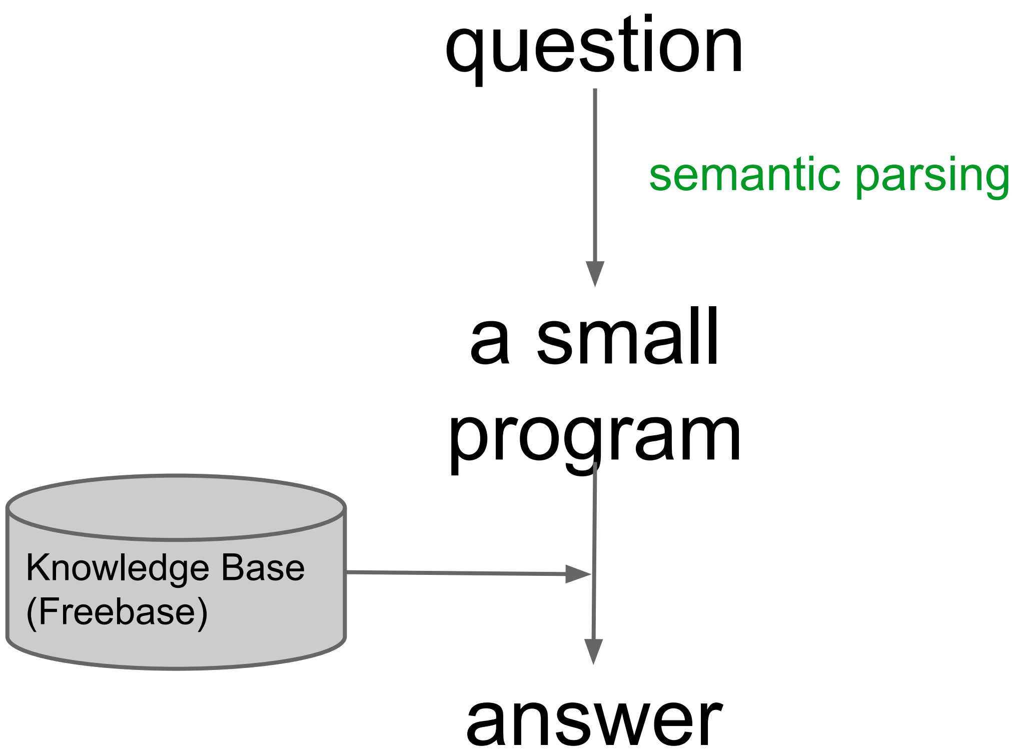
Figure 1: Parse trees where the prepositional phrase (PP) attaches to the noun, and to the verb.

Relations	Noun-Noun binary relations *(Paris, located in, France)* *(net, caught, butterfly)*
Nouns	Noun semantic categories *(butterfly, isA, animal)*
Verbs	Verb roles *caught(agent, patient, instrument)*
Prepositions	Preposition definitions *f(for)= used for, has purpose, ...* *f(with)= has, contains, ...*
Discourse	Context $n0 \in \{n0, v, n1, p, n2\}$

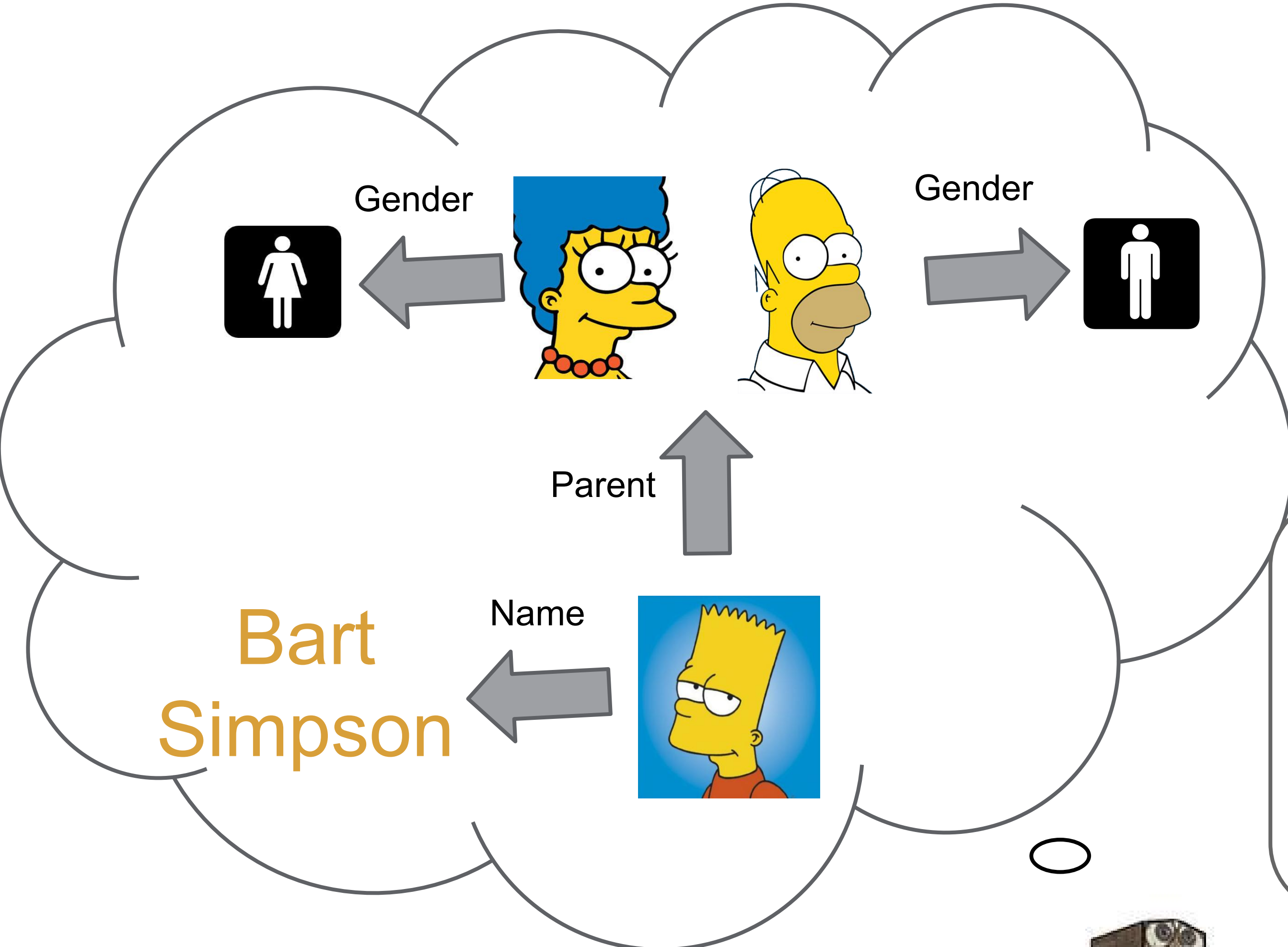
Table 1: Types of background knowledge used in this paper to determine PP attachment.

Semantic Parsing and Question Answering with Knowledge Base

[Berant+ 2013]
[Liang 2013]



[Berant+ 2013]

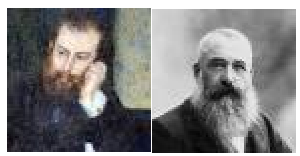


$$\lambda x.\, \exists e.\mathrm{Name}(e, \text{“Bart Simpson”}) \\ \wedge\ \mathrm{Parent}(e, x) \\ \wedge\ \mathrm{Gender}(x, \mathrm{Male/m/male})$$

Compositional Semantics

Text: impressionist painters during the 1920s

Denotations

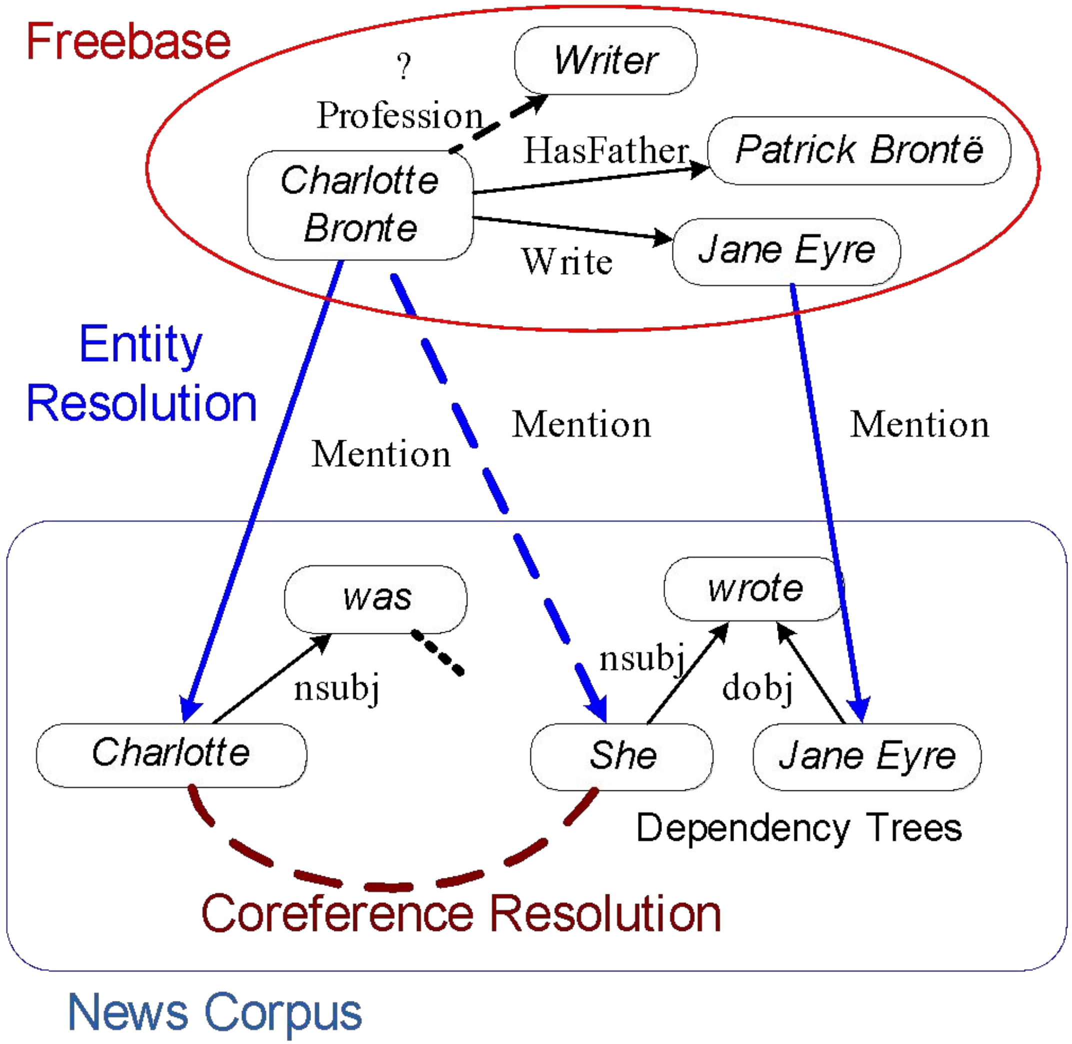


Grammar

painters	[/painting] !/art_forms
impressionist /visual_artist	[/associated_periods_or_movements = /impressionism]
/person during the 1920s	[/date_of_work < 1930; /date_of_work > 1920]

Relation extraction with KB

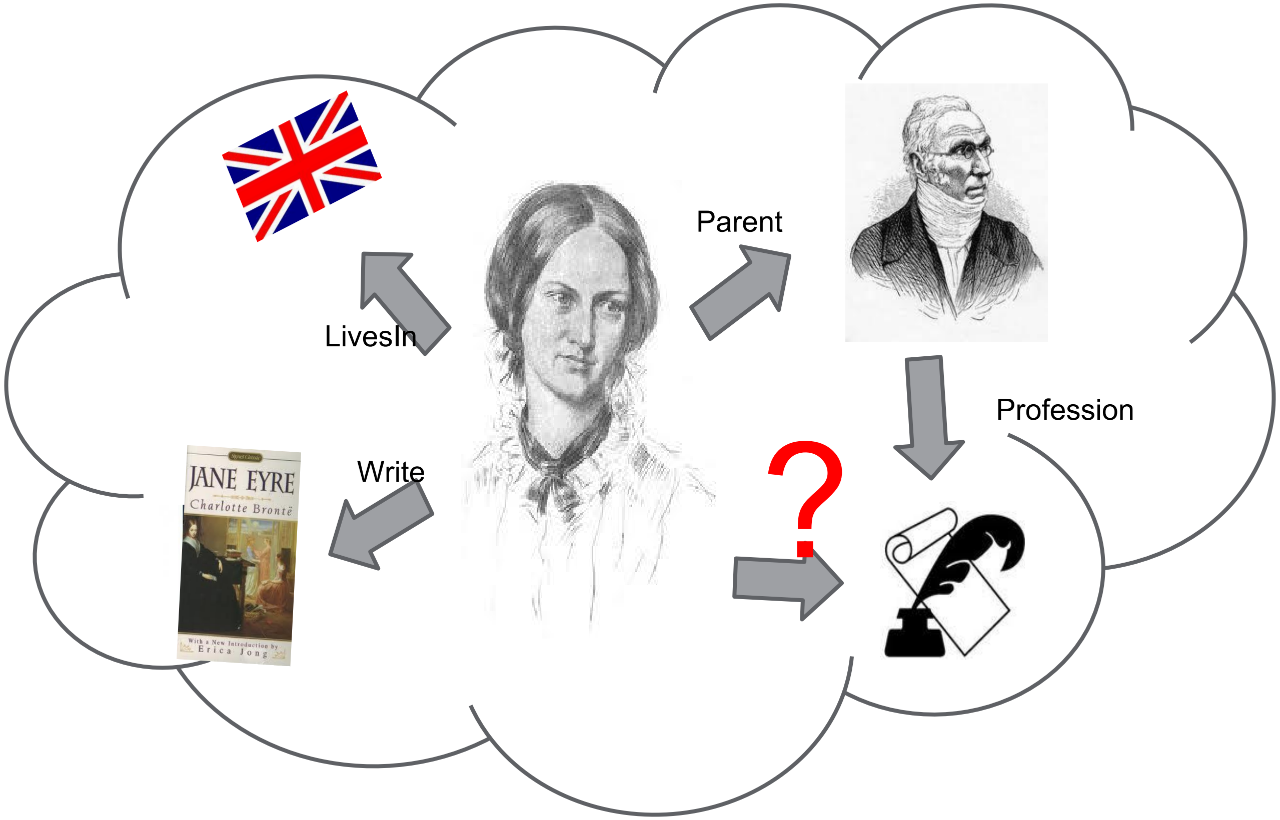
(21M concepts, 70M edges)



Outline

- KB and AI
  - Symbolism
  - Where does knowledge come from
- KB in action
  - Recommendation [Lao & Cohen 2010]
  - Natural language processing [Lao+ 2015][Zheng+ 2013][Nakashole & Mitchell 2015]
  - Question answering [Liang+ 2013+]
- KB Inference
  - KB completion [Lao+ 2011]
  - Path ranking algorithm [Lao& Cohen, 2010]
  - Efficient path finding [Lao& Cohen, 2011]
  - Longer paths, path with constant [Lao+, 2015]
  - First order logic [Wang+, 2015]
- KB construction & Vector space models
  - Relation extraction [Mintz+ 2009][Lao+ 2012][Dong+ 2014]
  - Open domain information extraction [Fader+ 2011][Fader+ 2013]
  - Vector space models [yao+ 2012][Guu+ 2015]
  - The Web as a KB [Pasupat & Liang 2015]
- Current trends in AI research
  - Modeless
  - Add memory
  - Unsupervised
  - Holistic
  - New applications

Knowledge base completion



[Lao & Cohen 2010]
[Lao+ 2011]

