Knowledge Acquisition

Ni Lao, Xipeng Qiu
8/28/2017

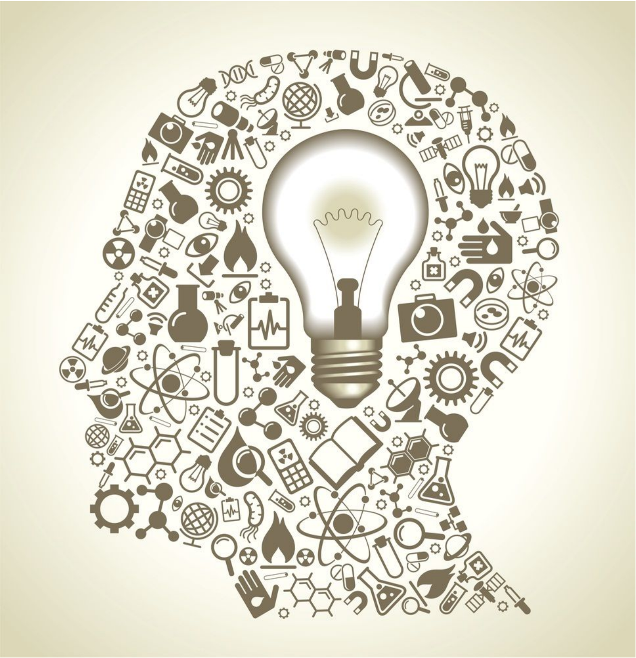
Partial Slides in the section of “Information Extraction” are provided by Dr Kang Liu.

Where does knowledge come from?

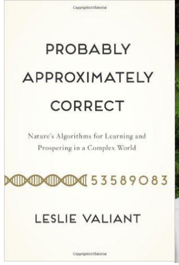
- The human brain contains roughly 100 billion neurons each capable of making around 1,000 connections

- 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

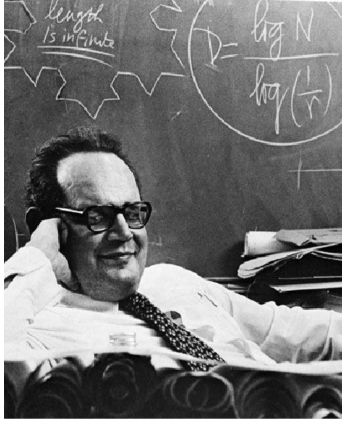


the world

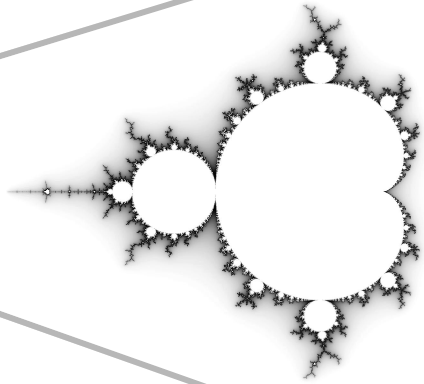
a small machine which can copy large amount of complexity from the world to the brain

a suitable representation

Mandelbrot Set



the nature
of complex
numbers



$$z_0 = 0$$

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

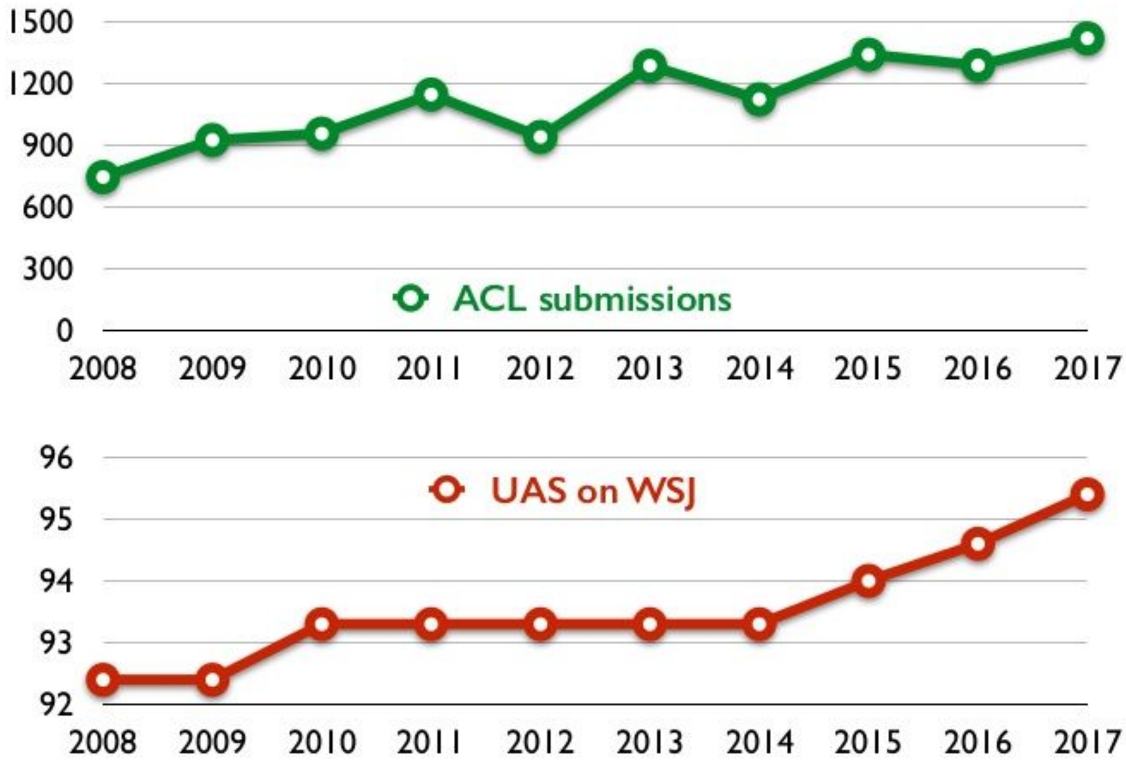
$$c \in M \iff \lim_{n\to\infty} |z_{n+1}| \leq 2$$

ACL 2017

- The premier conference in natural language understanding
- Vancouver, Canada
- 1,318 submissions, 302 accepted
- 1700 attendees

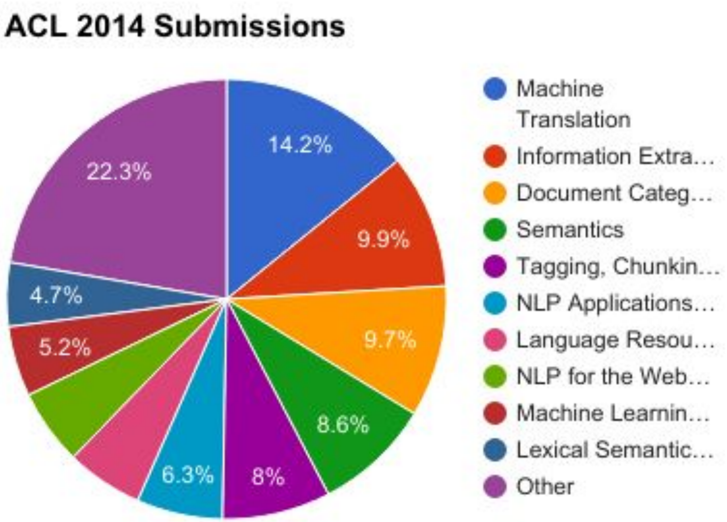
(A)CL is booming!

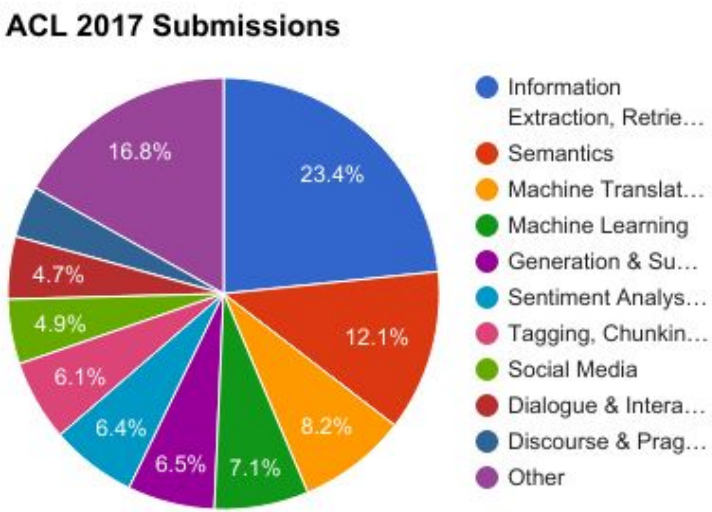
[Nivre ACL talk]



Recent changes

- More IE, summarization/generation, and dialogue

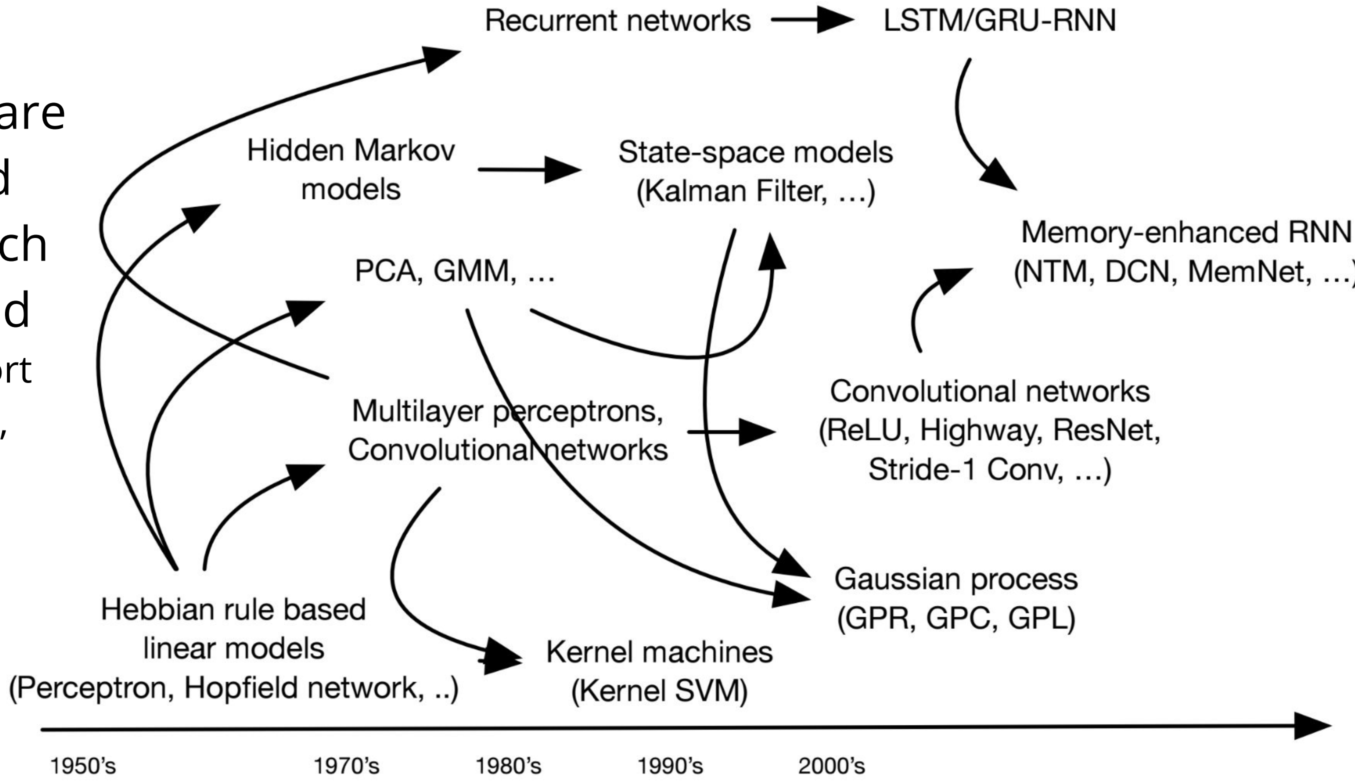




[kyunghyun cho 2017]

In perspective

- Sequence models are the most advanced and impactful, which ML has ever offered
  - A lot of recent effort in adding memory, but no impact yet