Link Prediction

Given

a directed, edge-labeled graph

a source node s 

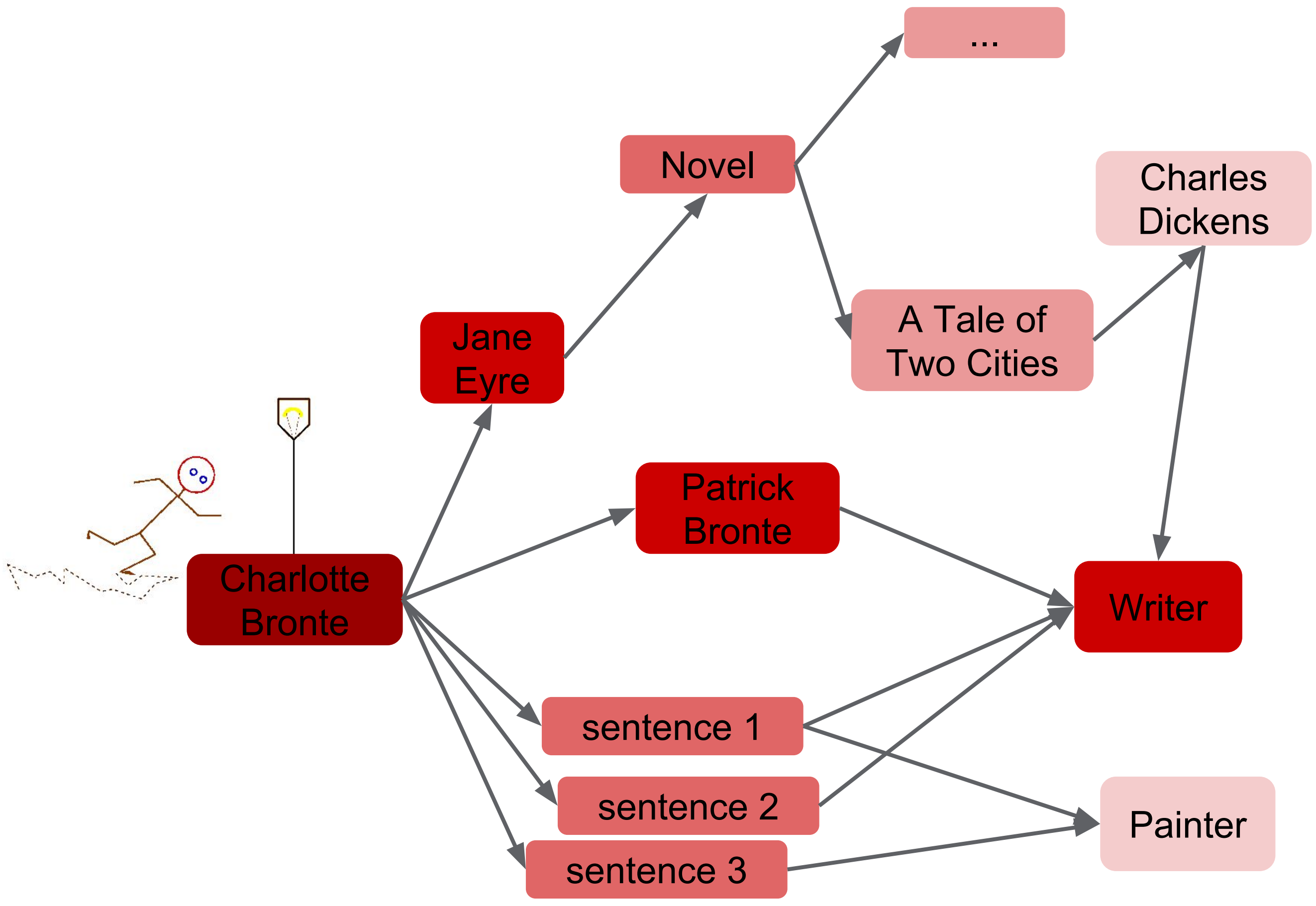
a edge label r Profession

Find

target node t  , s.t. $r(s,t)$

[Lao & Cohen 2010]
[Lao+ 2011]

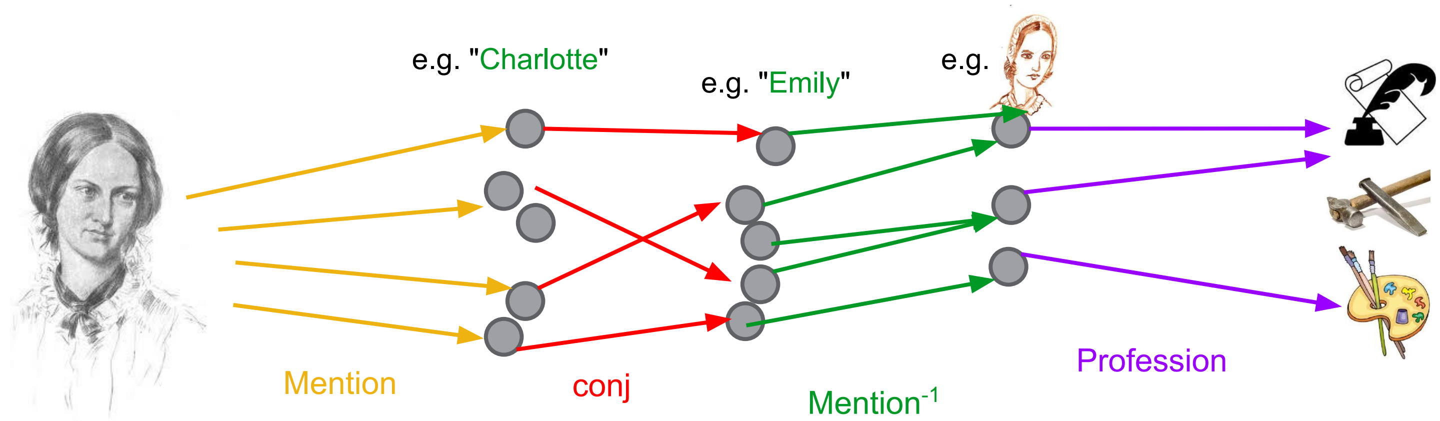
Random Walk with Restart



EMNLP [Lao+ 2012]

Path-Constrained Random Walks

P( →  | <Mention, conj, Mention^{-1},Profession>)



Calculated by dynamic programming or particle filtering

[Lao & Cohen 2010]

Path Ranking Algorithm

$$score(s,t) = \sum_{\pi_j \in \mathcal{P}_\ell} \theta_j P(s \to t; \pi_j),$$

Particle Filtering

A path type, e.g.,
<Father, Profession>

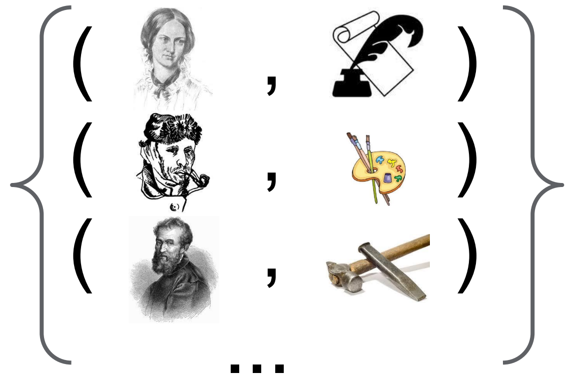
Path Finding

Logistic Regression

[Lao & Cohen 2010]
[Lao+ 2011]

Path Finding

Given a training set D



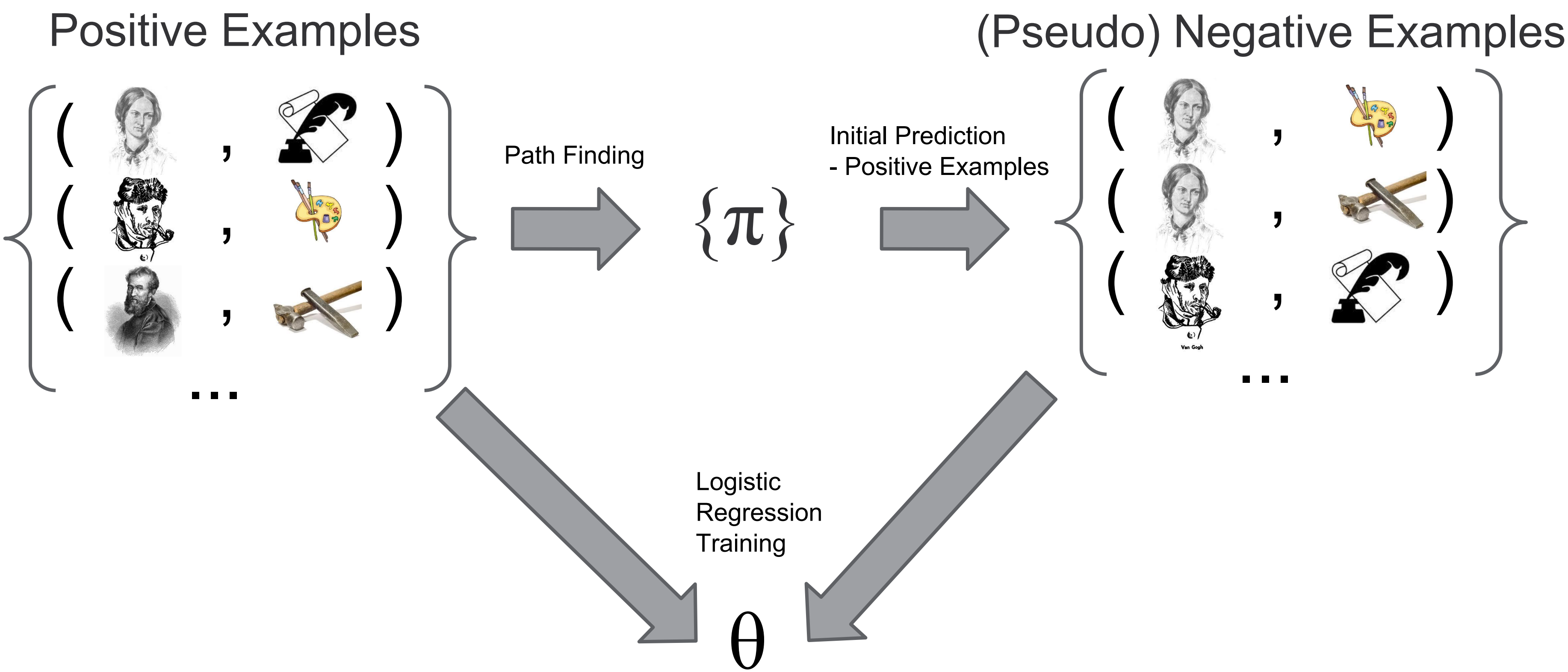
Find the set of path-types $P=\{\pi\}$, s.t. $E_D[P(s \rightarrow t|\pi)] > a$

A hard problem. More on this later...

[Lao & Cohen 2010]
[Lao+ 2011]
[Mintz+ 2009]

Distant Supervision

- Local Closed World Assumption

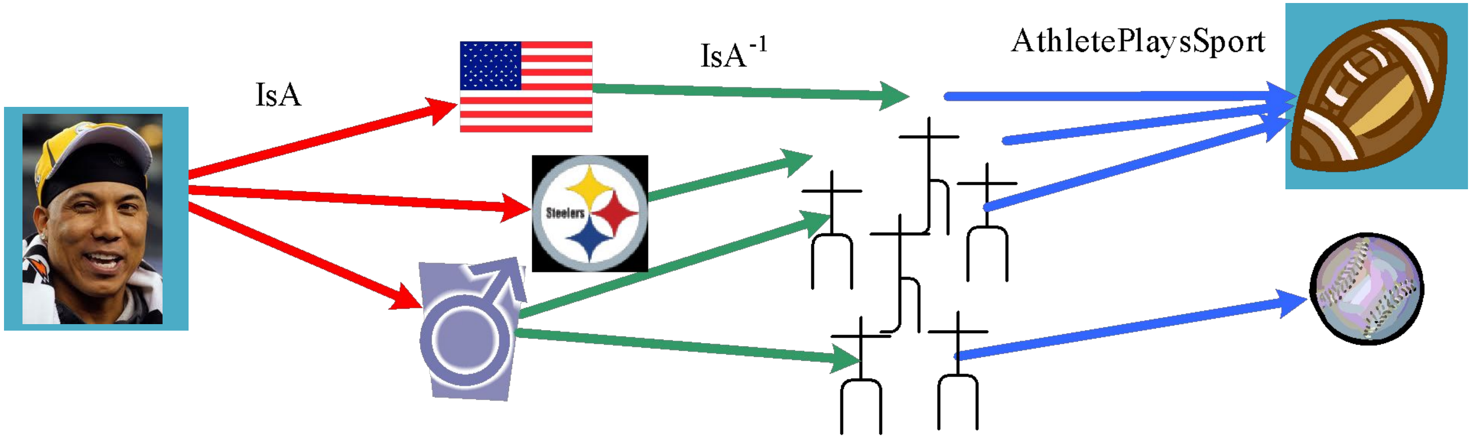


$$score(s,t) = \sum_{\pi \in Q} P(s \to t; \pi)\theta_\pi$$

Efficient Random Walks

- Exact calculation of random walks with dynamic programming results in non-zero probabilities for many internal nodes

AthletePlaysSport(HinesWard, ?)

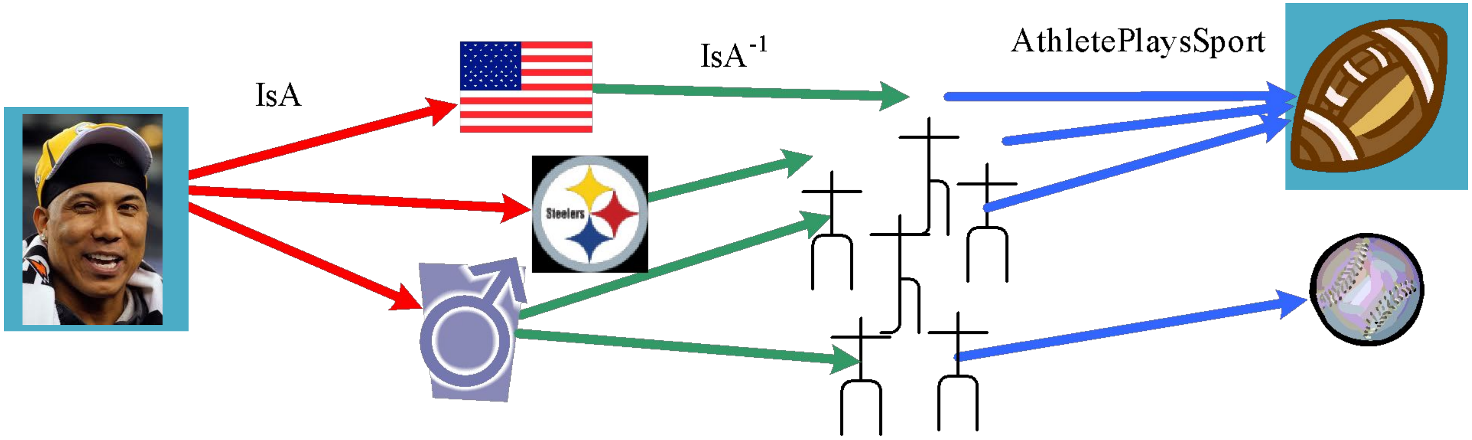


[Lao & Cohen 2010b]

Sampling

- A few sampling are enough to distinguish good target nodes from bad ones

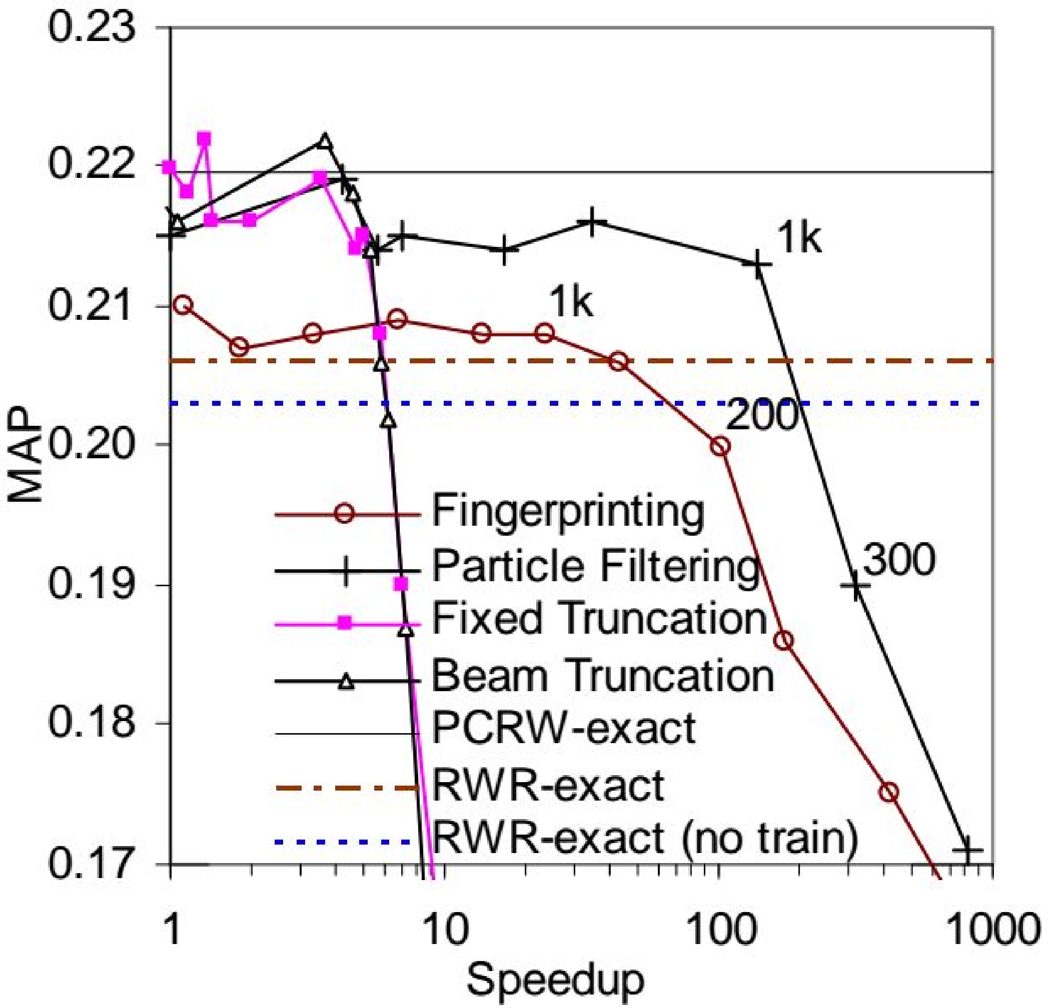
AthletePlaysSport(HinesWard, ?)