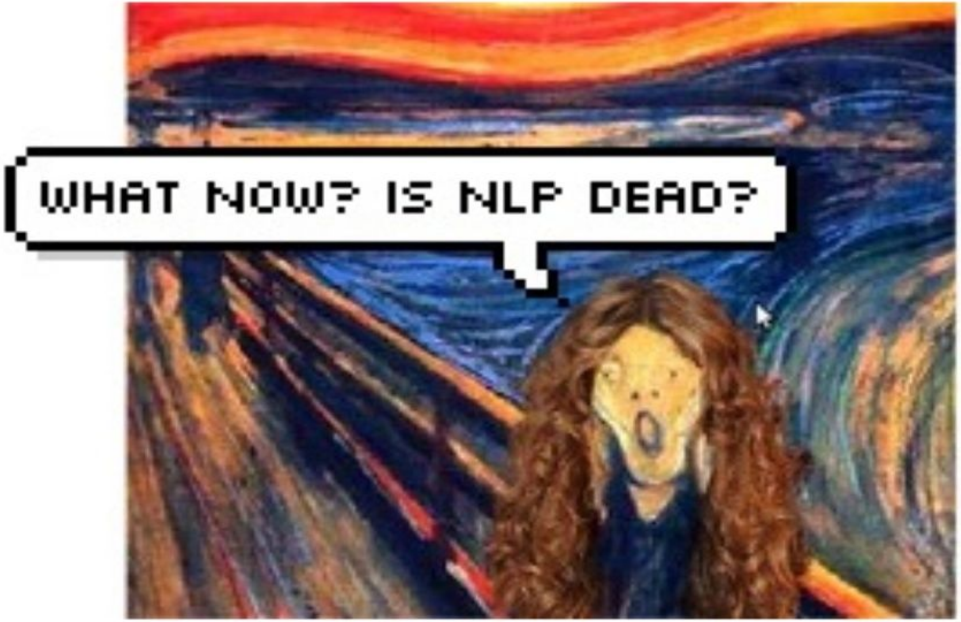
Very inaccurate illustration of the history of ML

[Lapata ACL talk]

Seq2seq models

[Lapata ACL talk]

Lapata's scream

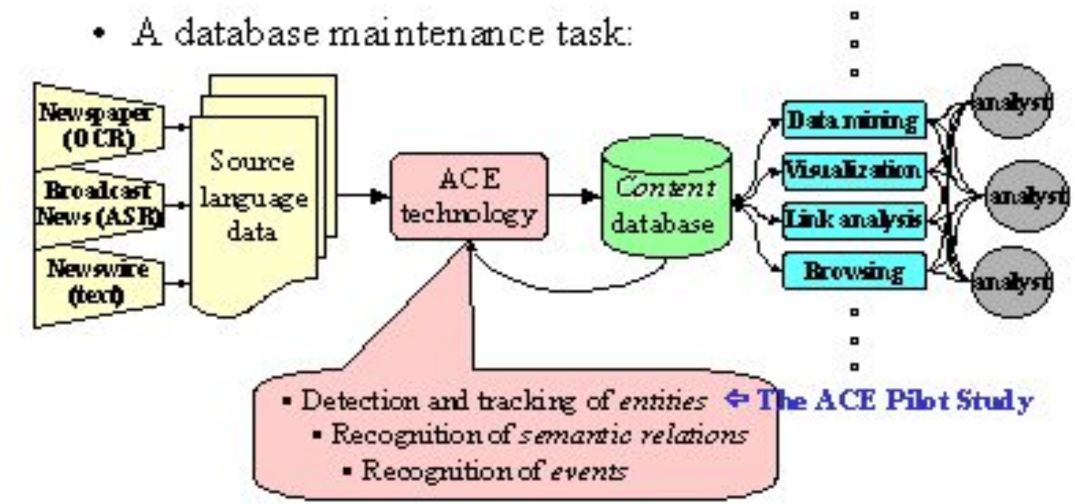


Plan

- **Information extraction**
- Semantic parsing
- Semantic representation

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 restricted domains
  - Supervised training is 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 supervision is scalable, but still limited by the KB schema
    - (Mints et al., 2009)

Problem Formulation

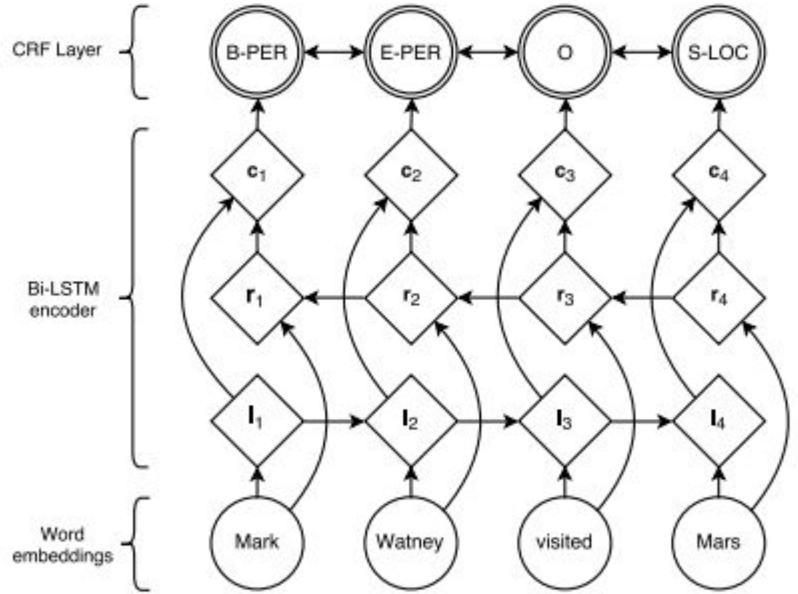
Entity->Relation->Event

1. Named Entity Recognition (NER)
2. Relation Classification: Binary and N-ary
3. Event Extraction

Neural Architectures for Named Entity Recognition

Supervised training

Neural Architecture: LSTM+CRF



https://github.com/clab/stack-lstm-ner

Extract Relations from Unstructured Text

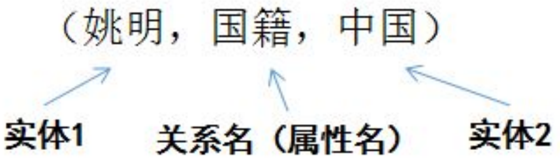
姚明

百科名片 求助编辑

姚明

姚明 1980年生于上海。美国NBA及世界篮球巨星。中国篮球史上里程碑式人物。CBA上海队老板。曾效力于中国国家篮球队，NBA火箭队。2011年7月20日退役。获7次NBA"全明星"，被美国《时代周刊》列入"世界最具影响力100人"，被中国体育总局授予"体育运动荣誉奖章""中国篮球杰出贡献奖"。姚明以高超球技，顽强进取精神，谦逊幽默气质与人格魅力，赢得了世界声誉。让世界对中国有了新的了解与认识，让更多的人关注、喜爱篮球。姚明成为东西方文化的桥梁，具有史无前例的个人影响力。姚明的意义与价值，超越了篮球运动，超越了国界。

中文名：	姚明	专业特点：	20英尺外精确跳投
外文名：	Yao Ming	主要奖项：	NBA全明星赛(7次)
别名：	小巨人 移动长城		ESPN全球最有潜力运动员奖(2000)
国籍：	中国		劳伦斯世界最佳新秀奖(2003)
民族：	汉族		中国篮球杰出贡献奖
出生地：	上海	重要事件：	专题影片《姚明年》发行
出生日期：	1980年9月12日	祖籍：	江苏苏州吴江震泽
身高：	2.23米（7.32英尺）	位置：	中锋
体重：	140.6kg	鞋码：	18码
运动项目：	篮球	自传：	《我的世界我的梦》
所属运动队：	NBA火箭队	生涯最高分：	41分



Extract Relations from Unstructured Text

Sentence Level

The [haft]$_{e1}$ of the [axe]$_{e2}$ is made of yew wood.

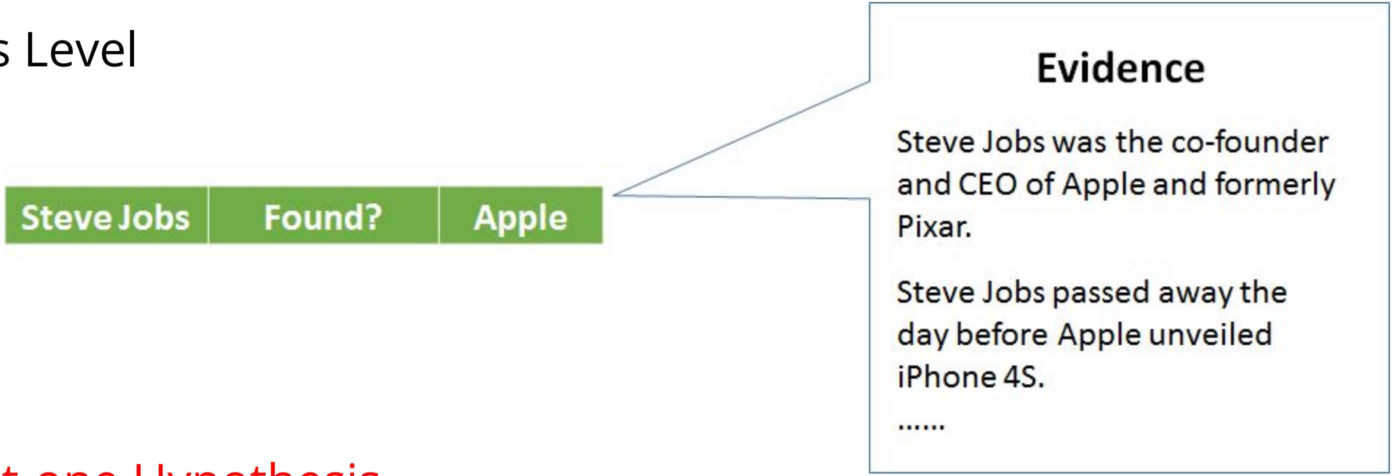
Component-Whole(e1,e2)

The [fire]$_{e1}$ inside WTC was caused by exploding [fuel]$_{e2}$.

Cause-Effect(e1,e2)

Extract Relations from Unstructured Text

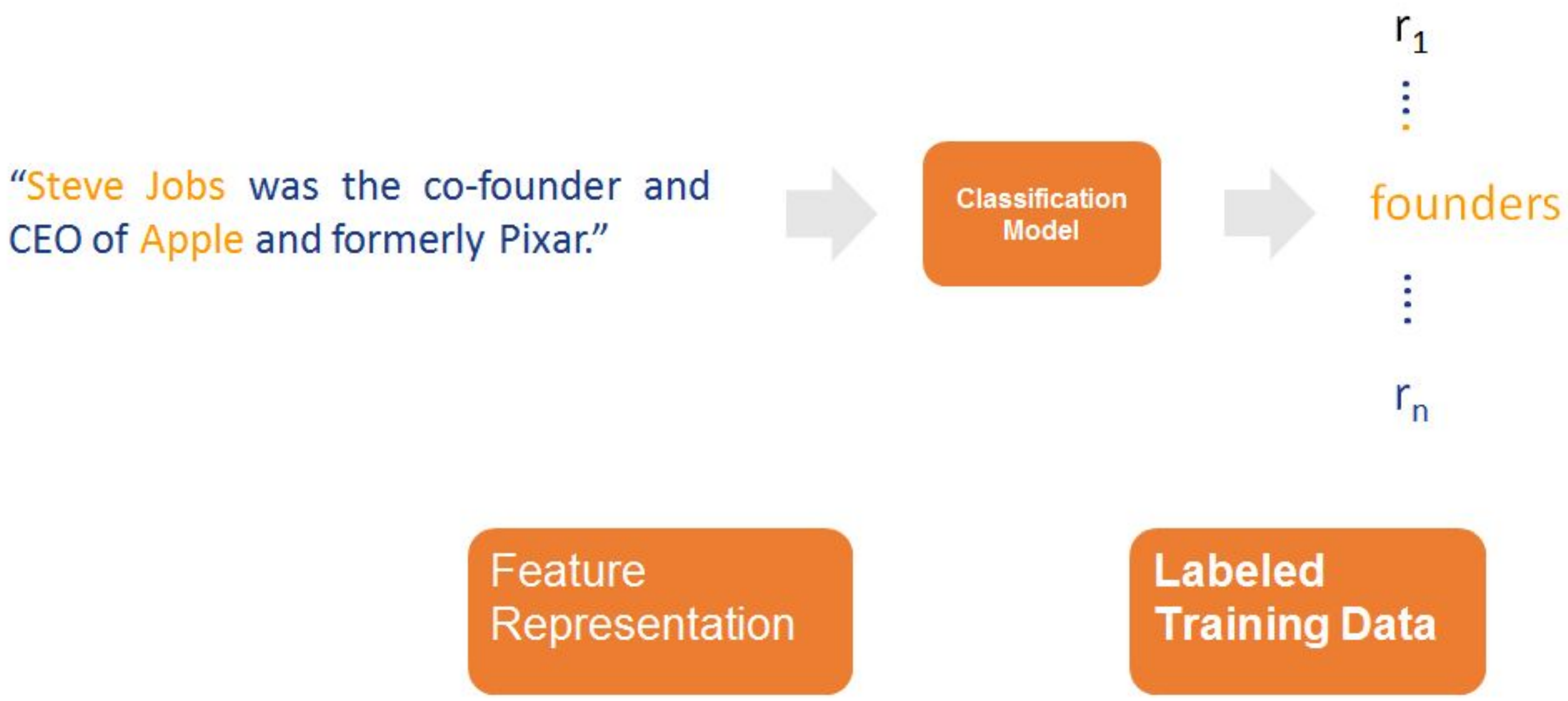
Corpus Level



At-least-one Hypothesis

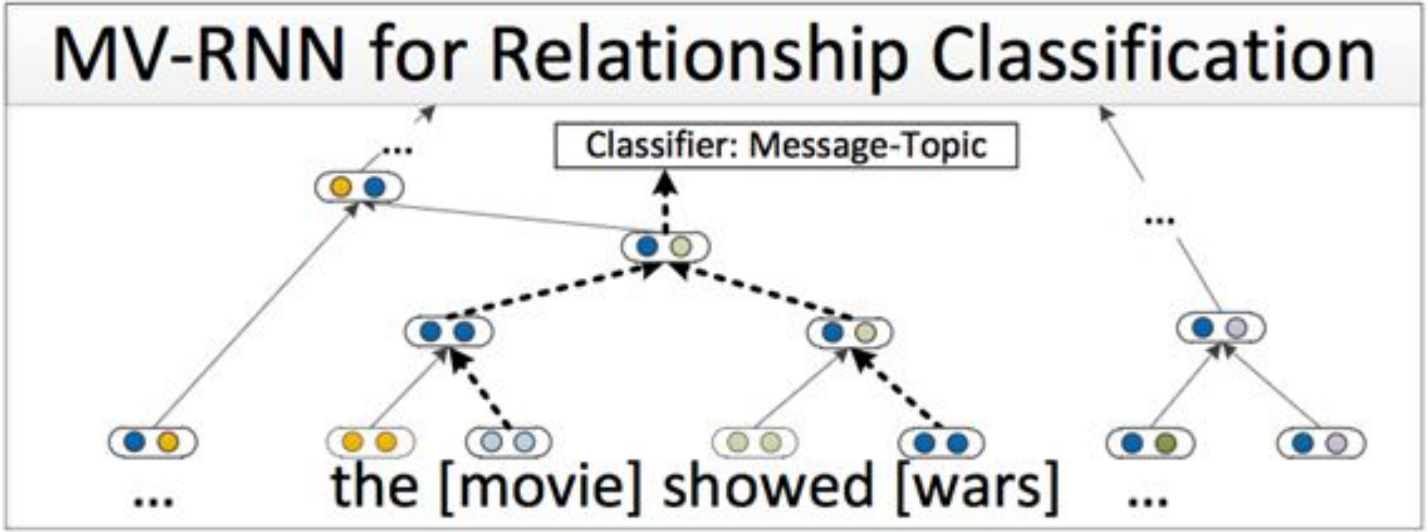
If two entities participate in a relation, at least one sentence that mentions these two entities might express that relation.

Convert into Classification Problem

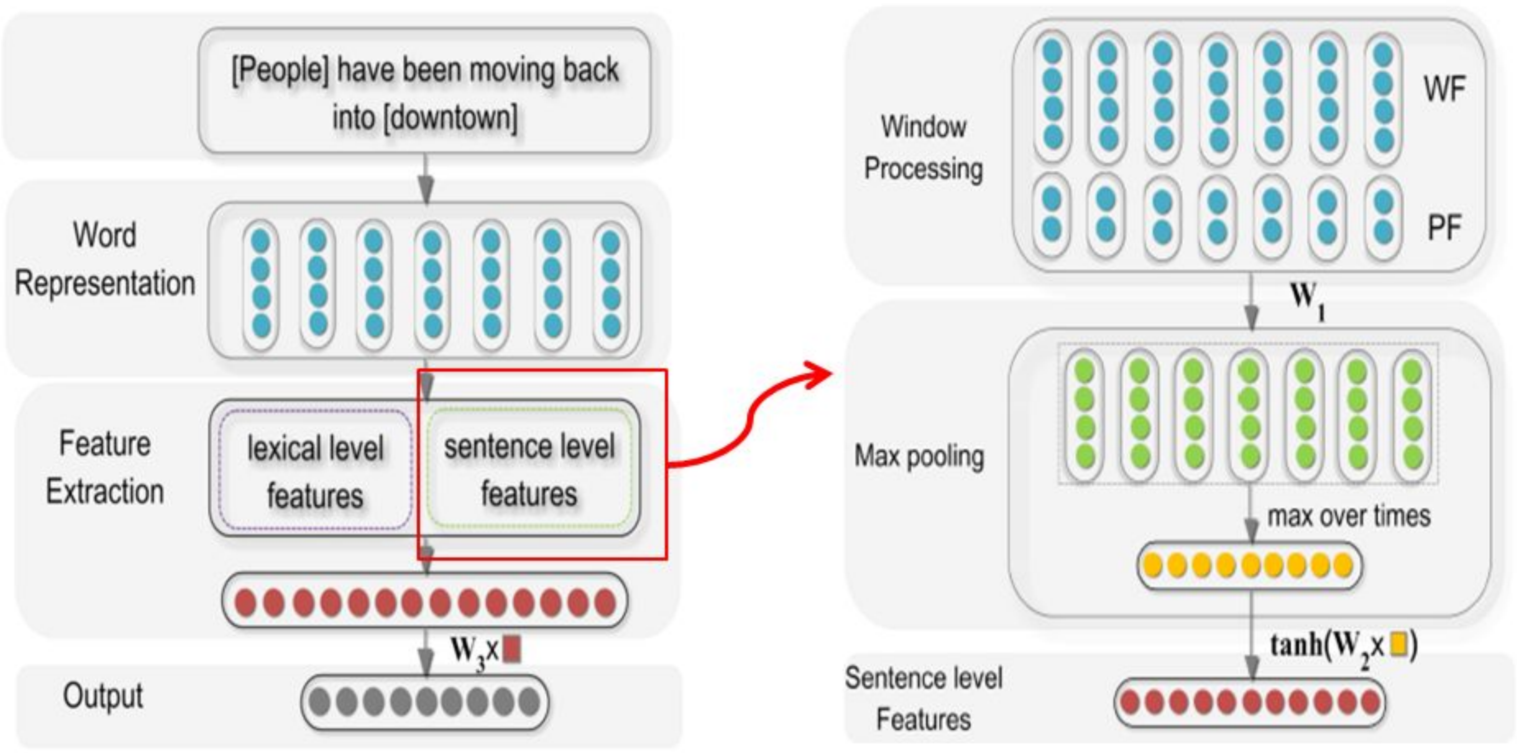


Socher, EMNLP 2012

Matrix-Vector Recursive Neural Network for Relation Classification

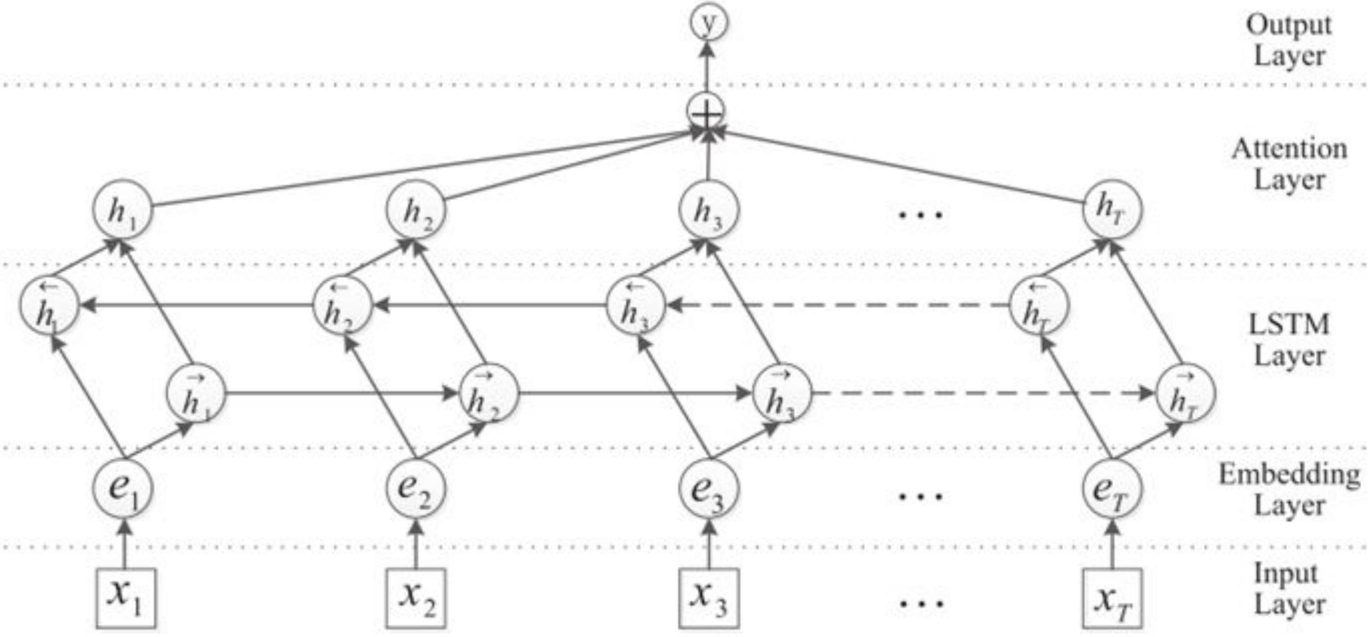


Convolutional Neural Network



Attention-Based Bidirectional Long Short-Term Memory Networks for Relation Classification

Zhou, ACL 2016



•SemEval-2010 Task 8

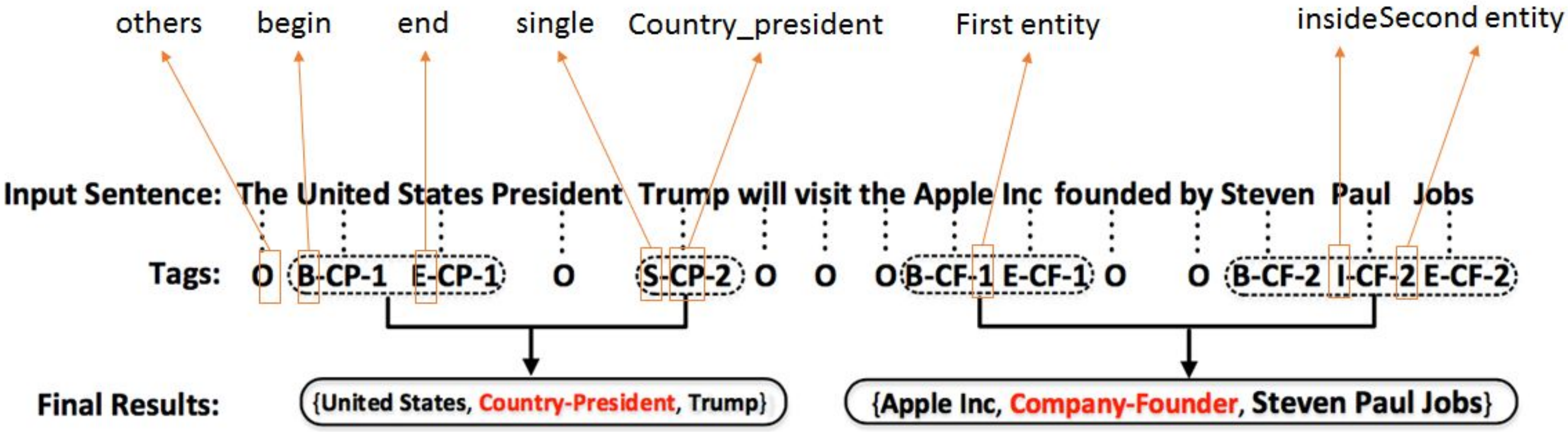
Performances

of training instance 8,000
# of test instance 2,717
# of relationships 19

Model	Feature Set	F1
SVM (Rink and Harabagiu, 2010)	POS, prefixes, morphological, WordNet, dependency parse, Levin classed, ProBank, FramNet, NomLex-Plus, Google n-gram, paraphrases, TextRunner	82.2
CNN (Zeng et al., 2014)	WV (Turian et al., 2010) (dim=50)	69.7
	+ PF + WordNet	82.7
RNN (Zhang and Wang, 2015)	WV (Turian et al., 2010) (dim=50) + PI	80.0
	WV (Mikolov et al., 2013) (dim=300) + PI	82.5
SDP-LSTM (Yan et al., 2015)	WV (pretrained by word2vec) (dim=200), syntactic parse	82.4
	+ POS + WordNet + grammar relation embeddings	83.7
BLSTM (Zhang et al., 2015)	WV (Pennington et al., 2014) (dim=100)	82.7
	+ PF + POS + NER + WNSYN + DEP	84.3
BLSTM	WV (Turian et al., 2010) (dim=50) + PI	80.7
Att-BLSTM	WV (Turian et al., 2010) (dim=50) + PI	82.5
BLSTM	WV (Pennington et al., 2014) (dim=100) + PI	82.7
Att-BLSTM	WV (Pennington et al., 2014) (dim=100) + PI	84.0

Zheng, ACL 2015

Joint Extraction of Entities and Relations



Number of tags: $2 * 4 * |R| + 1$

$|R|$ is the number of relation, 4 means begin, end, single, inside

Distant Supervision for Relation Extraction

Distant supervision automatically generates amount of training data, overcome the manually-labeling problem.