[Lao & Cohen 2010b]

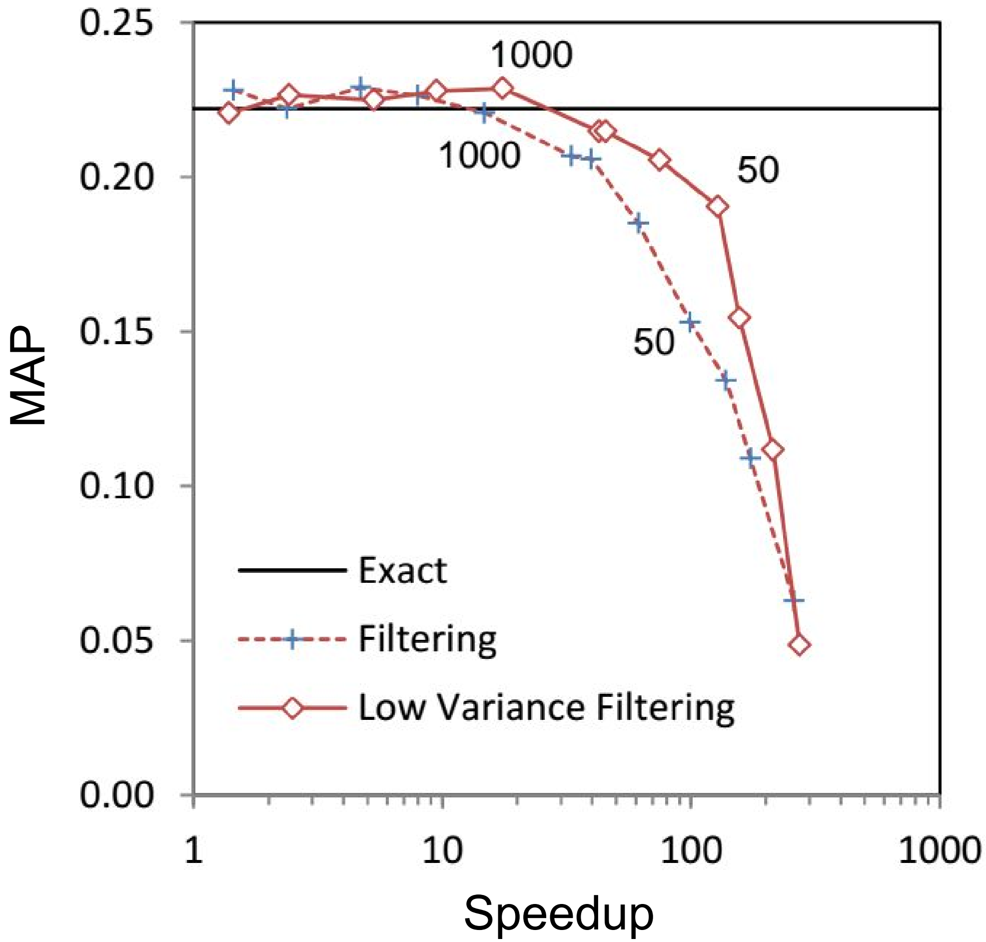
Compare Speedup Approaches

Reading recommendation tasks



Compare Speedup Approaches

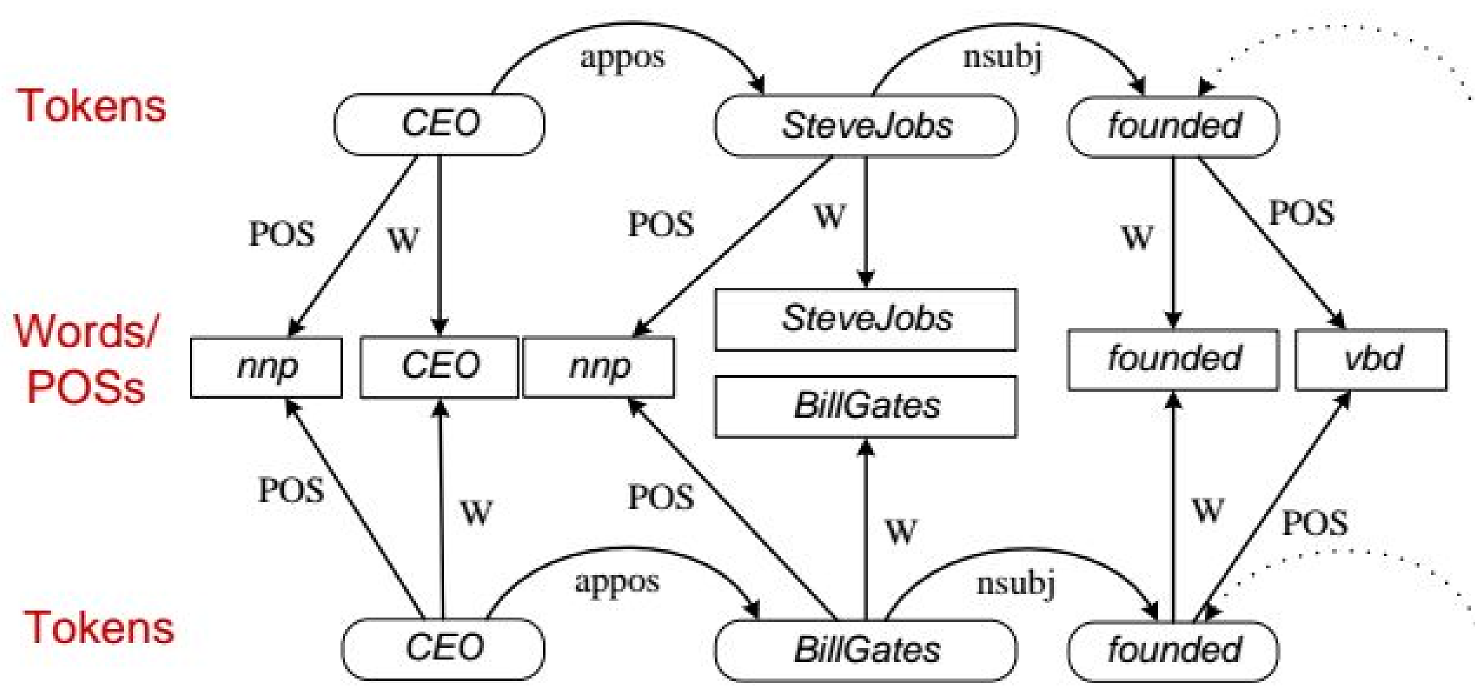
Reading recommendation tasks



[Lao+ 2015]

Long paths are very useful for the entity extraction task

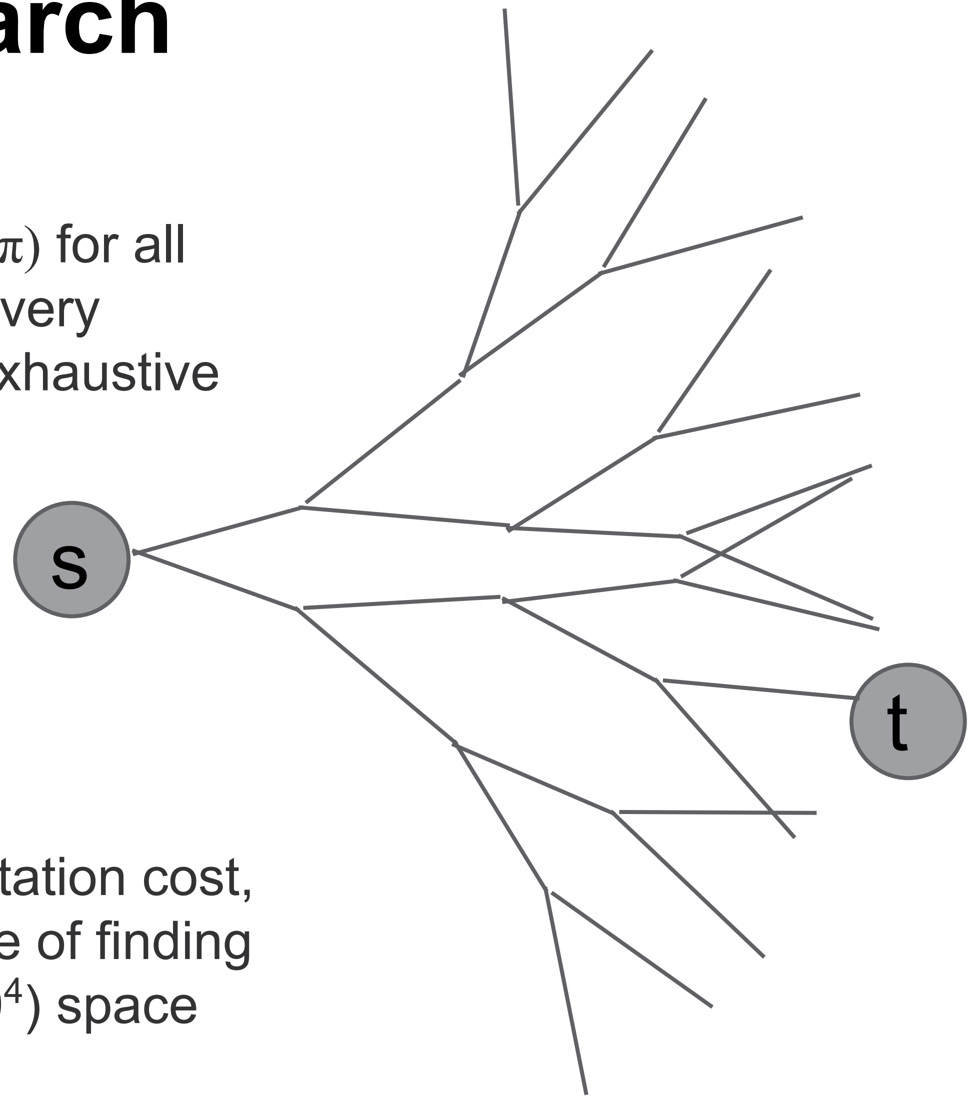
$$P(s \rightarrow t; W^{-1}, conj\_and^{-1}, W, W^{-1}, conj\_and, W)$$
$$P(s \rightarrow t; W^{-1}, nn, W, W^{-1}, appos^{-1}, W)$$
$$P(s \rightarrow t; W^{-1}, appos, W, W^{-1}, appos^{-1}, W)$$



[Lao+ 2015]

Forward Search

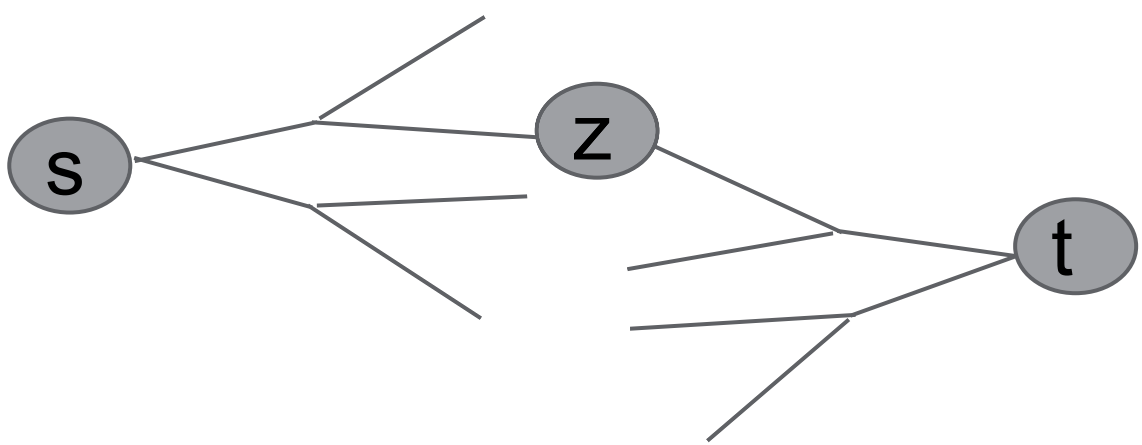
- Calculating $P(s \to t \mid \pi)$ for all possible π is either very expensive or non exhaustive



- With $O(10^2)$ computation cost, we have 1% chance of finding the target in a $O(10^4)$ space

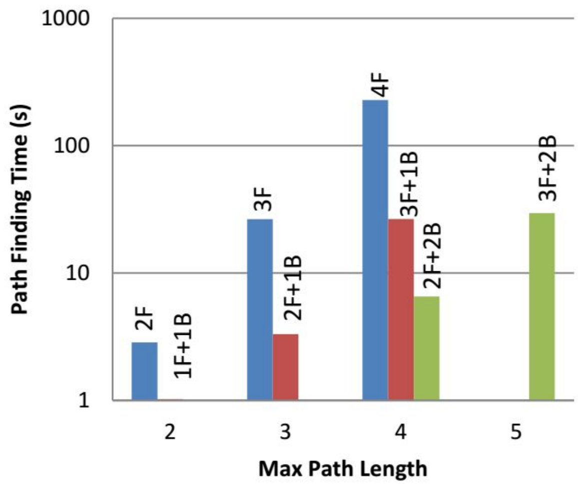
Combine Forward & Backward Random Walks

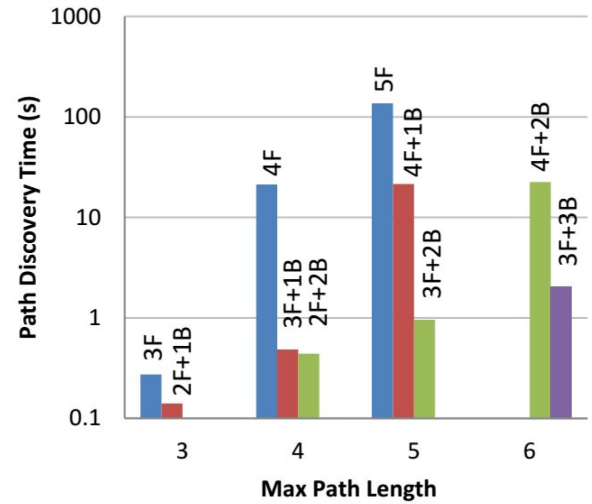
$$P(s \to t; \pi) = \sum_{z} P(s \to z; \pi')P(z \to t; r)$$



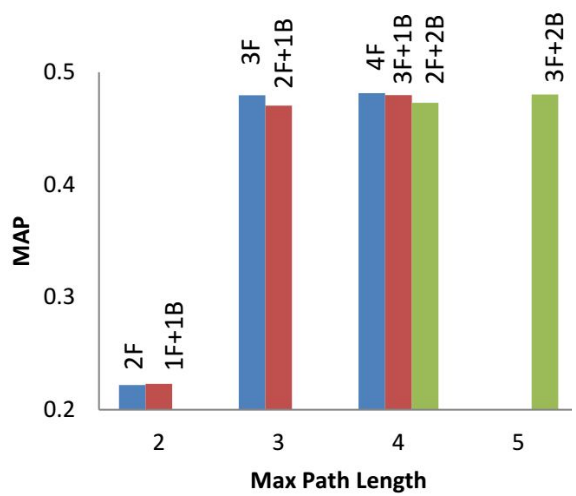
[Lao+ 2015]

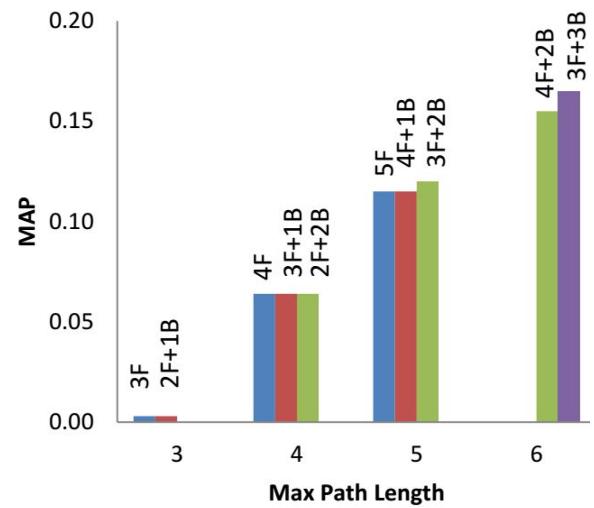
Path Finding Time





Prediction Quality





Knowledge Base Inference

Coordinate Term Extraction

[Lao+ 2015]

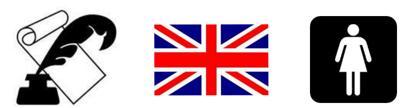
Path with Constants

Path features can be diffusive

MentionedAs

"writer"
"English"
"lady"

Word^{-1}



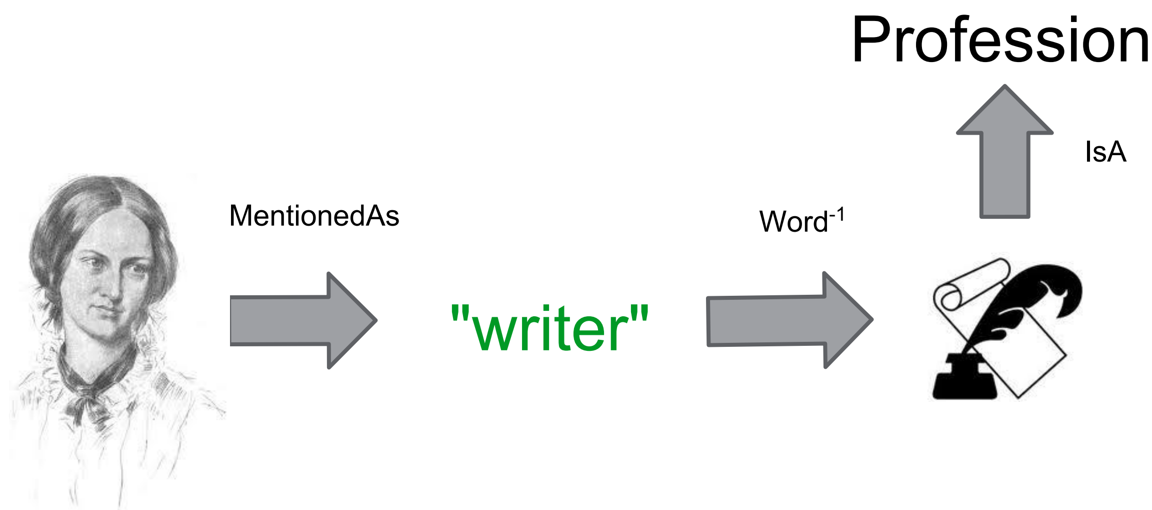
Query independent paths provide
useful prior about correct answers

HasInstance

Profession

Conjunction of Paths

P( → Profession | IsA) × P( →  | MentionedAs, Word^{-1})



- Conjunctions can improve accuracy

[Lao+ 2015]

Constant Paths for KB Completion

Constant path	Interpretation
r=athletePlaysInLeague	
$P(mlb \to t; \phi)$	Bias toward *MLB*.
$P(boston\_braves \to t; \langle athletePlaysForTeam^{-1}, athletePlaysInLeague\rangle)$	The leagues played by *Boston Braves* university team members.
r=competesWith	
$P(google \to t; \phi)$	Bias toward *Google*.
$P(google \to t; \langle competesWith, competesWith\rangle)$	Companies which compete with Google's competitors.
r=teamPlaysInLeague	
$P(ncaa \to t; \phi)$	Bias toward *NCAA*.
$P(boise\_state \to t; \langle teamPlaysInLeague\rangle)$	The leagues played by *Boise State* university teams.

[Lao+ 2015]

Constant Paths for Entity Extraction

Constant path	Interpretation
$P(said \leftarrow t; W^{-1}, nsubj, W)$	The subjects of 'said' or 'say'
$P(says \leftarrow t; W^{-1}, nsubj, W)$	are likely to be a person name.
$P(vbg \leftarrow t; POS^{-1}, nsubj, W)$	Subjects, proper nouns, and
$P(nnp \leftarrow t; POS^{-1}, W)$	nouns with apposition or
$P(nn \leftarrow t; POS^{-1}, appos^{-1}, W)$	possessive constructions, are
$P(nn \leftarrow t; POS^{-1}, poss, W)$	likely to be person names.

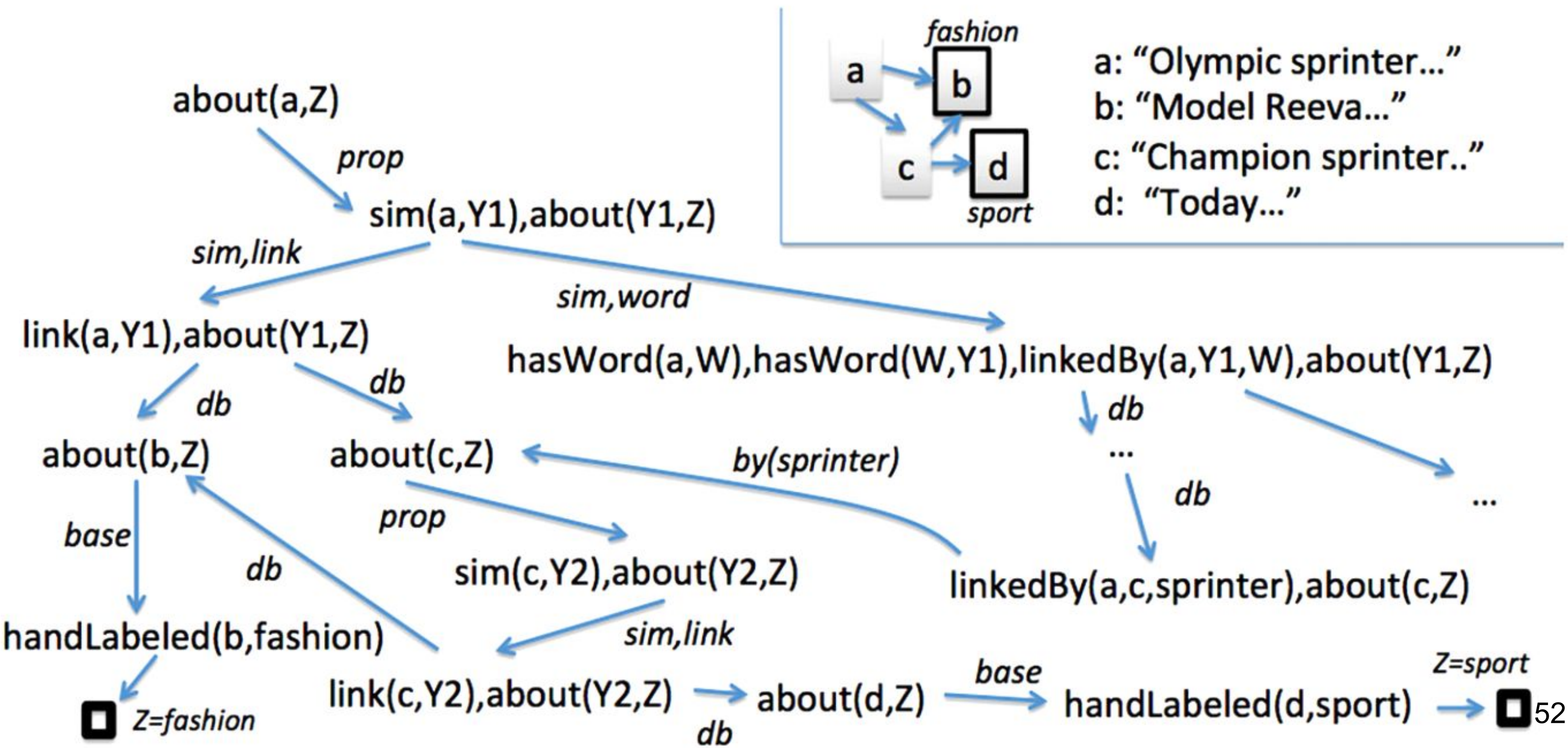
Evaluation for Constant Paths

	KB inference		NE extraction	
	Time	MAP	Time	MAP
RWR	25.6	0.429	7,375	0.017
FOIL	18918.1	0.358	366,558	0.167
PRA	10.2	0.477	277	0.107
CoR-PRA-$no\text{-}const$	16.7	0.479	449	0.167
CoR-PRA-$const_2$	23.3	0.524	556	0.186
CoR-PRA-$const_3$	27.1	**0.530**	643	**0.316**

First Order Logic Inference

[Wang+ 2015]

- SLD resolution by random walk with restart
  - SLD stands for Selective Linear Definite clause resolution
- Transition preferences trained from labeled samples

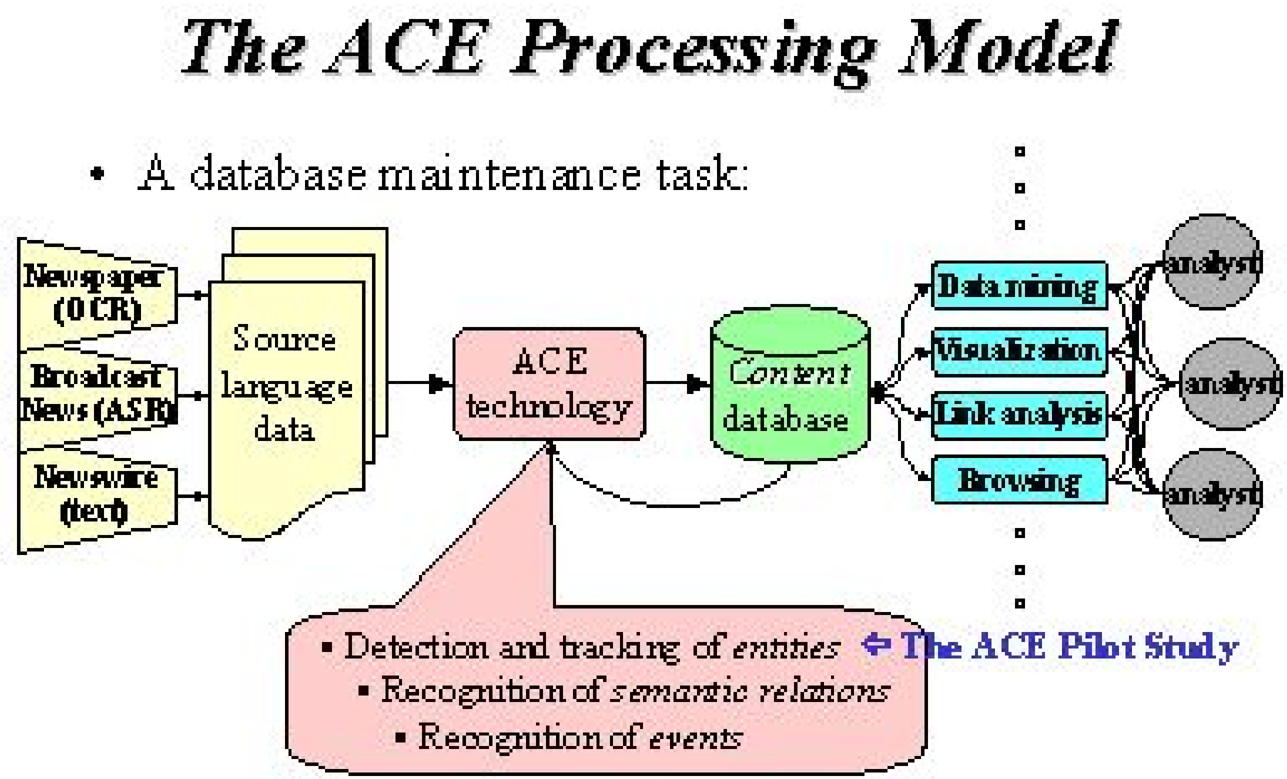


Outline

- KB and AI
  - Symbolism
  - Where does knowledge come from
- KB in action
  - Recommendation [Lao & Cohen 2010]
  - Natural language processing [Lao+ 2015][Zheng+ 2013][Nakashole & Mitchell 2015]
  - Question answering [Liang+ 2013+]
- KB Inference
  - KB completion [Lao+ 2011]
  - Path ranking algorithm [Lao& Cohen, 2010]
  - Efficient path finding [Lao& Cohen, 2011]
  - Longer paths, path with constant [Lao+, 2015]
  - First order logic [Wang+, 2015]
- KB construction & Vector space models
  - Relation extraction [Mintz+ 2009][Lao+ 2012][Dong+ 2014]
  - Open domain information extraction [Fader+ 2011][Fader+ 2013]
  - Vector space models [yao+ 2012][Guu+ 2015]
  - The Web as a KB [Pasupat & Liang 2015]
- Current trends in AI research
  - Modeless
  - Add memory
  - Unsupervised
  - Holistic
  - New applications

Information Extraction

- Has its root in DARPA
  - An intelligent agent monitoring a news data feed requires IE to transform unstructured data into something that can be reasoned with, e.g., (PERSON, works_for, ORGANIZATION)



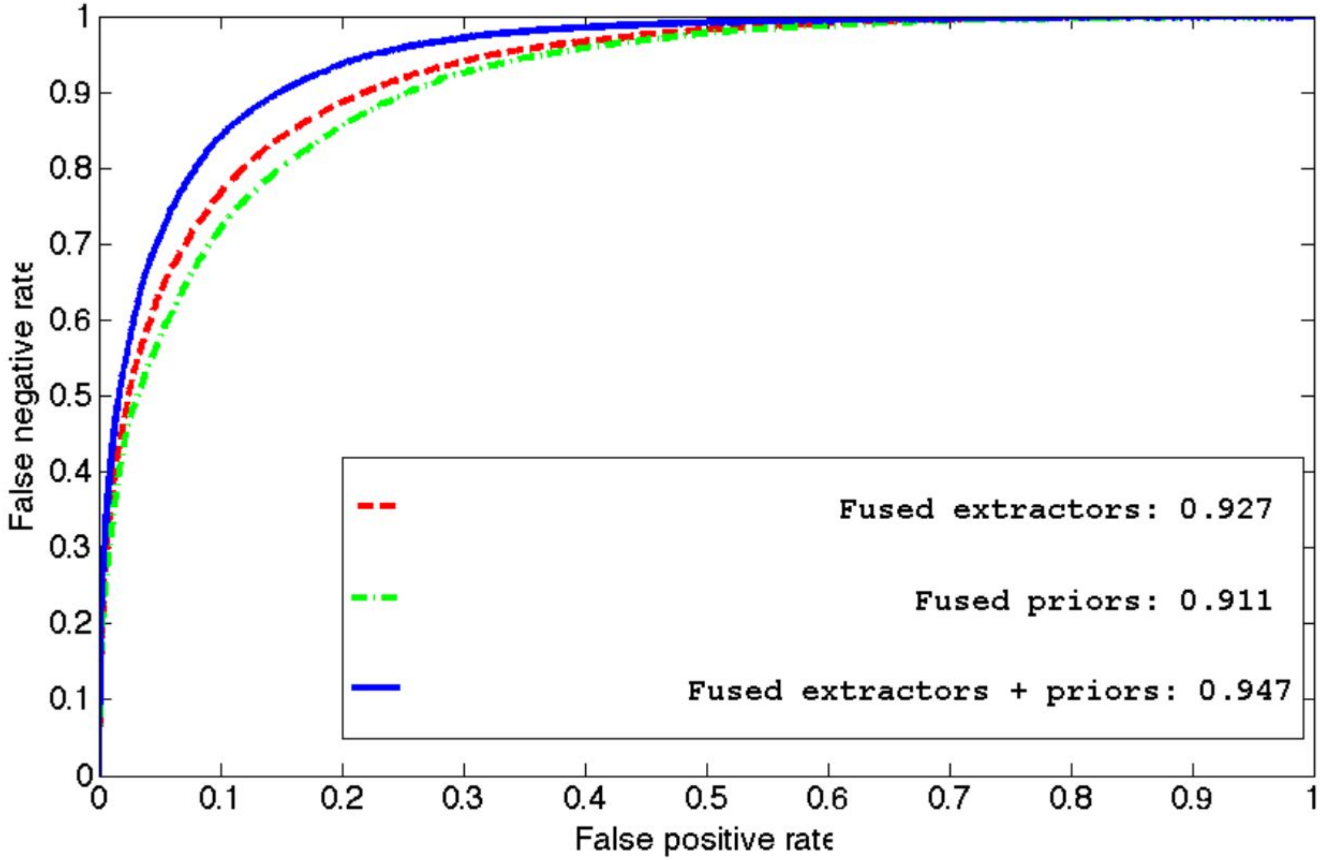
[Mintz+ 2009]

Information Extraction

- The result technologies can only be applied to very restricted domains
  - Supervised classifiers are limited by labeled data
    - (Zhou et al., 2005; Zhou et al., 2007; Sur-deanu and Ciaramita, 2007)
  - Unsupervised approaches can extract very large numbers of triple, but may not be easy to map to relations needed
    - (Shinyama and Sekine, 2006;Banko et al., 2007)
  - Distantly supervised classifiers are still limited by the KB schema
    - (Mints et al., 2009)

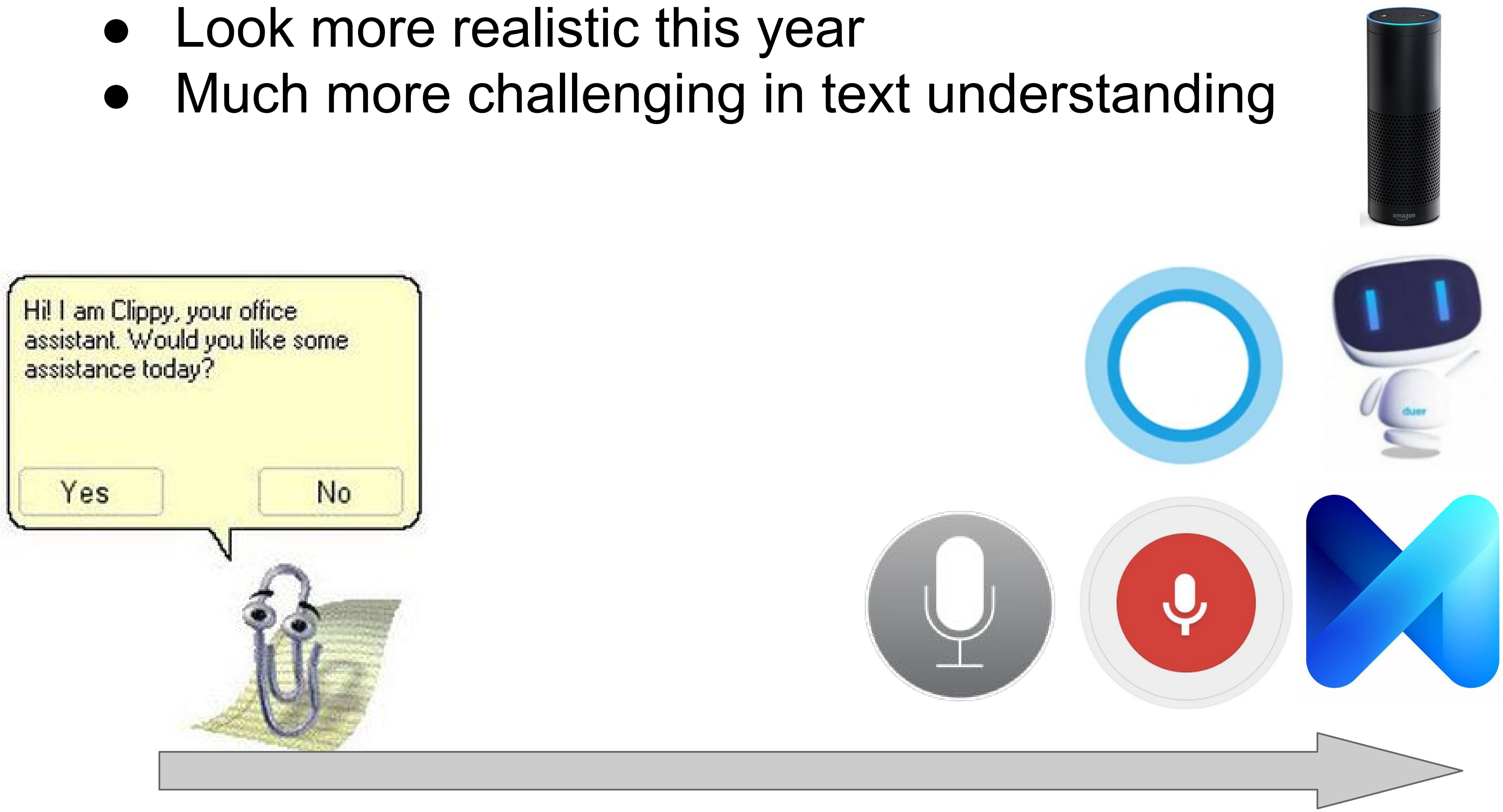
[Dong+ 2014]

Combine KB completion models with relation extractions



Intelligent personal assistants

- Look more realistic this year
- Much more challenging in text understanding



1997

2015

Challenges of KB construction

- Open domain
  - what is the form of worldly KB which enables efficient reasoning?
- Representation
  - symbolism vs connectionism
- The web as a KB
  -
- Unsupervised/semi-supervised training
  - where to get training data?