Freebase

Knowledge base		
Relation	Entity 1	Entity 2
Founder	Steve Jobs	Apple
...	...	...

The New York Times

Sentence
Steve Jobs was the co-founder and CEO of Apple and formerly Pixar.
Steve Jobs passed away the day before Apple unveiled iPhone 4S.
...

Multi-instance Learning

Zeng, EMNLP 2015

- T Bags
- i-th bag has q_i instances

$$M_i = \{m_i^1, m_i^1, \cdots m_i^{q_i}\}$$

- Objective function:

$$J(\theta) = \sum_{i=1}^{T} \log p(y_i | m_i^j; \theta)$$

where

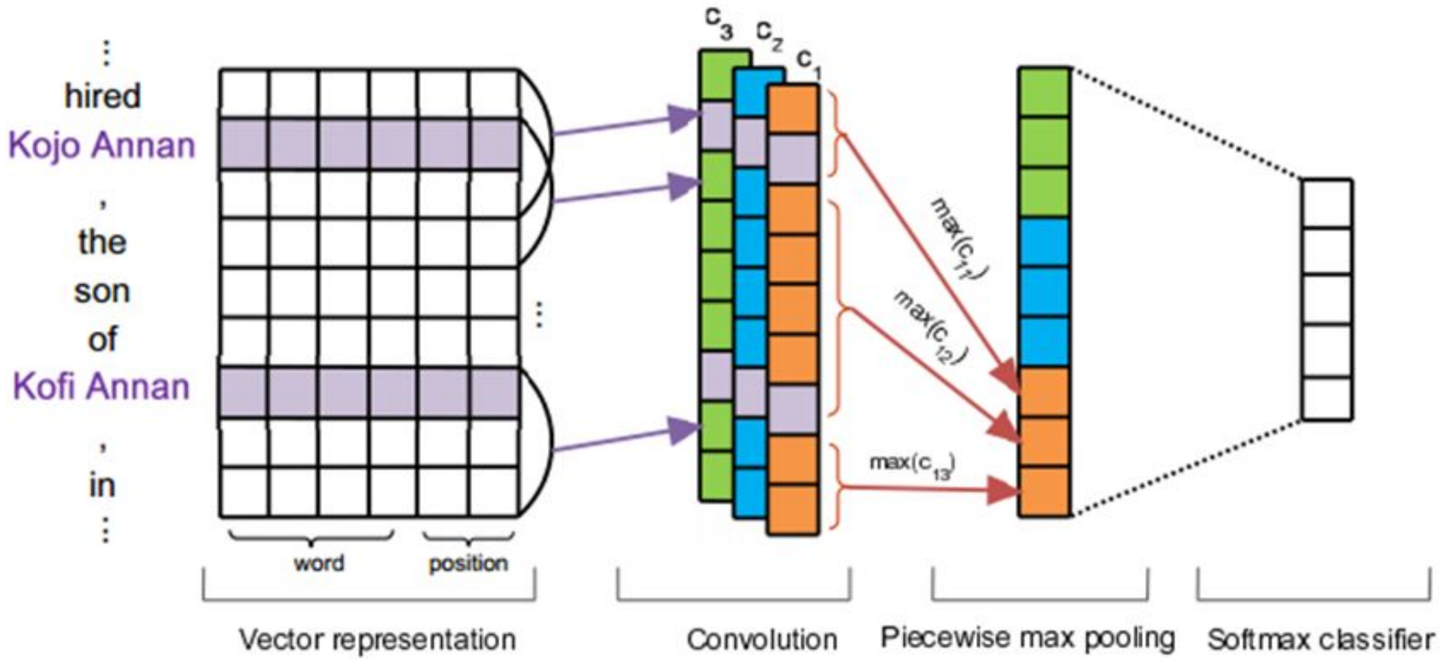
$$j^* = \arg\max_j p(y_i | m_i^j; \theta) \quad 1 \le j \le q_i$$

Algorithm 1 Multi-instance learning

1: Initialize θ. Partition the bags into mini-batches of size b_s.
2: Randomly choose a mini-batch, and feed the bags into the network one by one.
3: Find the j-th instance m_i^j $(1 \le i \le b_s)$ in each bag according to Eq. (9).
4: Update θ based on the gradients of m_i^j $(1 \le i \le b_s)$ via Adadelta.
5: Repeat steps 2-4 until either convergence or the maximum number of epochs is reached.

Piece-wise CNN Model

Zeng, EMNLP 2015



Selective Attention over Instances

Lin, ACL 2016

Selective Attention

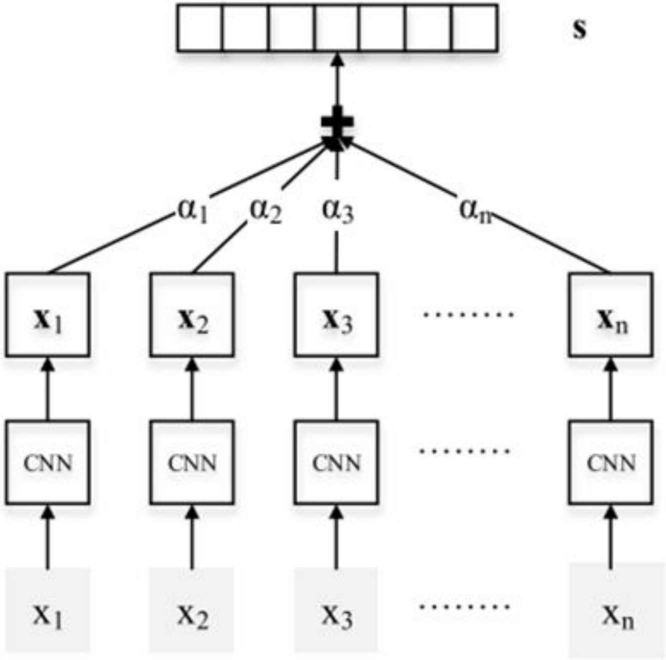
$$\alpha_i = \frac{\exp(e_i)}{\sum_k \exp(e_k)}$$
$$e_i = x_i A r$$

A is a weighted diagonal matrix

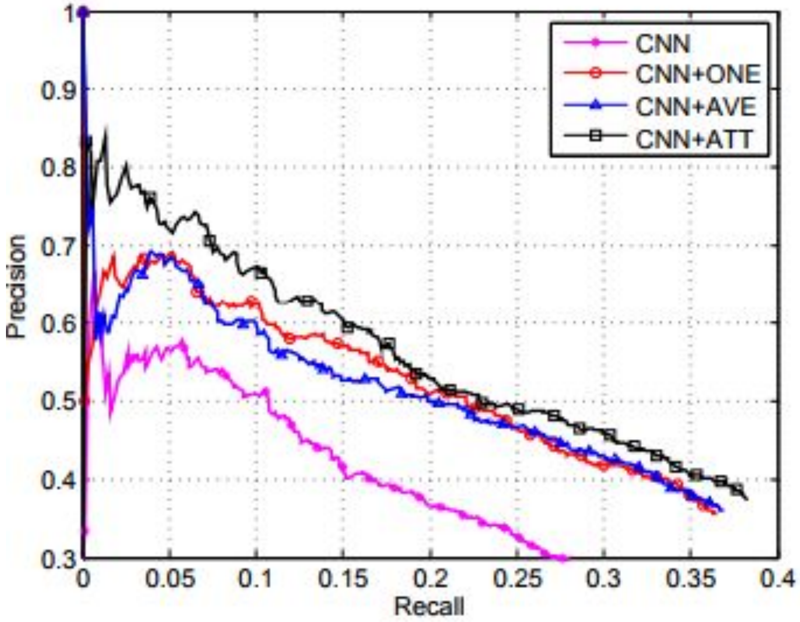
•*r* is the query vector associated with relation r

The final set vector s

$$\mathbf{s} = \sum_i \alpha_i \mathbf{x}_i$$



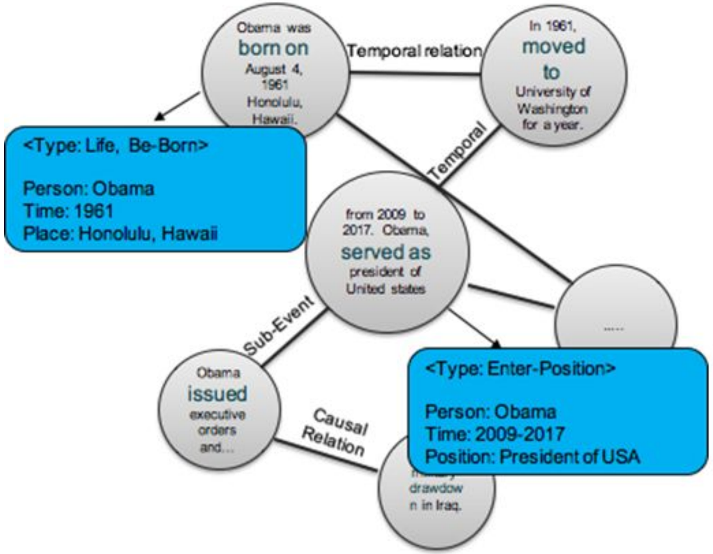
Case Study



Relation	employer_of
Low	When Howard Stern was preparing to take his talk show to **Sirius Satellite Radio**, following his former boss, **Mel Karmazin**, Mr. Hollander argued that ...
High	**Mel Karmazin**, the chief executive of **Sirius Satellite Radio**, made a lot of phone calls ...
Relation	place_of_birth
Low	**Ernst Haefliger**, a Swiss tenor who ... roles , died on Saturday in **Davos**, Switzerland, where he maintained a second home.
High	**Ernst Haefliger** was born in **Davos** on July 6, 1919, and studied at the Wettinger Seminary ...

From Static Knowledge to Dynamic Knowledge

Dynamic Knowledge: Event-Centric Knowledge Graph



出生事件
- 出生日期
- 出生地点
- 姓名

结婚事件
- 结婚日期
- 结婚地点
- 男方
- 女方

离职事件
- 离职日期
- 公司
- 职位

地震事件
- 震中
- 震级
- 震源
- 伤亡人数
- 财产损失

暴恐事件
- 地点
- 时间
- 伤亡人数
- 被攻击方
- 实施方

收购事件
- 收购金额
- 收购方
- 被收购方
- 时间

事件框架（脚本）

Extract Event from Unstructured Text

Trigger	Quit (a "Personnel/End-Position" event)	
Arguments	Role = Person	Barry Diller
	Role = Organization	Vivendi Universal Entertainment
	Role = Position	Chief
	Role = Time-within	Wednesday (2003-03-04)

Definition of Event Extraction

Definition:

Event trigger, Event Type, Event argument, Argument role

Barry Diller on Wednesday *quit* as chief of Vivendi Universal Entertainment.