Open domain extraction

[Fader+ 2011]
[Fader+ 2013]

- Reverb
  - introduced two syntactic and lexical constraints to overcome uninformative and incoherent extractions
  - e.g., “Faust made a deal with the devil.” !==> (Faust, made, a deal)

- WikiAnswers
  - a large, community-authored, question-paraphrase corpus

- A small manually design seed lexicon
  - e.g., "who r e" ==> r(?, e)
  - e.g., big ==> population

- Expand lexicon and tune ranking function by question pairs
  - e.g., "who r e" ==> r(?, e)

still have data sparsity issues

Why Relation Extraction Worked

- In very restricted domains

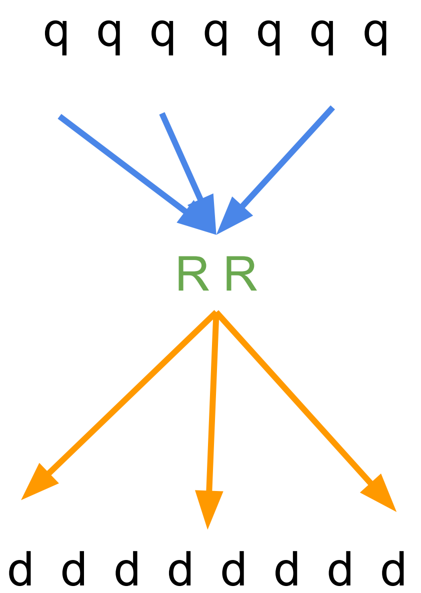
Closed domain queries

Semantic parser

KB relations

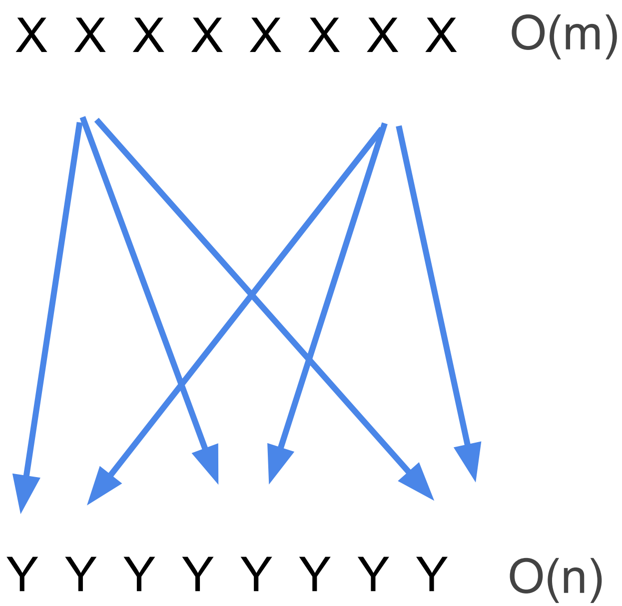
Text or html patterns

Web docs

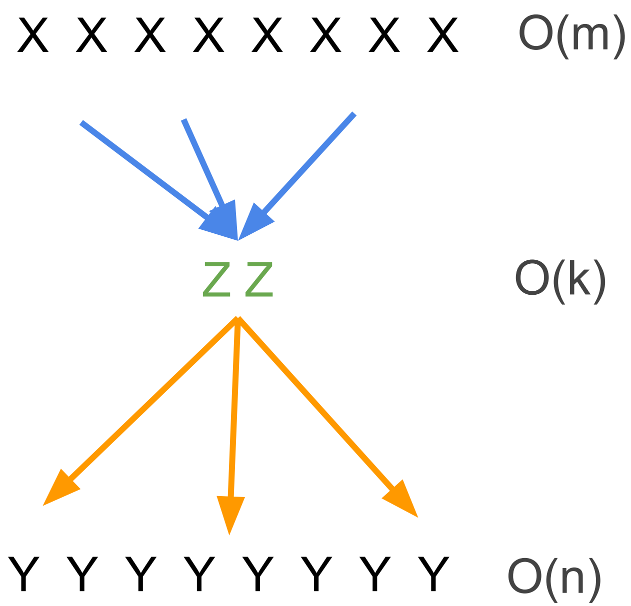


Why Graphical Models Need Hidden Variables

- We want to model the correlations between variables Xs and variables Ys



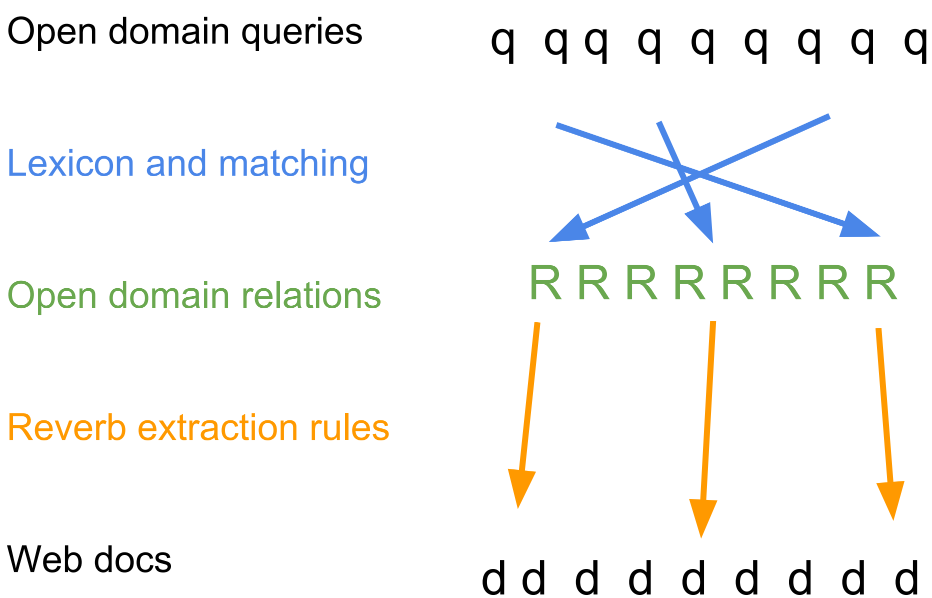
#param: O(m*n)



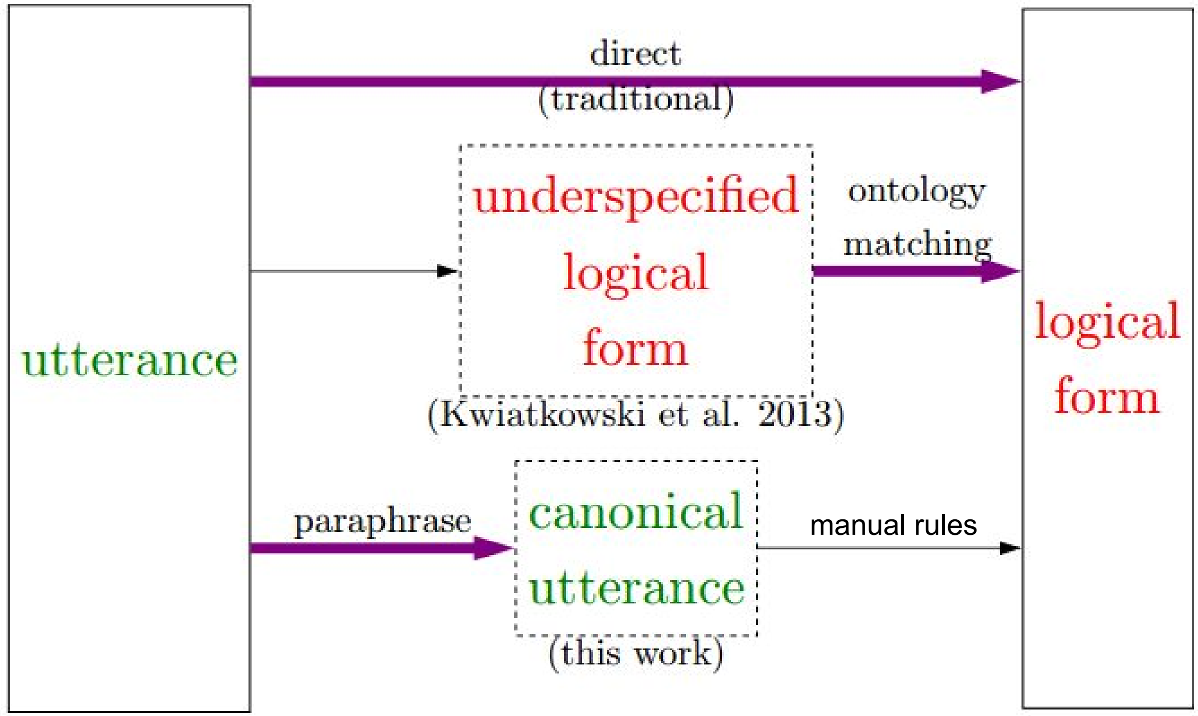
#param: O(m*k + n*k)

Why Open Domain Relation Extraction Is Hard

- Open domain schemas are not compact enough



From open IE to matching problems

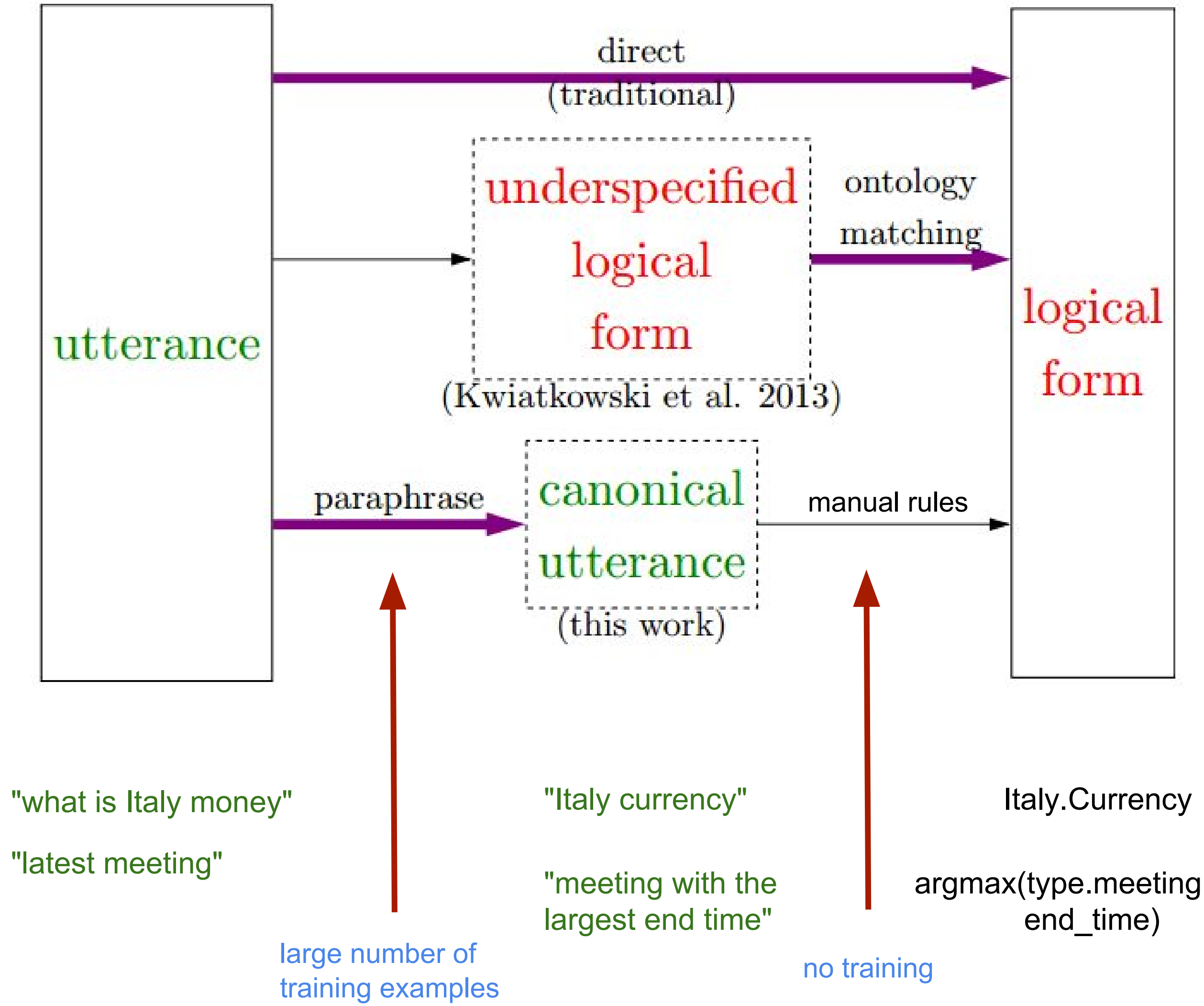


"what is Italy money"	"Italy currency"	Italy.Currency
"latest meeting"	"meeting with the largest end time"	argmax(type.meeting, end_time)

[Berant & Liang 2014]

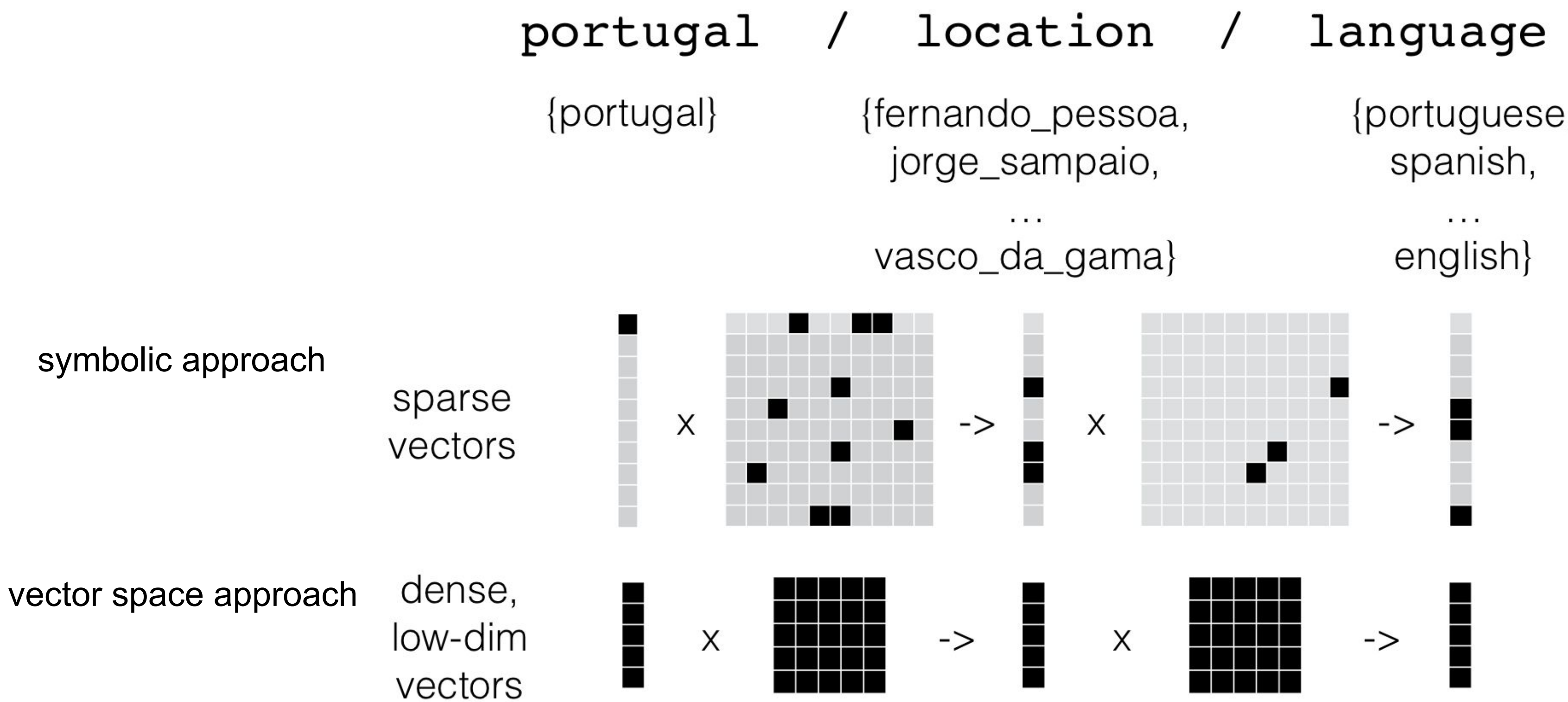
From open IE to matching problems

The beauty of the proposed approach



[Guu, Miller, Liang 2015]

Vector Space Models

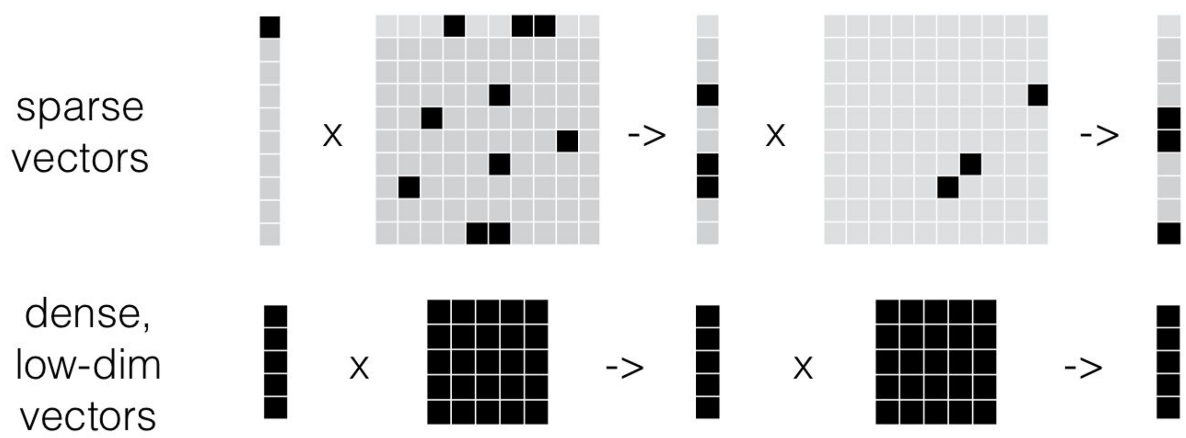


[Guu, Miller, Liang 2015]

Vector Space Models

`portugal / location / language`

{portugal} {fernando_pessoa, jorge_sampaio, ... vasco_da_gama} {portuguese, spanish, ... english}

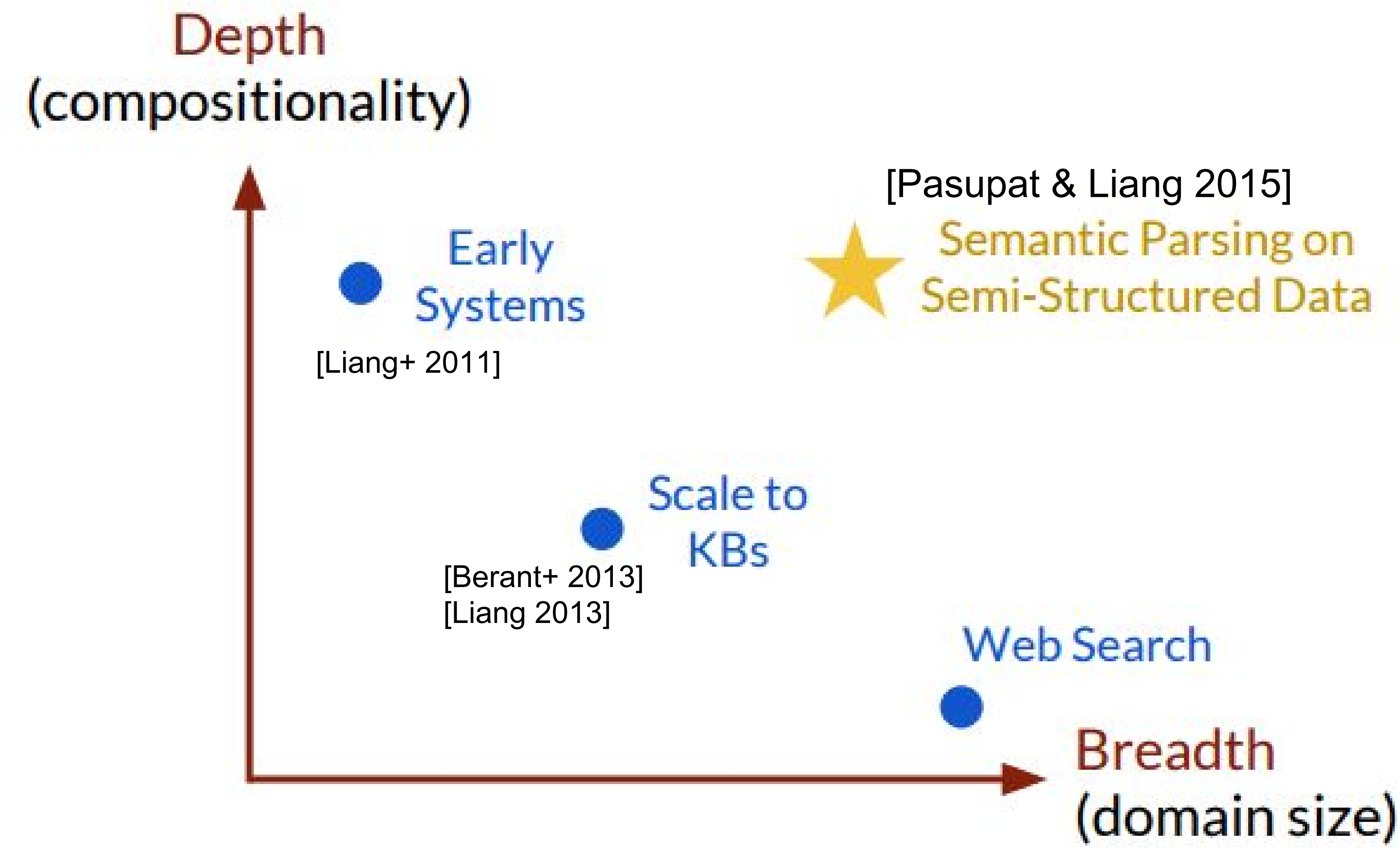


- propositionalization is done too early
  - most graph information cannot be crammed into node vectors and is lost
  - "which college did barack obama's second child's english teach graduate?"

heated discussion about what should be in the "Ray Mooney vector" @the vector space model workshop, ACL 2015

[Pasupat & Liang 2015]

The Web as a KB



[Pasupat & Liang 2015]

The Web as a KB

Early systems: Parse very compositional questions into database queries

How many rivers are in the state with the largest population?

```
answer(A,
  count(B,
    (river(B), loc(B, C),
     largest(D, (state(C), population(C, D)))),
    A)))
```

Compositionality: High

Knowledge source: Database

- few entities / relations
- fixed schema

[Zelle + Mooney, 1996 / Wong + Mooney, 2007 / Zettlemoyer + Collins, 2007 / Kwiatkowski et al., 2011 / ...]

The Web as a KB

Scaling to large knowledge bases (KBs): Answer open-domain questions using curated KBs

In which comic book issue did Kitty Pryde first appear?

R[FirstAppearance].KittyPryde

Compositionality: Lower

Knowledge source: Large KBs

- lots of entities / relations
- fixed schema

[Cai + Yates, 2013 / Berant et al., 2013 + 2014 / Fader et al., 2014 / Reddy et al., 2014 / ...]

The Web as a KB

QA on semi-structured data

Input: utterance x and HTML table t

Output: answer y

Training data: list of (x, t, y) — no logical form

WikiTableQuestions dataset:

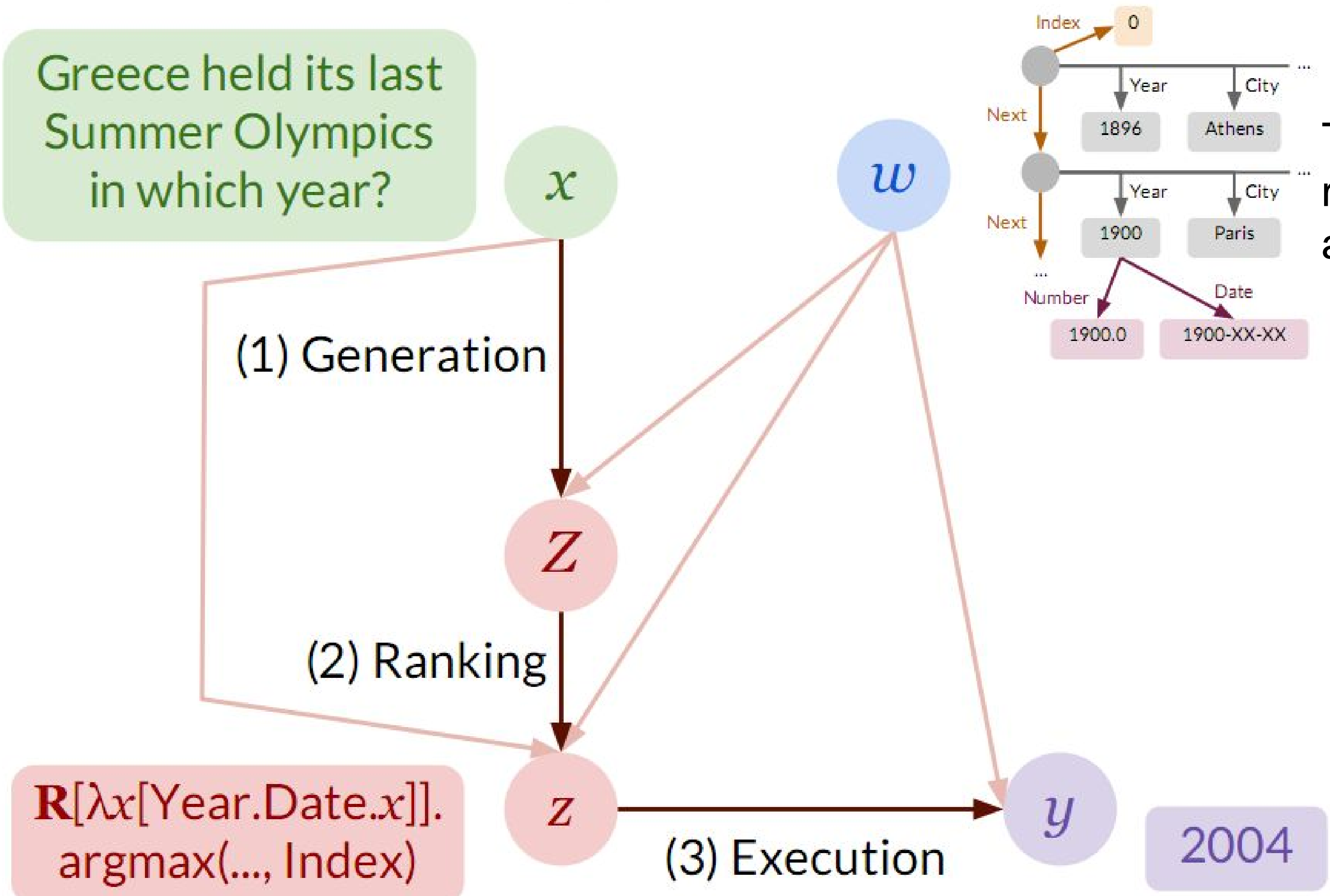
- Tables t are from Wikipedia
- Questions x and answers y are from Mechanical Turk

— Prompts are given to encourage compositionality

e.g. Prompt: The question must contains "last" (or a synonym)

In what city did Piotr's last 1st place finish occur?

The Web as a KB



Tables are represented as graphs

Outline

- KB and AI
  - Symbolism
  - Where does knowledge come from
- KB in action
  - Recommendation [Lao & Cohen 2010]
  - Natural language processing [Lao+ 2015][Zheng+ 2013][Nakashole & Mitchell 2015]
  - Question answering [Liang+ 2013+]
- KB Inference
  - KB completion [Lao+ 2011]
  - Path ranking algorithm [Lao& Cohen, 2010]
  - Efficient path finding [Lao& Cohen, 2011]
  - Longer paths, path with constant [Lao+, 2015]
  - First order logic [Wang+, 2015]
- KB construction & Vector space models
  - Relation extraction [Mintz+ 2009][Lao+ 2012][Dong+ 2014]
  - Open domain information extraction [Fader+ 2011][Fader+ 2013]
  - Vector space models [yao+ 2012][Guu+ 2015]
  - The Web as a KB [Pasupat & Liang 2015]
- Current trends in AI research
  - Modeless
  - Add memory
  - Unsupervised
  - Holistic
  - New applications

Current trends in AI

- Modeless
  - connectionism vs symbolism
- Add memory
  -
- Unsupervised
  -
- Holistic
  - Blind Men and An Elephant
- New applications
  - OpenAI, Atomic energy

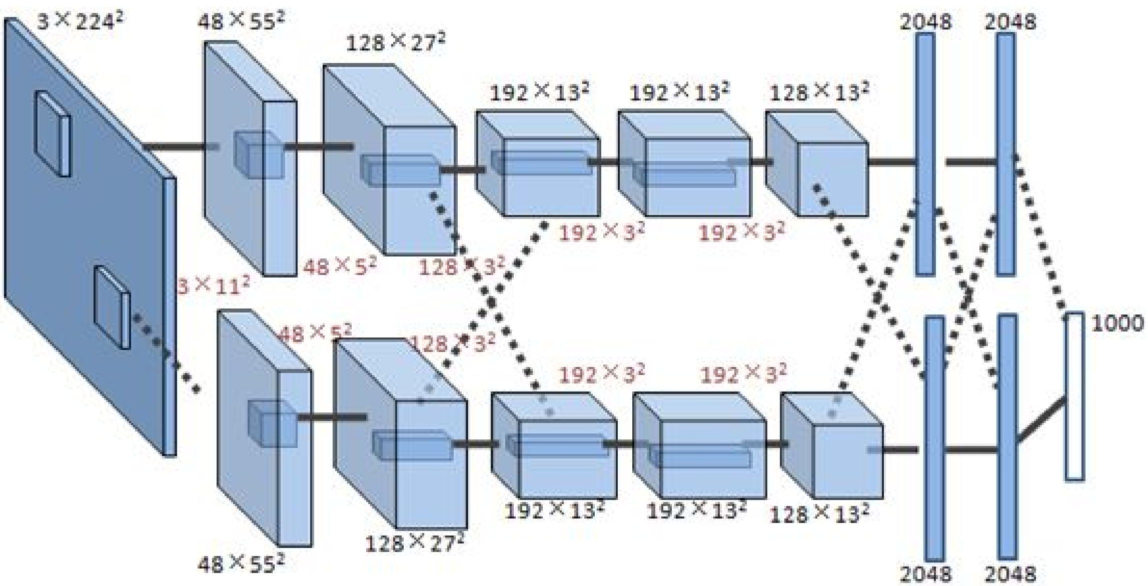
The connectionism is coming back

"NLP is kind of like a rabbit in the headlights of the deep learning machine, waiting to be flattened"

-- Neil Lawrence @ICML2015

Image recognition

- Deep NN beats predominant approaches by large margin
- The key is DNN's ability of feature engineering

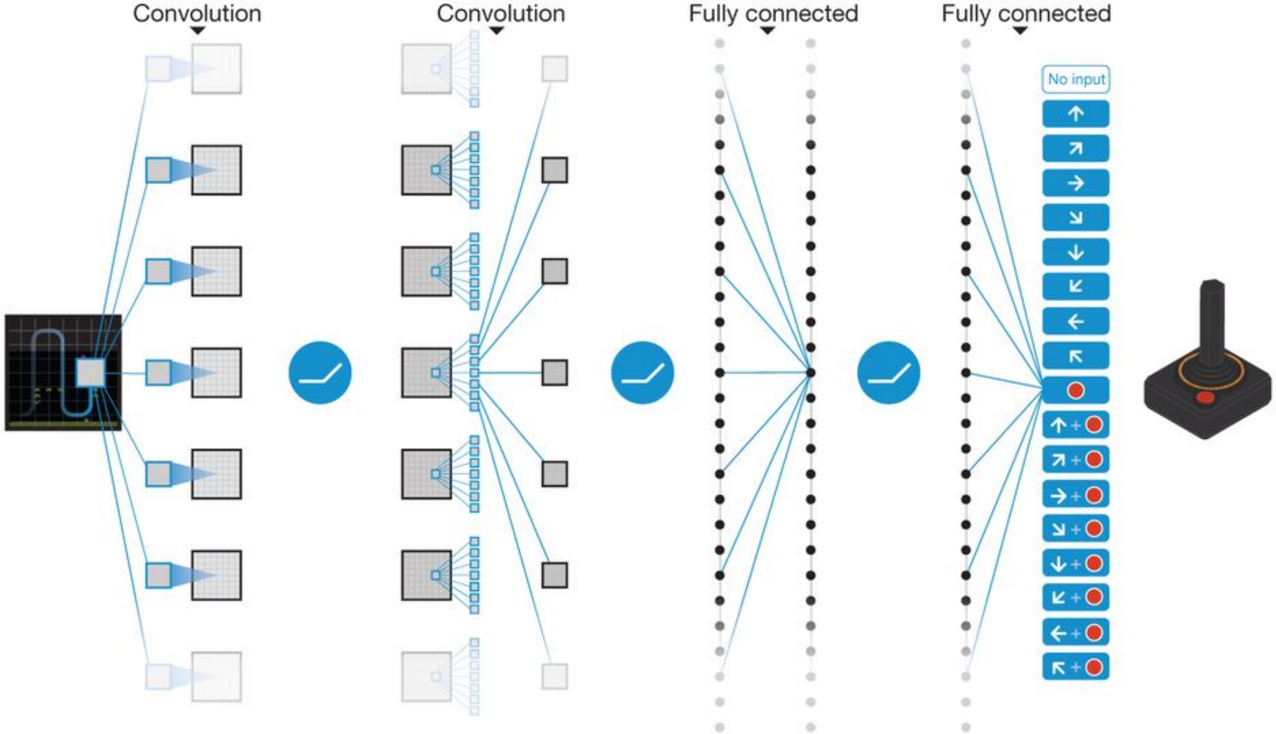


Model	Top-1	Top-5
Sparse coding [2]	*47.1%*	*28.2%*
SIFT + FVs [24]	*45.7%*	*25.7%*
CNN	**37.5%**	**17.0%**

Table 1: Comparison of results on ILSVRC-2010 test set. In *italics* are best results achieved by others.