Trigger	Quit (a "Personnel/End-Position" event)	
Arguments	Role = Person	Barry Diller
	Role = Organization	Vivendi Universal Entertainment
	Role = Position	Chief
	Role = Time-within	Wednesday (2003-03-04)

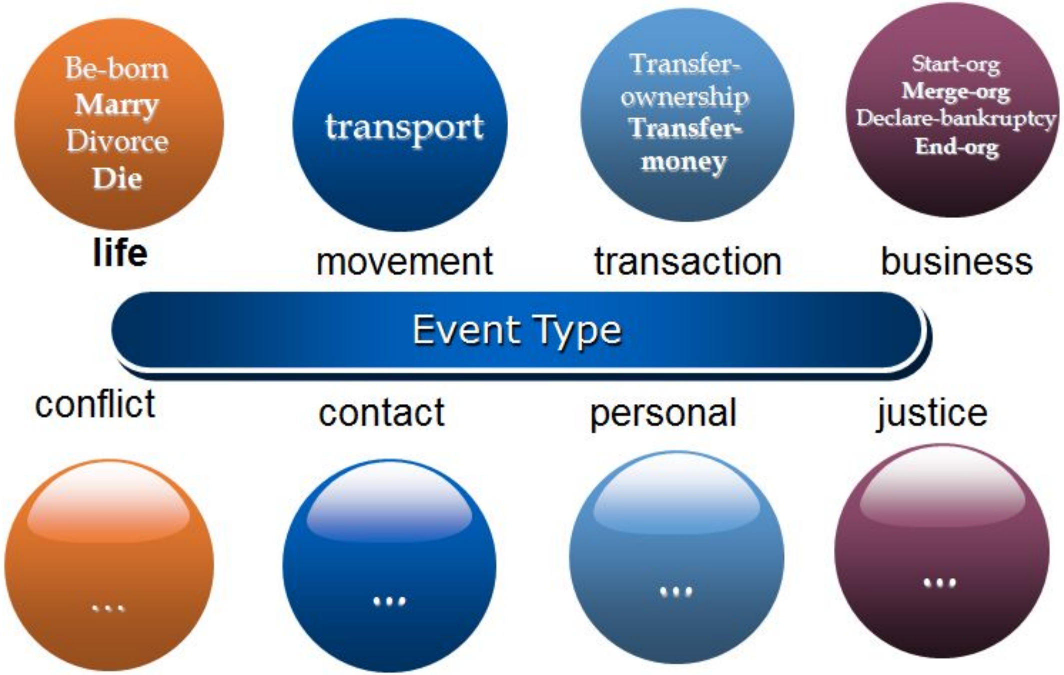
1. Event Identification (Trigger Words)
2. Event Type Identification
3. Argument Identification
4. Argument Role Identification

Event Extraction vs. Relation Extraction

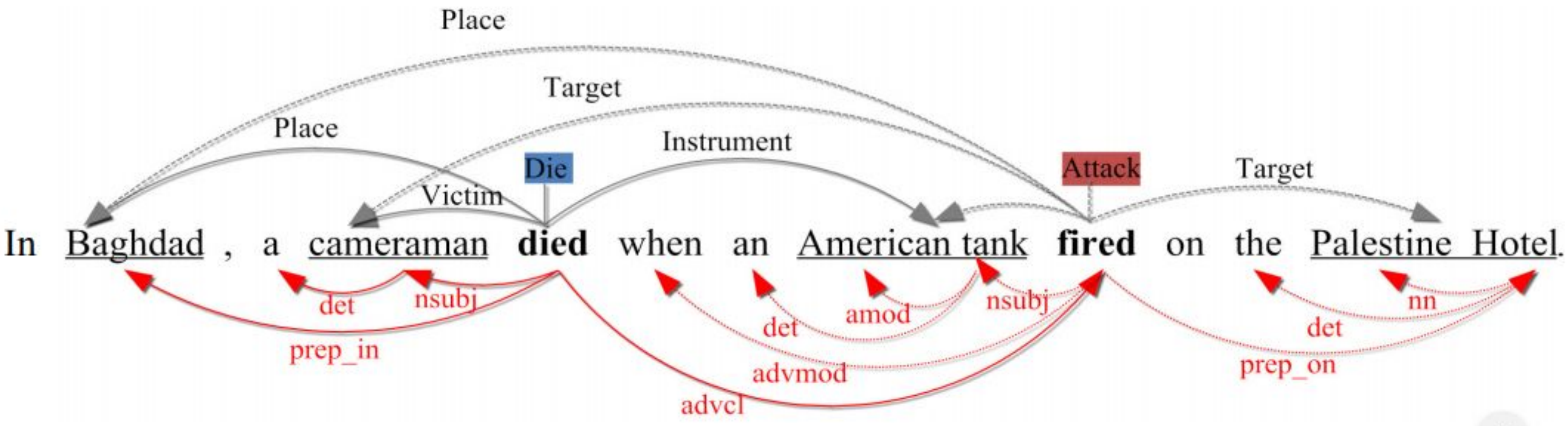
- Relation Extraction
  - Identify the relation between **two given entities**

- Event Extraction
  - Identify the relation between **an event and an entity**

Type of Events



Example

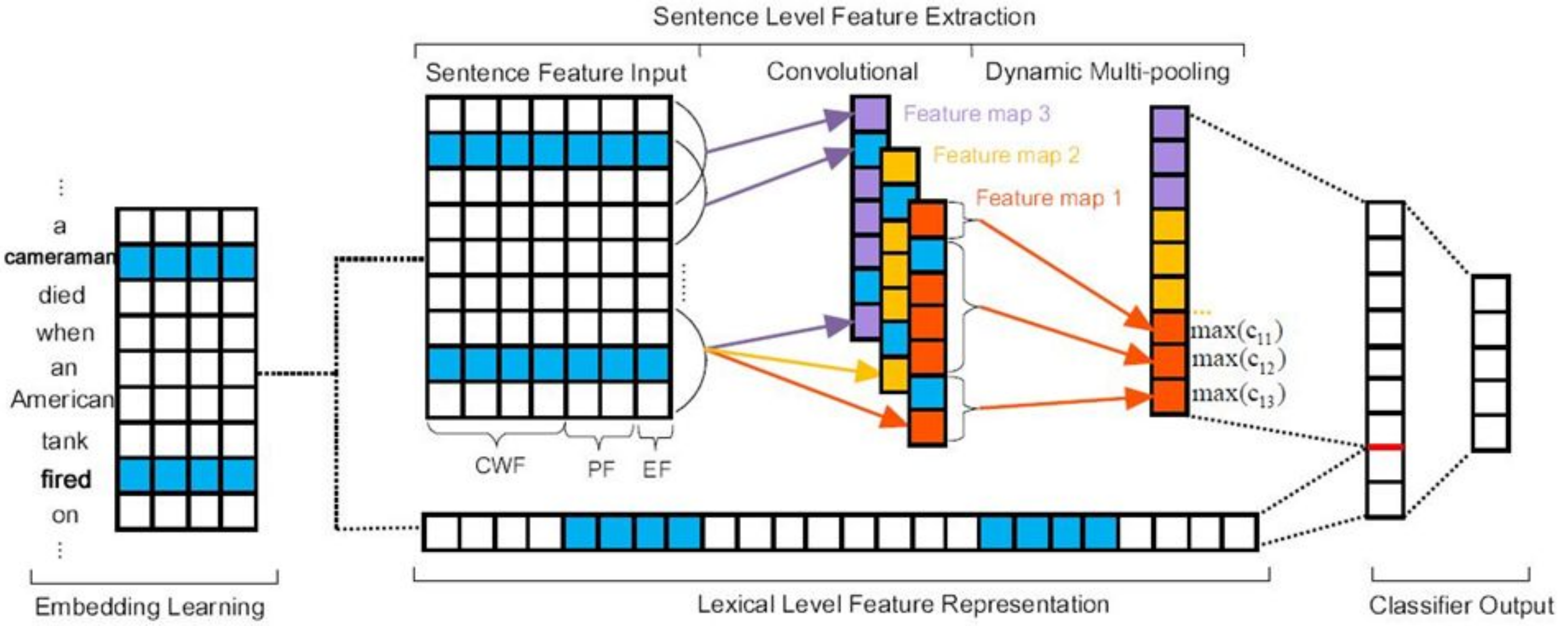


He has fired his air defense chief.

Position (End-Position)

Event Extraction via Dynamic Multi-Pooling Convolutional Neural Networks

Chen, ACL 2015



Experiments

Dataset：ACE 2005

Testing: 40 newswire articles

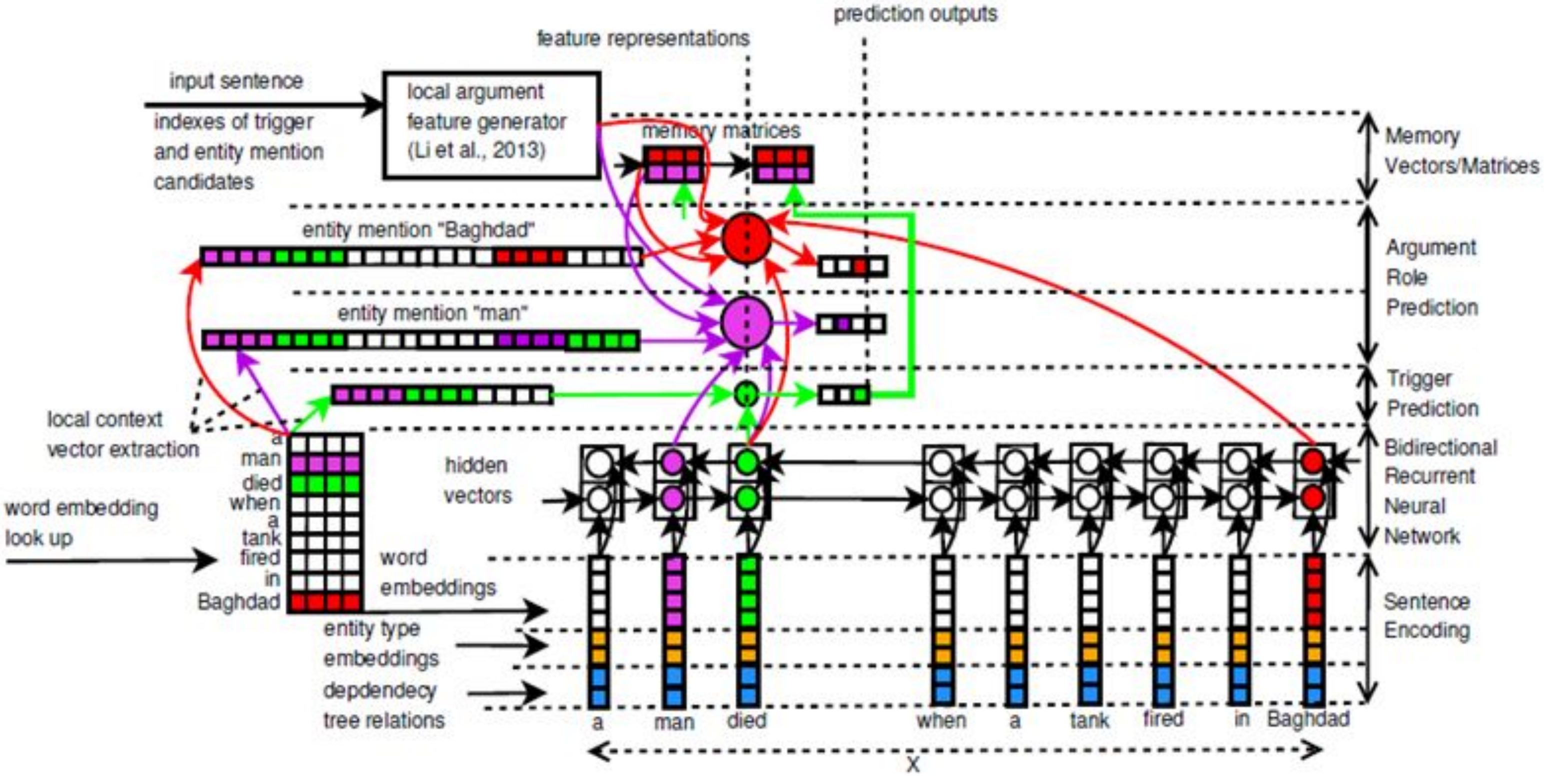
Development: 30 documents

Training: The rest (529) documents

Methods	Trigger Identification(%)			Trigger Identification + Classification(%)			Argument Identification(%)			Argument Role(%)		
	P	R	F	P	R	F	P	R	F	P	R	F
Li's baseline	76.2	60.5	67.4	74.5	59.1	65.9	74.1	37.4	49.7	65.4	33.1	43.9
Liao's cross-event	N/A			68.7	68.9	68.8	50.9	49.7	50.3	45.1	44.1	44.6
Hong's cross-entity	N/A			72.9	64.3	68.3	53.4	52.9	53.1	51.6	45.5	48.3
Li's structure	76.9	65.0	70.4	73.7	62.3	67.5	69.8	47.9	56.8	64.7	44.4	52.7
DMCNN model	80.4	67.7	**73.5**	75.6	63.6	**69.1**	68.8	51.9	**59.1**	62.2	46.9	**53.5**