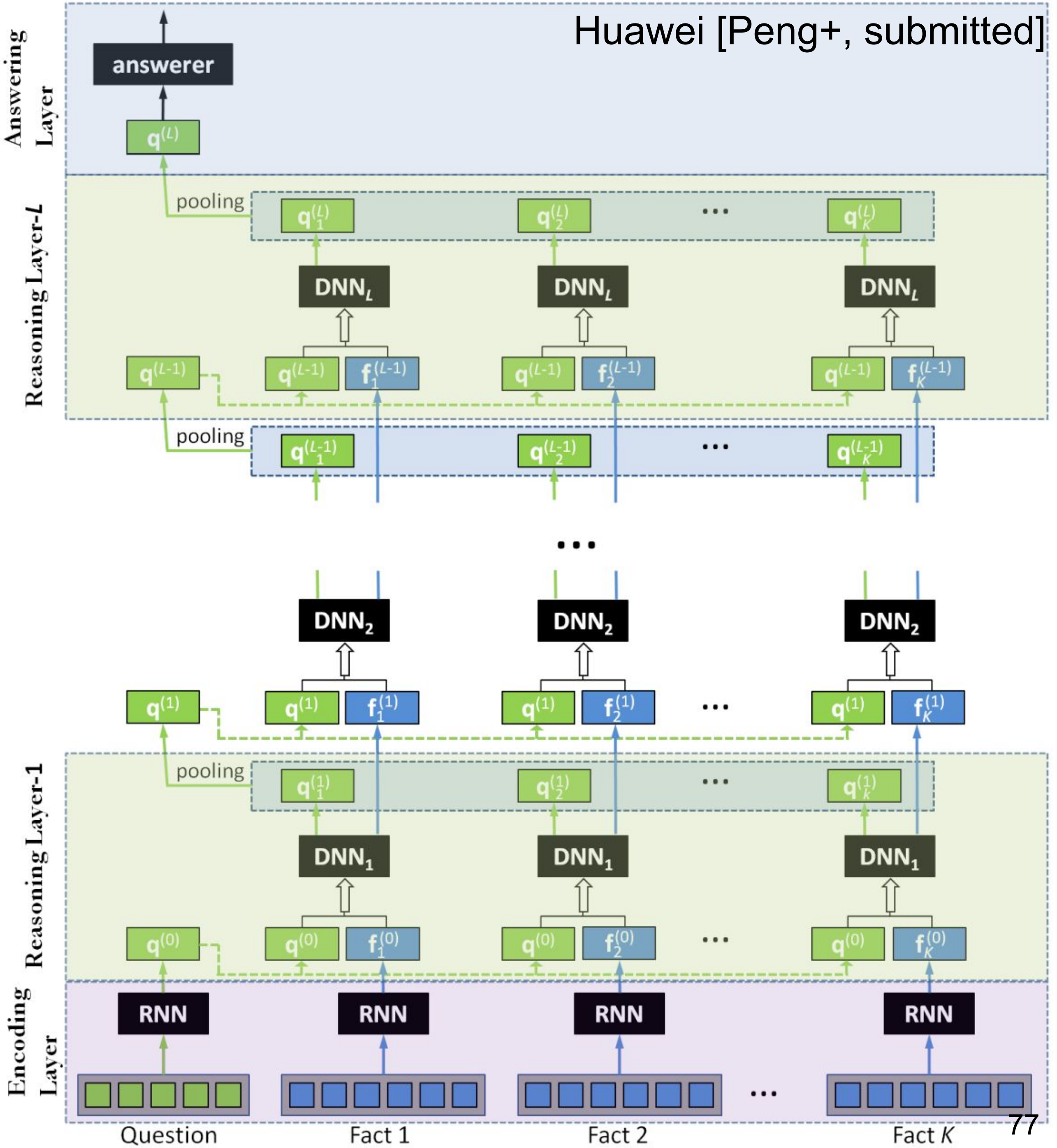
Reinforcement learning

- Once dominating symbolic approaches (MDP, POMD) have been abandoned for deep NN



Reasoning

- Shows striking similarity between a neural reasoner and SLD-resolution



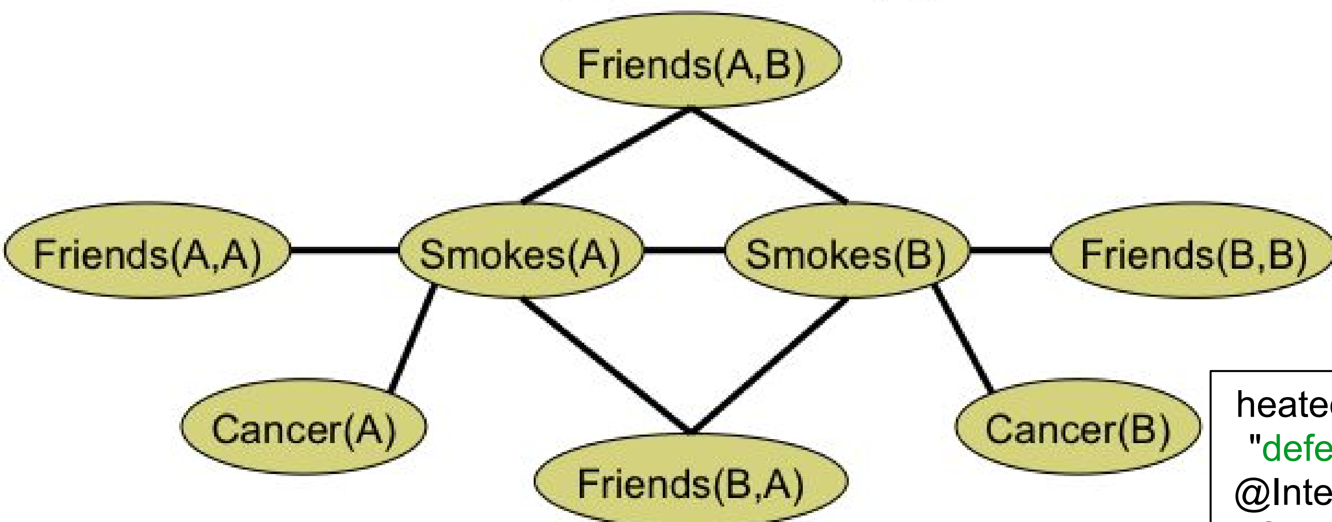
[Richardson & Domingos 2004]

Defending symbolism

- Simple rules can be unrolled into big models
- but these rules need experts to come up with ...

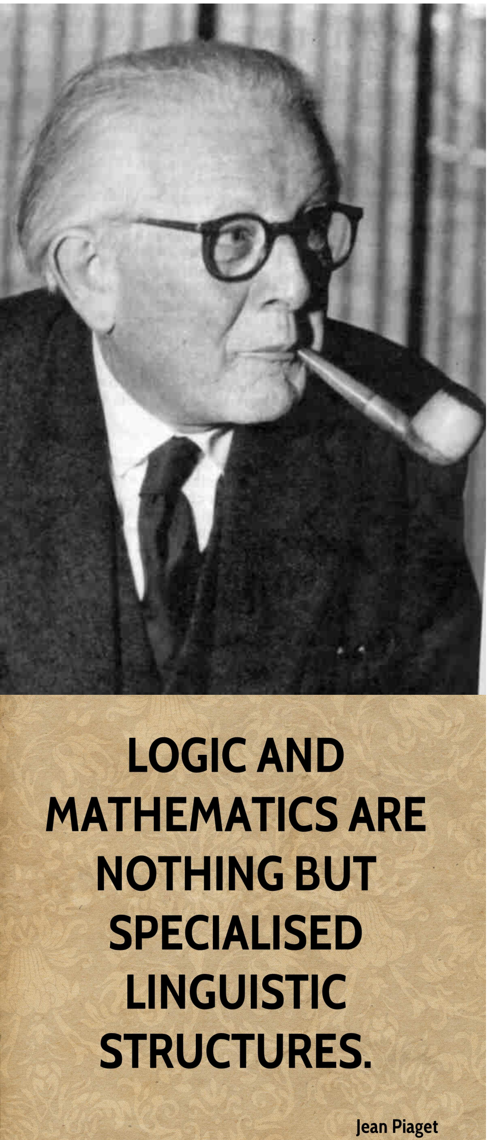
1.5	$\forall x\ Smokes(x) \Rightarrow Cancer(x)$
1.1	$\forall x, y\ Friends(x,y) \Rightarrow \left(Smokes(x) \Leftrightarrow Smokes(y)\right)$

Two constants: **Anna** (A) and **Bob** (B)



heated discussion about "defending symbolism" @Integrating Neural and Symbolic Approaches workshop, NIPS 2015

Connectionism vs. Symbolism

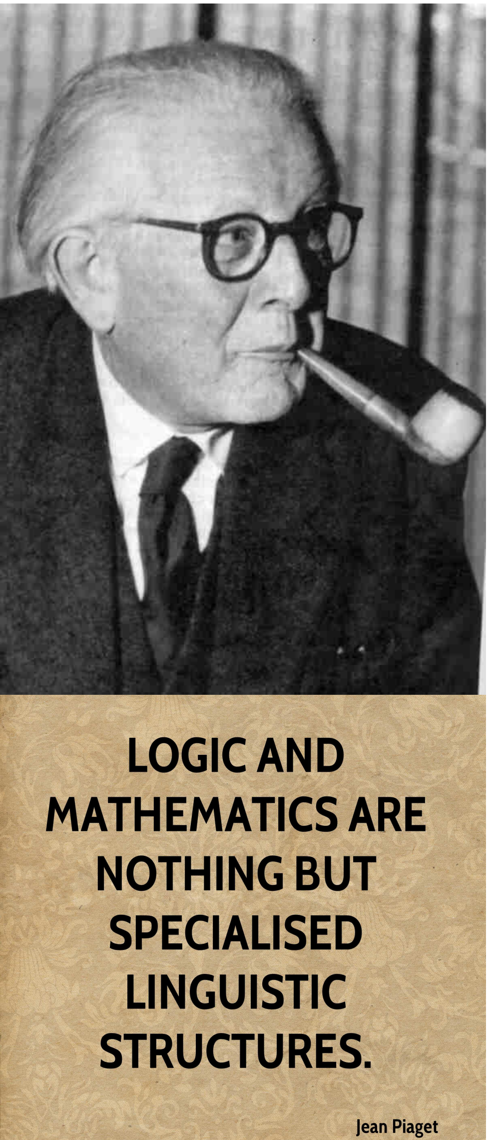


The symbolic models represents elegant solutions to problems, and have been dominating AI for a very long time

VS.

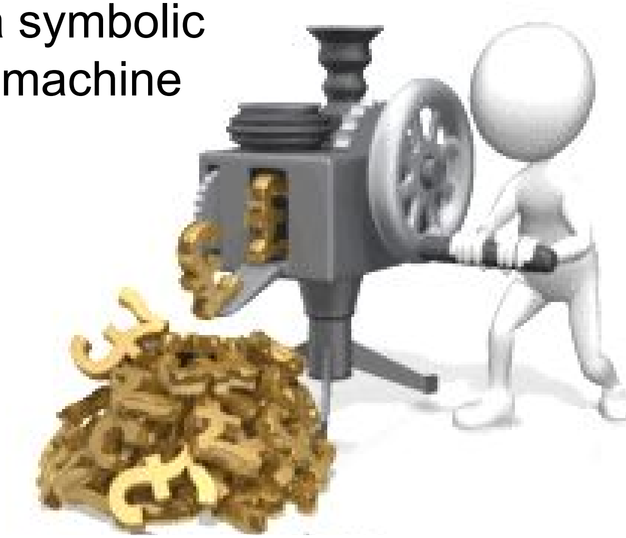
Once we have figured out how to train them (after 30 years), the connectionism approaches starts to win and do not need genius scientists to come up with

Connectionism vs. Symbolism



Since symbolism is very useful for human, it should also be useful to connectionism

a symbolic machine



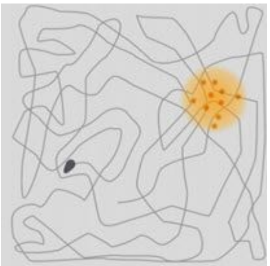
a neural controller

a sequence of symbols

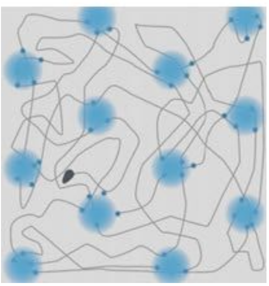
[Stensola+ 2012]

Ghost in the shell

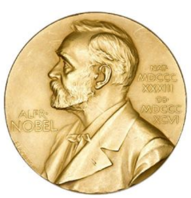
- 2014 Nobel Prize in Physiology or Medicine awarded for ‘inner GPS’ research



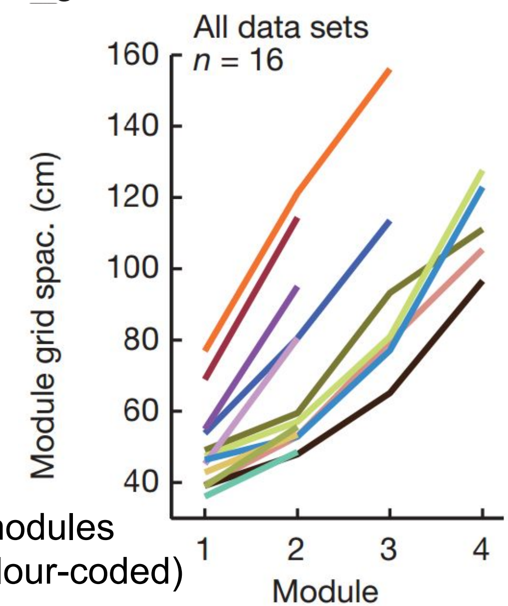
- John O’Keefe discovered, in 1971, that certain nerve cells in the brain were activated when a rat assumed a particular place in the environment. Other nerve cells were activated at other places. He proposed that these “place cells” build up an inner map of the environment. Place cells are located in a part of the brain called the hippocampus.



- May-Britt and Edvard I. Moser discovered in 2005 that other nerve cells in a nearby part of the brain, the entorhinal cortex, were activated when the rat passed certain locations. Together, these locations formed a hexagonal grid, each “grid cell” reacting in a unique spatial pattern. Collectively, these grid cells form a coordinate system that allows for spatial navigation.



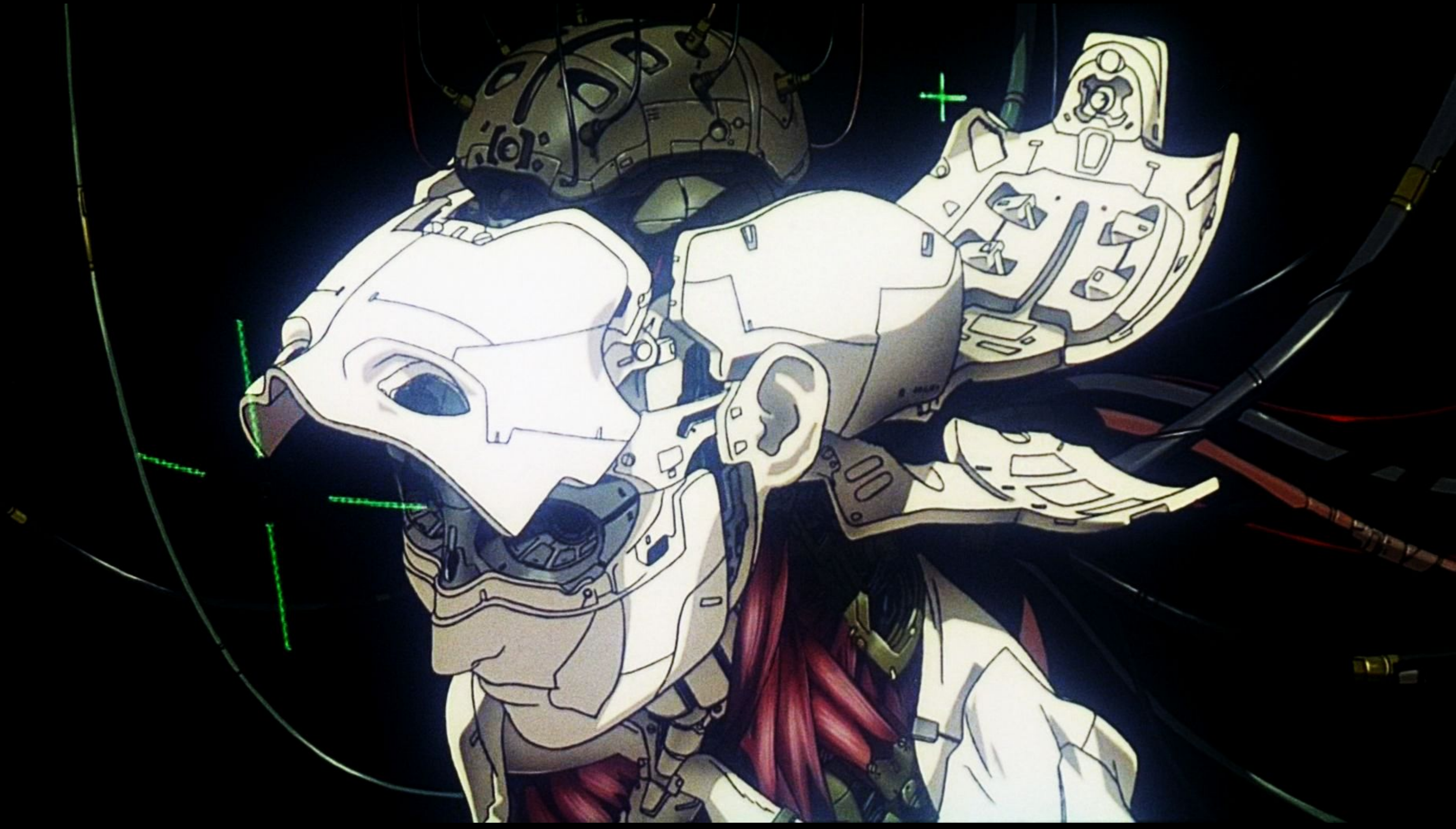
- Locations are represented as discrete numbers in animals' brains, which enable accurate and autonomous calculations