Nguyen, NAACL 2016

Joint Event Extraction via Recurrent Neural Networks



Brief Summary of IE

- Deep Learning
  - Sentence Representation (CNN/RNN)
  - Attention

- Data
  - Human Labeled
  - Distant Supervision

A representative IE domain

- Its conclusions can be applied to other domains

Natural Language Processing for Precision Medicine

Hoifung Poon, Chris Quirk, Kristina Toutanova, Scott Wen-tau Yih

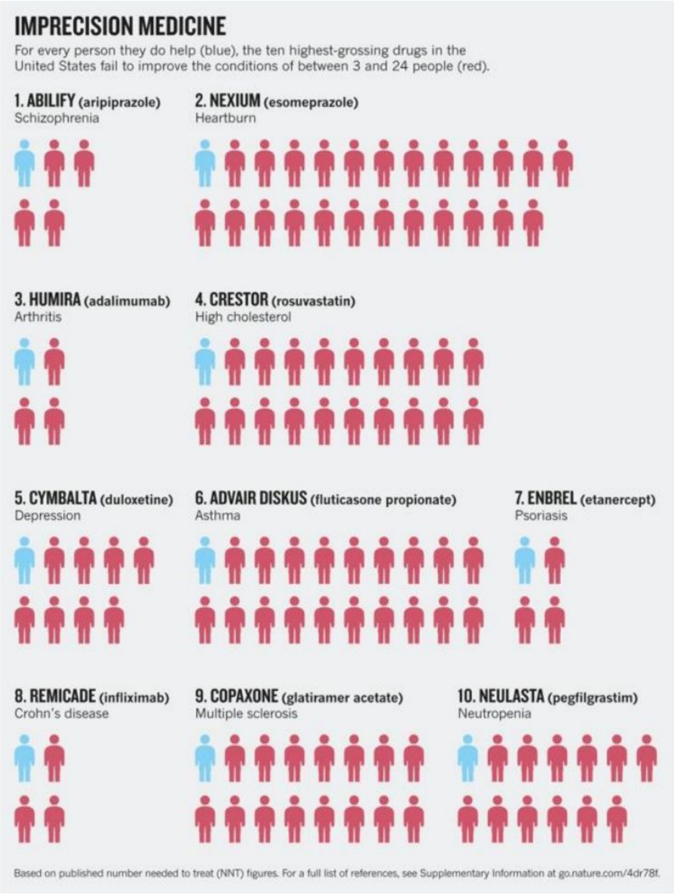
[Poon+ 2017]

IE

“~~Biomedicine~~ is an ocean that’s one meter deep”

Medicine Today Is Imprecise

[Poon+ 2017]

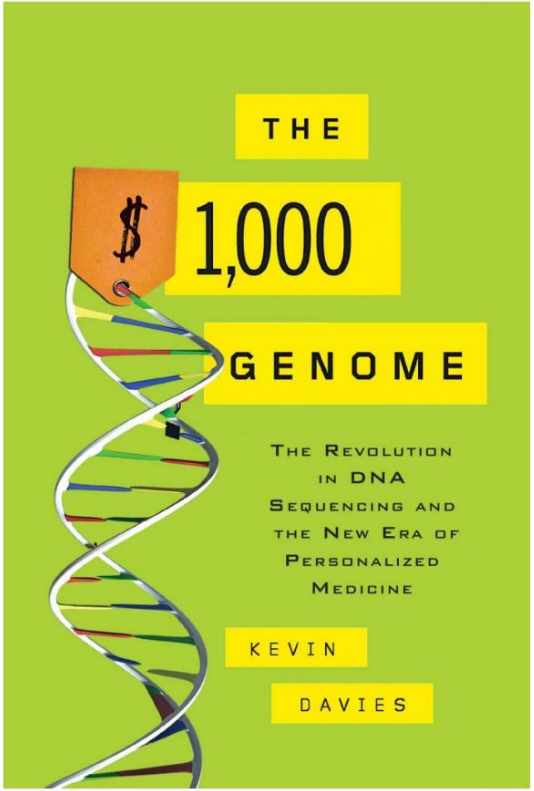


Top 20 drugs
80% non-responders

Wasted
1/3 health spending
$1 Trillion / year

[Poon+ 2017]

Disruption: Big Data



Accenture study: 93% of US doctors using EMRs

May 14, 2013 IHQRE informatics, IHQRE Journal Club EHR, EMR, Meaningful Use

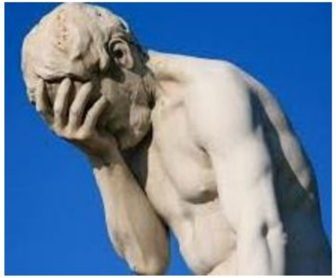
2009 – 2013: 40% → 93%

[Poon+ 2017]

Why We Haven't Solved Precision Medicine?

... ATTCGGATATTTAAGGC ...
... ATTCGGGTATTTAAGCC ...
... ATTCGGATATTTAAGGC ...
... ATTCGGGTATTTAAGCC ...
... ATTCGGATATTTAAGGC ...
... ATTCGGGTATTTAAGCC ...

High-Throughput Data



Discovery

Bottleneck #1: Knowledge

Bottleneck #2: Reasoning

AI is the key to overcome these bottlenecks

[Poon+ 2017]

Key Scenario: Molecular Tumor Board

Problem: Hard to scale

U.S. 2016: 1.7 million new cases, 600K deaths

902 cancer hospitals

Knowledge bottleneck

E.g., given a tumor sequence, determine:

- What genes and mutations are important
- What drugs might be applicable

Can do manually but hard to scale

Reasoning bottleneck

E.g., personalize drug combinations

Can't do manually, ever

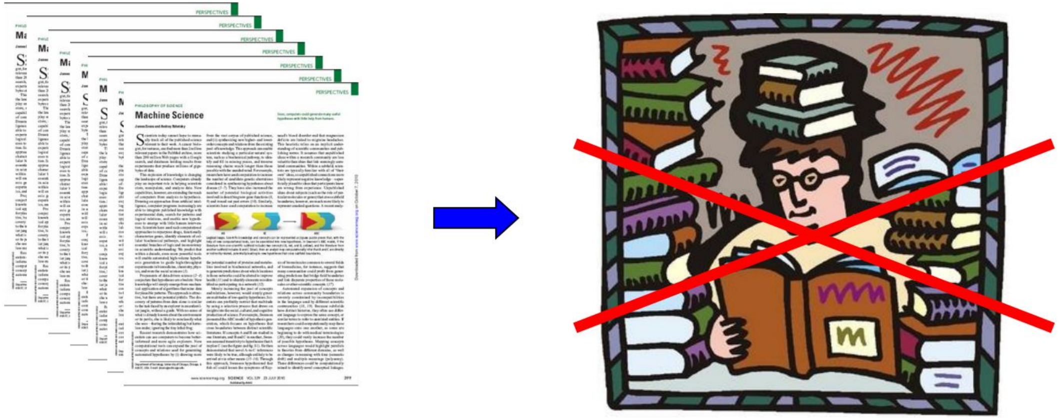
PubMed

[Poon+ 2017]

27 million abstracts

Two new abstracts every minute

Adds over one million every year



[Poon+ 2017]

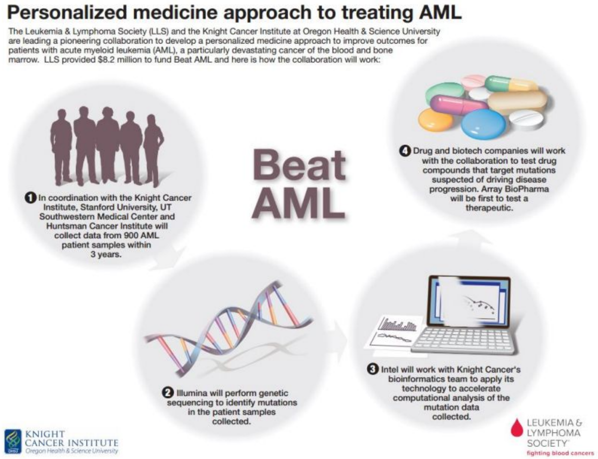
Example: Personalize Drug Combos

Targeted drugs: 149

Pairs: 11,026

Tested: 102 (in two years)

Unknown: 10,924



Can we find good combos in months, not centuries?

[Poon+ 2017]

Challenge: Cross-Sentence Relation Extraction

The deletion mutation on exon-19 of EGFR gene was present in 16 patients, while the L858E point mutation on exon-21 was noted in 10.

All patients were treated with gefitinib and showed a partial response.

Gefitinib could be used to treat tumors w. EGFR mutation L858E.

TREAT(Gefitinib, EGFR, L858E)

[Poon+ 2017]

Generalize to N-ary Relations

The deletion mutation on exon-19 of EGFR gene was present in 16 patients, while the L858E point mutation on exon-21 was noted in 10.

All patients were treated with gefitinib and showed a partial response.

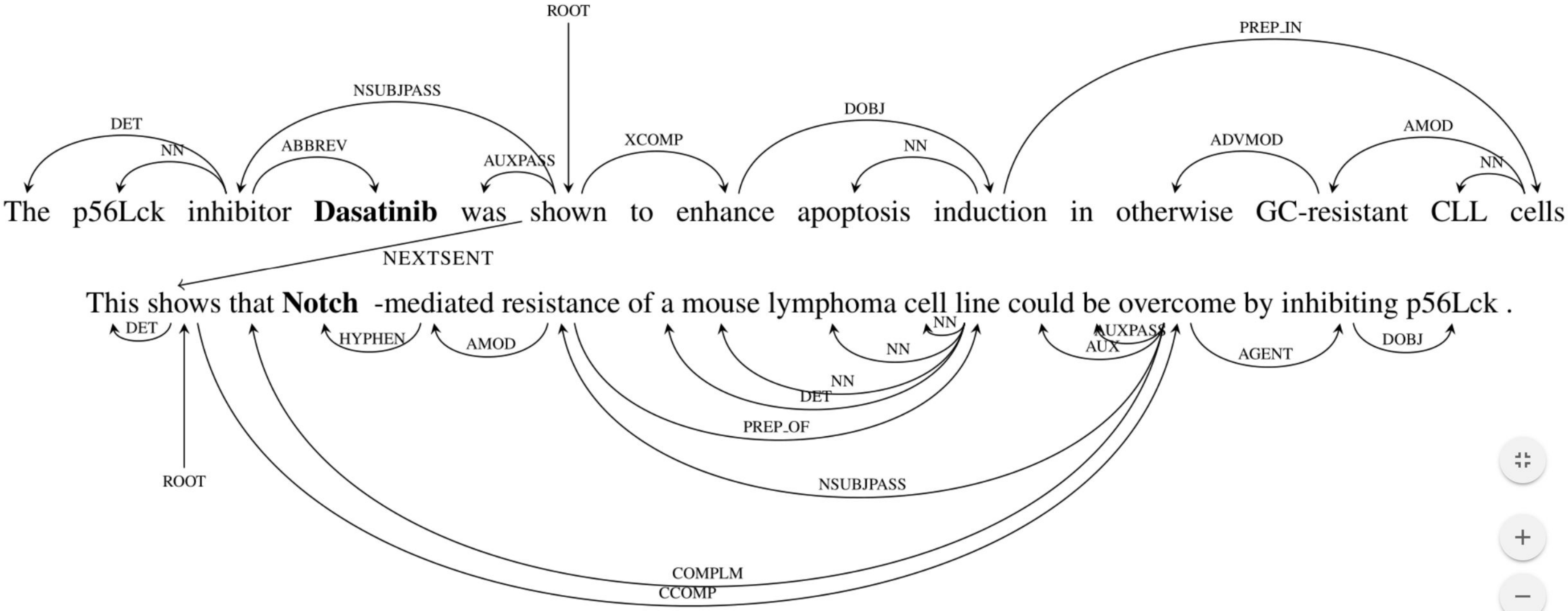
Peng et al. "Cross-Sentence N-ary Relation Extraction with Graph LSTM", *TACL-17*.