Mean grid spacing for all modules (M1–M4) in all animals (colour-coded)

Ghost in the shell

- Symbolic machines are given to and modified by the neural controller



Adding Memories

- We want our models to store a lot of knowledge while having very few parameters
    - cool models, but no impressive application yet

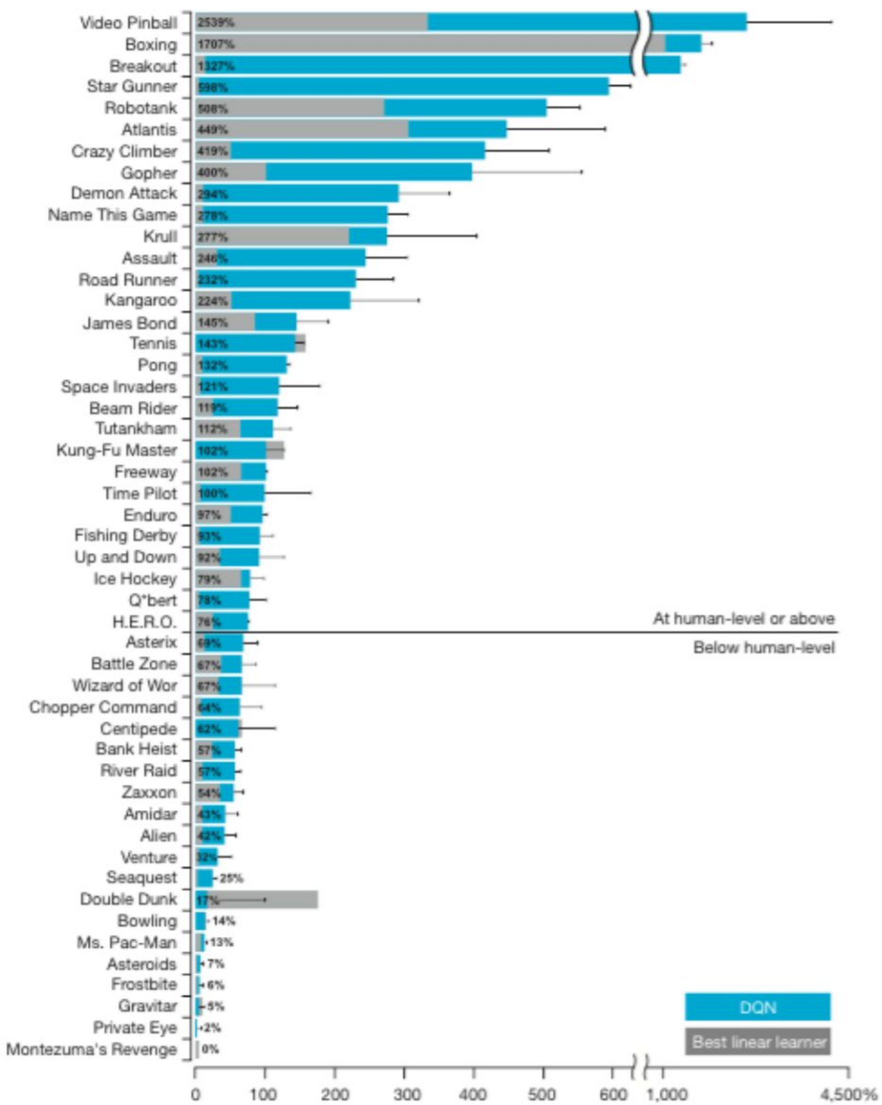
- Uniform model structures
    - DNN, RNN

- Separate memory modules
    - Turing machine [Graves+ 2014]
    - Memory network [Sukhbaatar+ 2015]
    - Dynamic memory network [Kumar+ 2015]
    - Queue & stack [Grefenstette+ 2015]

- Separate memory & computation modules
    - Random-access machines [Kurach+ 2015]

Unsupervised Training

[Silver, 2015]



- Good at:
  - quick-moving, complex, short-horizon games
  - Semi-independent trails within the game
  - Negative feedback on failure
  - Pinball
- Bad at:
  - long-horizon games that don't converge
  - Ms. Pac-Man
  - Any "walking around" game
- Needs intrinsic reward

Holistic AI approach

- I used to fancy an image of “The old Man and The Sea” with
  - a giant sail fish that represents the holly algorithm of “intelligence”
- This image has gradually faded away after years of graduate study

Holistic AI approach

- Intelligence is everywhere, and everyone can feel some aspect of it
- Each subfield of AI holds certain truth, but not all of it

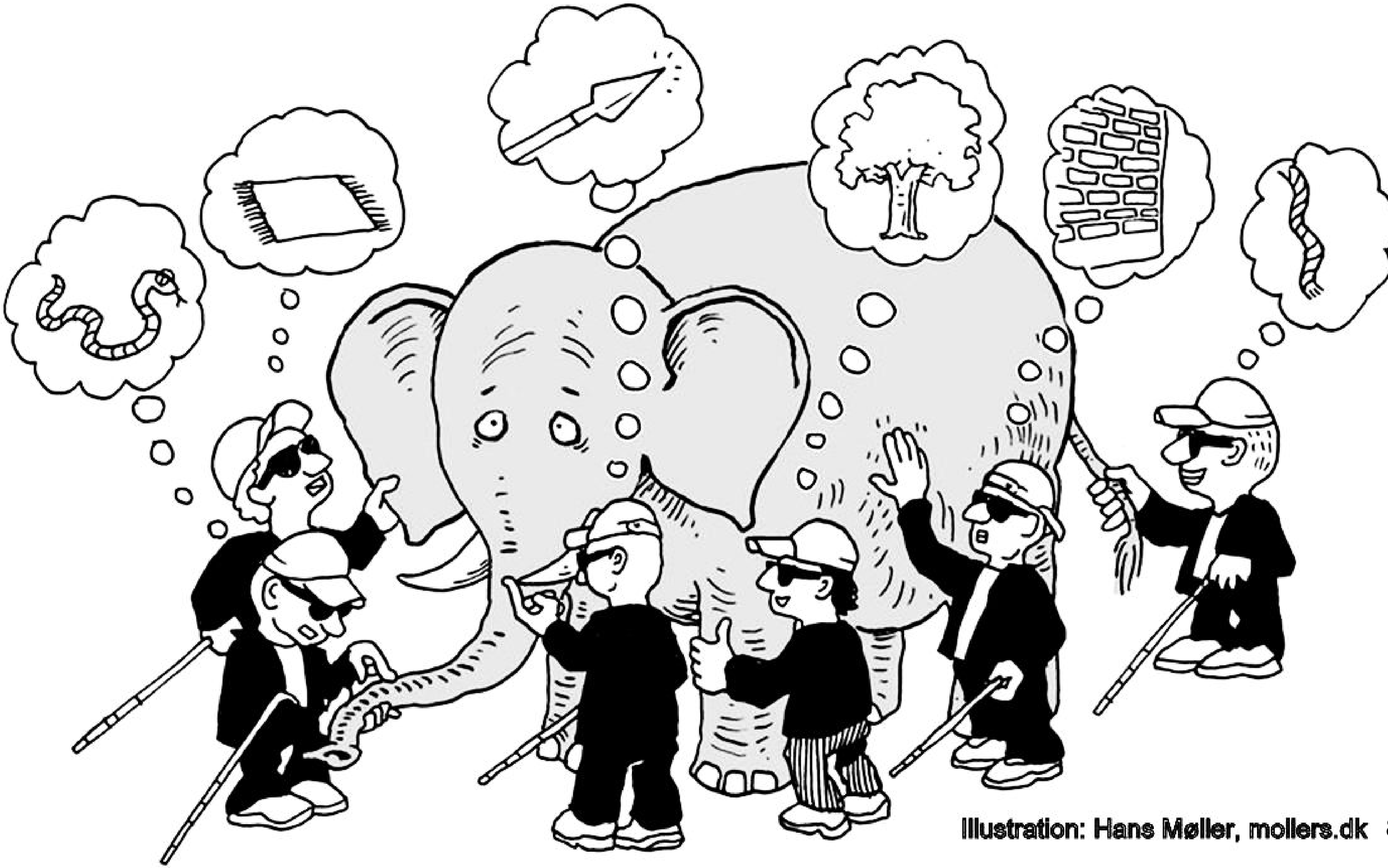
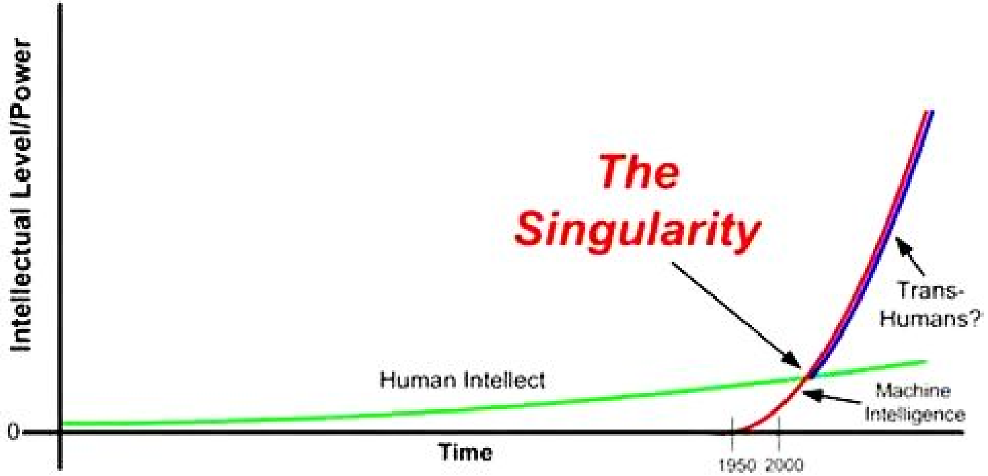


Illustration: Hans Møller, mollers.dk

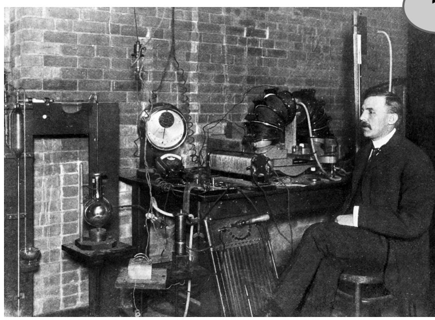
What is coming next?

- Intelligence explosion
  - an uncontrolled hyper-leap in the cognitive ability of AI that Musk and Hawking worry could one day spell doom for the human race
- Open letter
  - Musk+ signed an open letter pledging to conduct AI research for good (http://futureoflife.org/ai-open-letter/)
- OpenAI
  - Musk and Altman create a billion-dollar not-for-profit company that will maximize the power of AI—and then share it with anyone who wants it



New applications? -- the deep, the cool, and the impactful
The story of atomic energy

1



1908 Nobel Prize awarded to Ernest Rutherford in McGill Univ. (Canada) for discovering radioactive half-life

2

After the discovery of the neutron in the 1930s, several teams raced to create elements heavier than uranium for the next Nobel Prize

- Ernest Rutherford (Britain)
- Irène Joliot-Curie (France)
- Enrico Fermi (Italy)
- Meitner & Hahn (Germany)

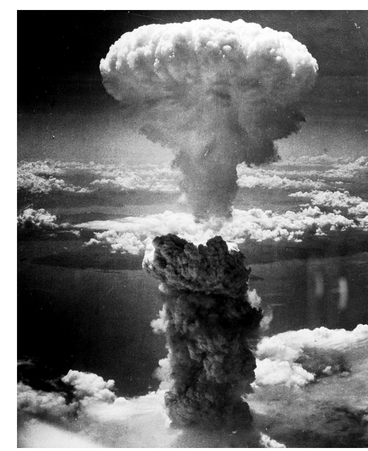
3

1938 Otto Hahn and Lise Meitner discovered nuclear fission

4

1939 Einstein–Szilárd letter was sent to the US government

1942 Chicago Pile-1 (CP-1) is the world's first artificial nuclear reactor



1945 Nagasaki, Japan

References

1. Chris Knight, Camilla Powera and Ian Wattsa, The Human Symbolic Revolution: A Darwinian Account, Cambridge Archaeological Journal 5:1 (1995), pp. 75-114
2. Stanley I. Greenspan, Stuart G. Shanker, The First Idea: How Symbols, Language, and Intelligence Evolved from Our Primate Ancestors to Modern Humans
3. Forrest Wickman, Your Brain's Technical Specs. How many megabytes of data can the human mind hold? 2012
4. Semantic parsing on Freebase from question-answer pairs. Jonathan Berant, Andrew Chou, Roy Frostig, Percy Liang. Empirical Methods in Natural Language Processing (EMNLP), 2013.
5. Lambda dependency-based compositional semantics. Percy Liang. arXiv:1309.4408, 2013.
6. Andrea Gesmundo, Keith Hall, Projecting the Knowledge Graph to Syntactic Parsing. EACL 2014
7. Ndapandula Nakashole, and Tom Mitchell, A Knowledge-Intensive Model for Prepositional Phrase Attachment. ACL 2015
8. J. Zheng, L. Vilnis, S. Singh, J. D. Choi, A. McCallum. Dynamic Knowledge-Base Alignment for Coreference Resolution. CoNLL 2013
9. Jonathan Berant, Percy Liang Semantic Parsing via Paraphrasing, ACL 2014.
10. Panupong Pasupat, Percy Liang, Compositional Semantic Parsing on Semi-Structured Tables. ACL 2015
11. Percy Liang, Michael I. Jordan, Dan Klein, Learning dependency-based compositional semantics. ACL 2011
12. Kelvin Guu, John Miller, Percy Liang, Traversing Knowledge Graphs in Vector Space. EMNLP 2015

References

1. Xin Luna Dong, Evgeniy Gabrilovich, Geremy Heitz, Wilko Horn, Ni Lao, Kevin Murphyy, Thomas Strohmann, Shaohua Sun, Wei Zhang . KDD, 2014 Kevyn B Collins-thompson, Ni Lao, Context-Aware Query Alteration, US Patent 20120233140, 2012
2. Ni Lao, Einat Minkov and William Cohen, Learning relational features with backward random walks , ACL 2015
3. William Yang Wang, Kathryn Mazaitis, Ni Lao, Tom M. Mitchell, William W. Cohen, Efficient Inference and Learning in a Large Knowledge Base: Reasoning with Extracted Information using a Locally Groundable First-Order Probabilistic Logic, to appear in Machine Learning Journal (MLJ 2015), Springer.
4. Ni Lao, Amarnag Subramanya, Fernando Pereira, William W. Cohen Reading The Web with Learned Syntactic-Semantic Inference Rules. EMNLP, 2012
5. Ni Lao, William W. Cohen, Personalized Reading Recommendations for Saccharomyces Genome Database. DILS, 2012
6. Ni Lao, Tom Mitchell, William W. Cohen, Random Walk Inference and Learning in A Large Scale Knowledge Base. EMNLP, 2011
7. Ni Lao, William W. Cohen, Relational retrieval using a combination of path-constrained random walks, Machine Learning, 2010, Volume 81, Number 1, Pages 53-67
8. Ni Lao, William W. Cohen, Fast Query Execution for Retrieval Models based on Path Constrained Random Walks. KDD, 2010

References

1. Michele Banko, Michael J Cafarella, Stephen Soderland, Matt Broadhead and Oren Etzioni, Open Information Extraction from the Web.
2. Mike Mintz, Steven Bills, Rion Snow, Dan Jurafsky, Distant supervision for relation extraction without labeled data. ACL 2009
3. R. Socher, D. Chen, C. Manning, and A. Ng. Reasoning with Neural Tensor Networks for Knowledge Base Completion. In NIPS, 2013.
4. A Fader, S Soderland, O Etzioni, Identifying relations for open information extraction. EMNLP 2011
5. A Fader, LS Zettlemoyer, O Etzioni, Paraphrase-Driven Learning for Open Question Answering. ACL, 2013
6. Limin Yao, Sebastian Riedel and Andrew McCallum, Probabilistic Databases of Universal Schema, Proceedings of the AKBC-WEKEX Workshop at NAACL 2012
7. Mnih etc., Human-level control through deep reinforcement learning. Nature 2015
8. David Silver, “Advanced Topics: Reinforcement Learning” class notes, UCL 2015
9. Baolin Peng, Zhengdong Lu, Hang Li, Kam-Fai Wong, Towards Neural Network-based Reasoning, submitted
10. Pengcheng Yin, Zhengdong Lu, Hang Li, Ben Kao, Neural Enquirer: Learning to Query Tables. submitted
11. Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks. NIPS 2012
12. Richardson, M., Domingos, P. Markov logic: a unifying framework for statistical relational learning. ICML-2004 Workshop on Statistical Relational Learning and its Connections to Other Fields
13. Stensola H, Stensola T, Solstad T, Frøland K, Moser M-B and Moser EI (2012). The entorhinal grid map is discretized. Nature, 492, 72-78.