TACL 2017

[Poon+ 2017]

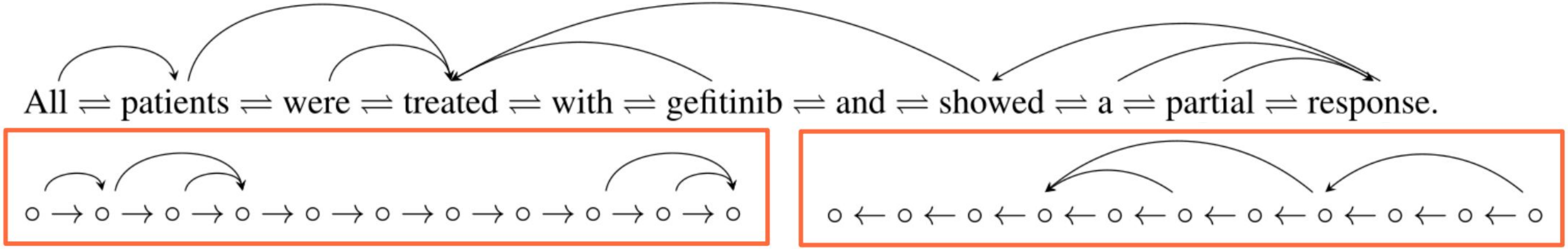
Document Graph

Sequence, syntax, discourse



[Poon+ 2017]

Asynchronous Update



Forward Pass

Backward Pass

PubMed-Scale Extraction

[Poon+ 2017]

Relations	Single-Sent.	Cross-Sent.
Candidates	169,168	332,969
$p \geq 0.5$	32,028	64,828
$p \geq 0.9$	17,349	32,775
GDKD	162	

Orders of magnitude more knowledge by machine reading

[Poon+ 2017]

So far: Relationships Directly Expressed in Text

Tumor suppressor P53 down-regulates the activity of BCL-2 proteins.

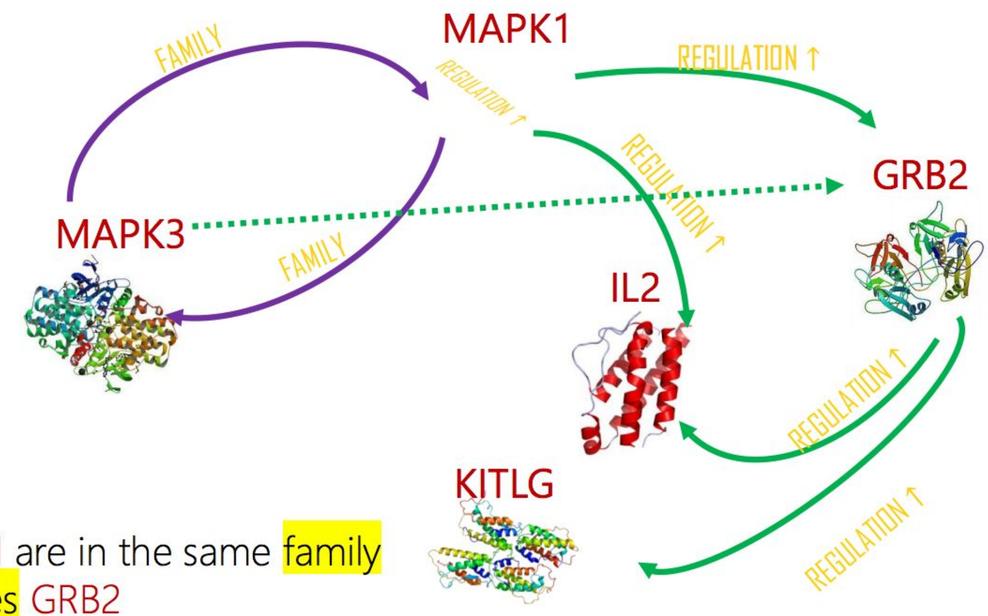
↓

negative_regulation(P53,BCL-2)

Reasoning: combining several pieces of relevant information.

Genomics Knowledge Base (Network)

[Poon+ 2017]



MAPK3 and MAPK1 are in the same family
MAPK1 up-regulates GRB2

Likely that MAPK3 up-regulates GRB2

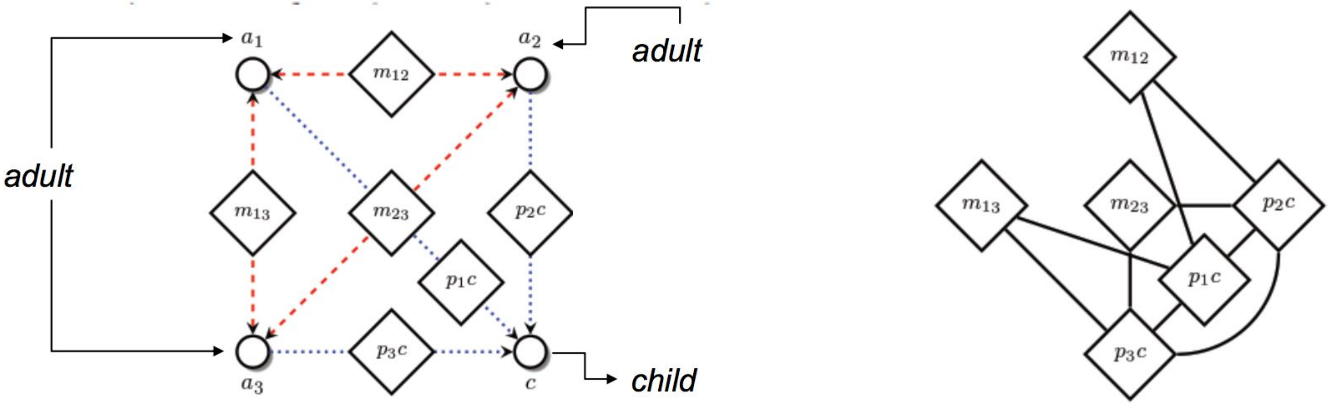
[Poon+ 2017]

Graphical models are expensive

Statistical relational learning [Getoor & Taskar, 2007]

- Modeling dependencies among the truth values of multiple possible relations

$$F_1 : (x, parentOf, z) \wedge (y, parentOf, z) \Rightarrow (x, marriedTo, y)$$



- Can be prohibitively expensive (e.g. marginal inference is exponential in the treewidth for Markov Random Fields)

[Poon+ 2017]

Embeddings and random walks are more scalable

Knowledge base embedding

- Assumes truth values of facts are independent given latent features (embeddings) of entities and relations
- Can be very efficient (e.g. matrix multiplication for prediction)
- Has difficulty generalizing when graph has many small cliques

Path ranking methods (e.g., random walk) [e.g., Lao+ 2011]

- Assumes truth values of unknown facts are independent given observed facts
- Difficulty capturing dependencies through long relation paths
- Sparsity when number of relation types is large

Hybrid of path ranking and embedding methods

[Poon+ 2017]

WideOpen: "Make Public Data Public"

NLP: Automate detection of overdue datasets

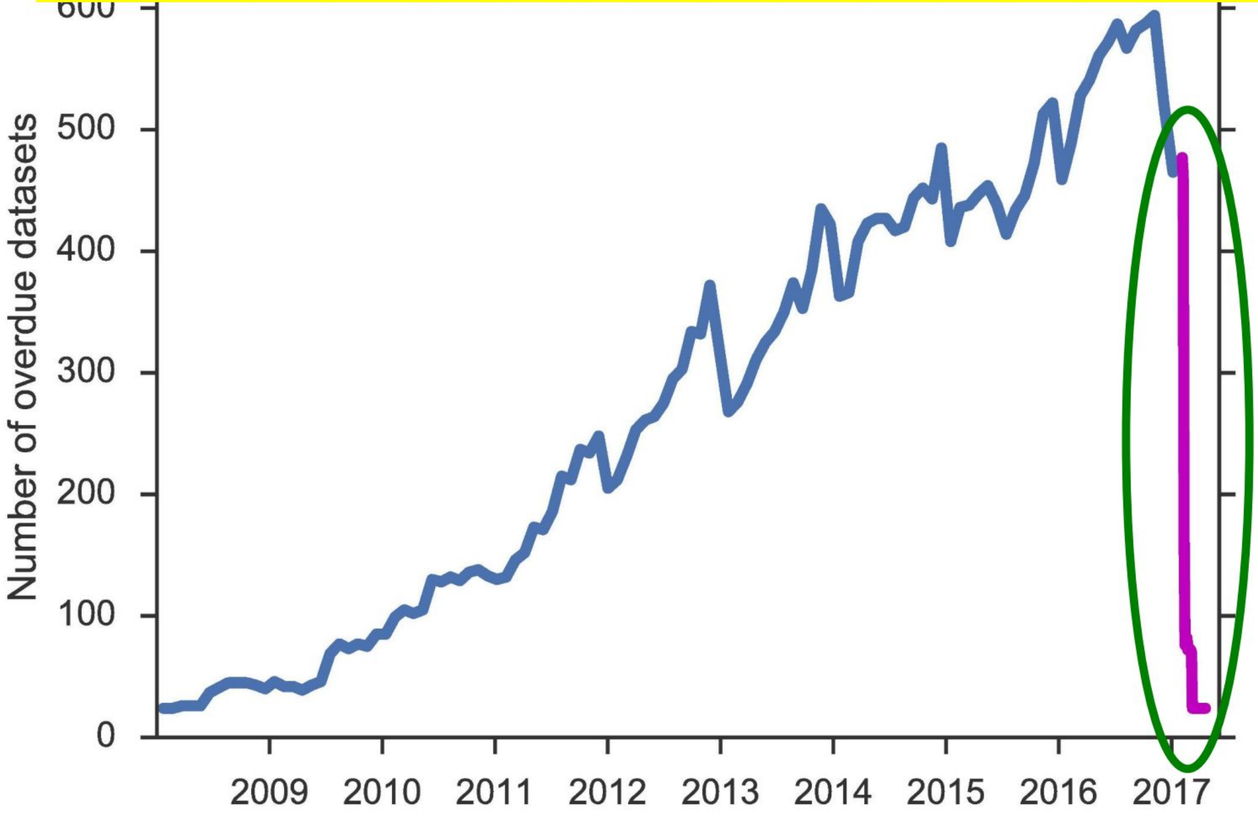
PubMed: Identify dataset mentions

Repo: Parse query output to determine if overdue

Grechkin et al. "Wide-Open: accelerating public data release by automating detection of overdue datasets". *PLOS Biology, 2017*.

[Poon+ 2017]

Enabled GEO to release 400 datasets in a week

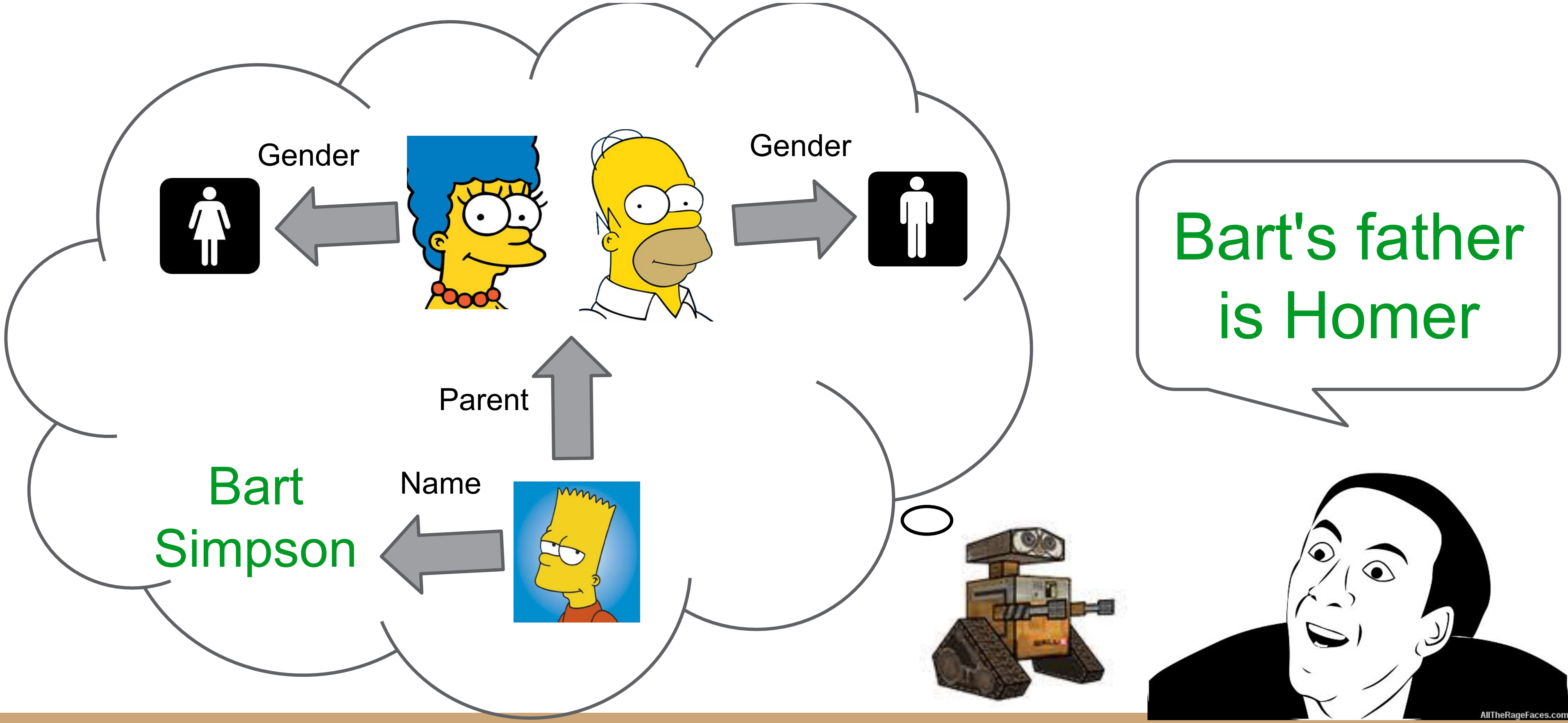


Plan

- Information extraction
- **Semantic parsing**
- Semantic representation

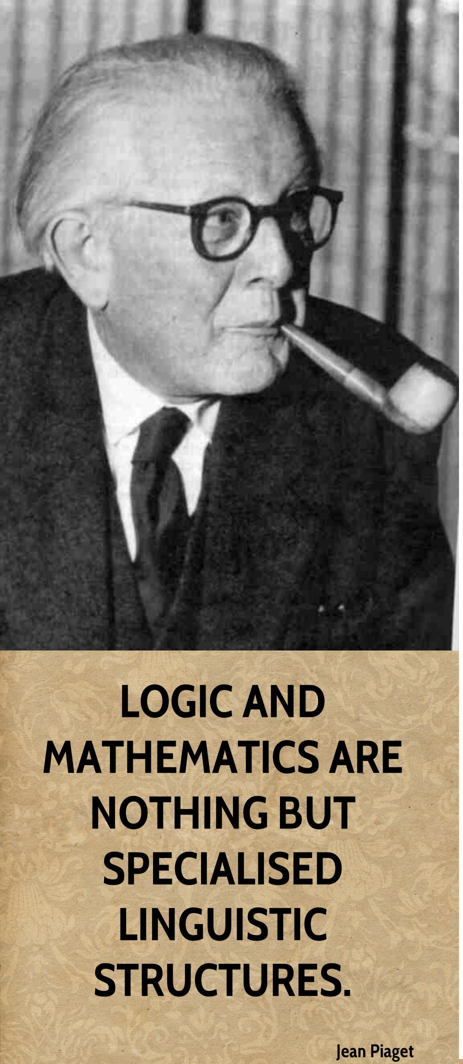
[Liang+ 2017]

When reasoning is needed to understand text



[Liang+ 2017]

Language, Translation & Control



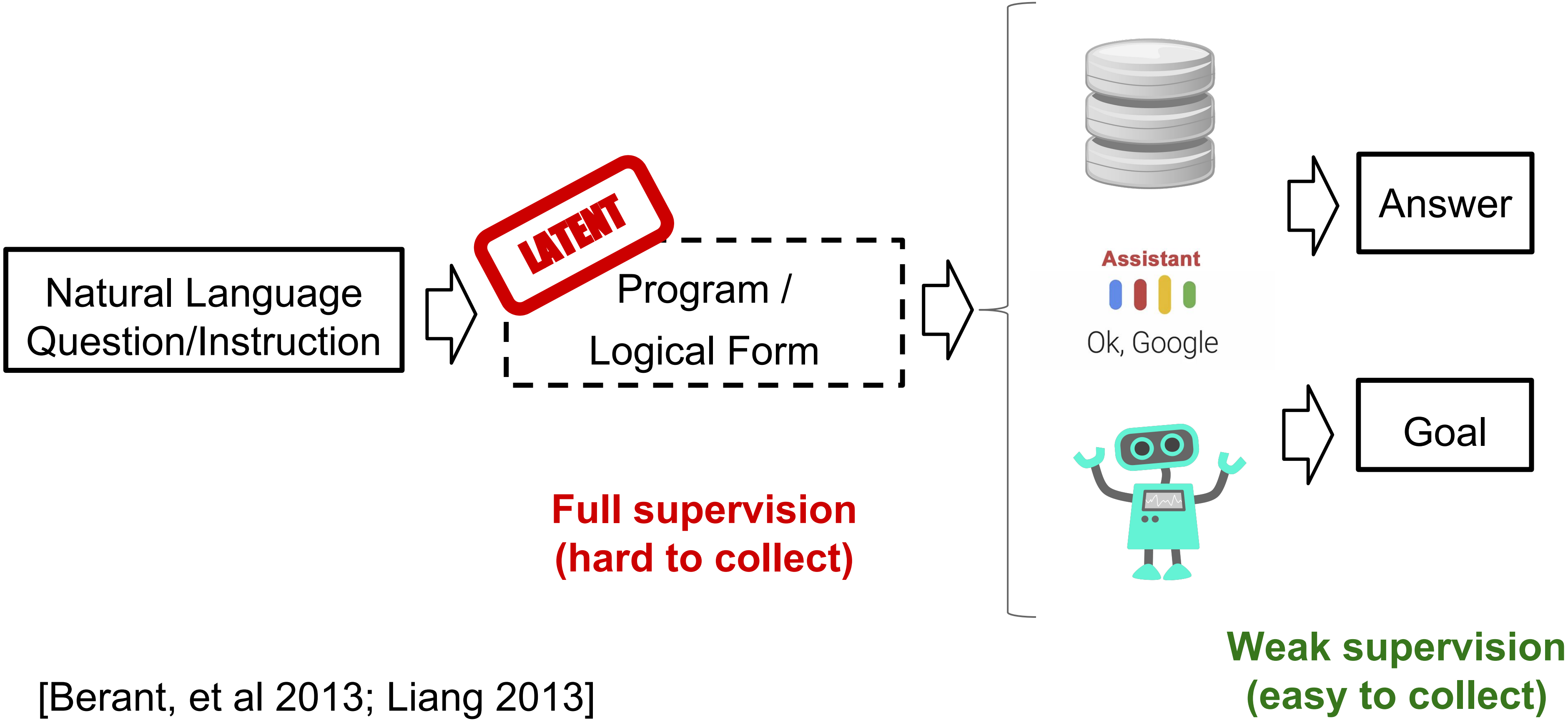
1) Natural languages are programming languages to control human behavior

2) For machines to understand natural languages, they just need a translation model, which converts questions (statements) to programs

3) The programs find answers when "executed" against KB

[Liang+ 2017]

Semantic Parsing: Language to Programs



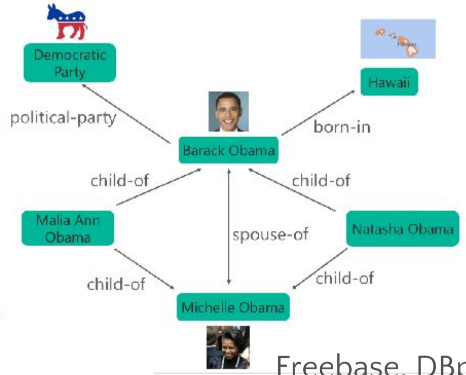
[Berant, et al 2013; Liang 2013]

[Liang+ 2017]

Question Answering with Knowledge Base

Largest city in US?

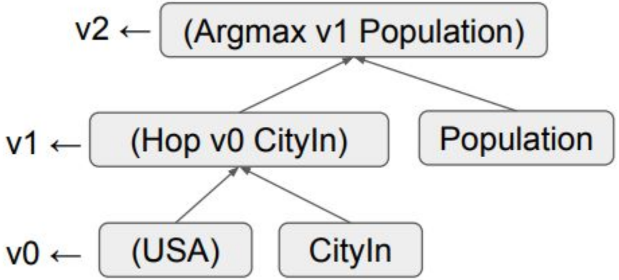
```
GO
(Hop V1 CityIn)
(Argmax V2 Population)
RETURN
```



Freebase, DBpedia, YAGO, NELL

NYC

1. Compositionality



2. Large Search Space

Freebase:
23K predicates, 82M entities, 417M triplets

[Liang+ 2017]

WebQuestionsSP Dataset

- 5,810 questions Google Suggest API & Amazon MTurk[1]
- Remove invalid QA pairs[2]
- 3,098 training examples, 1,639 testing examples remaining
- Open-domain, and contains grammatical error
- Multiple entities as answer => macro-averaged F1

Grammatical error

Multiple entities

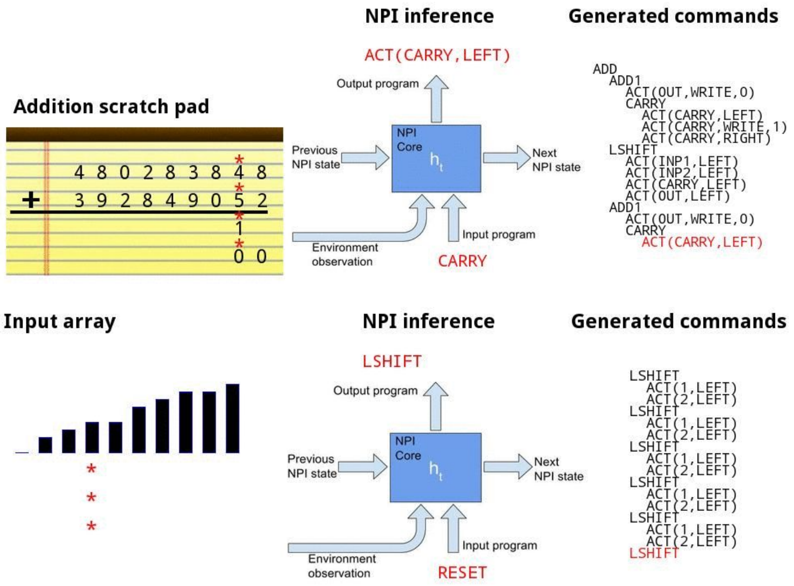
• What do Michelle Obama do for a living?	writer, lawyer
• What character did Natalie Portman play in Star Wars?	Padme Amidala
• What currency do you use in Costa Rica?	Costa Rican colon
• What did Obama study in school?	political science
• What killed Sammy Davis Jr?	throat cancer

[Berant et al, 2013; Yih et al, 2016]

[Liang+ 2017]

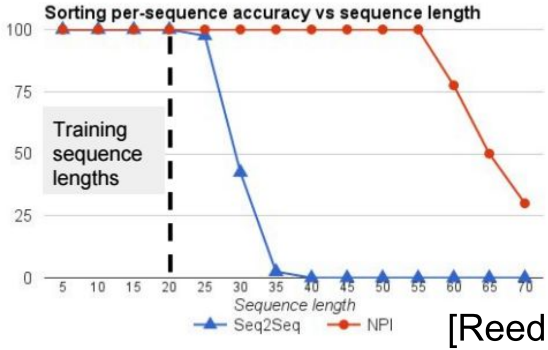
(Scalable) Neural Program Induction

- Impressive works to show NN can learn addition and sorting, but...



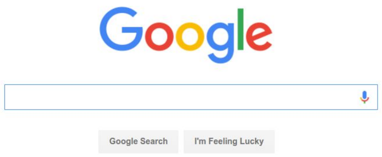
[Zaremba & Sutskever 2016]

- The learned operations are not as scalable and precise.



[Reed & Freitas 2015]

- Why not use existing modules that are scalable, precise and interpretable?



Connectionism + Symbolism

[Liang+ 2017]

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, the connectionism approaches starts to win

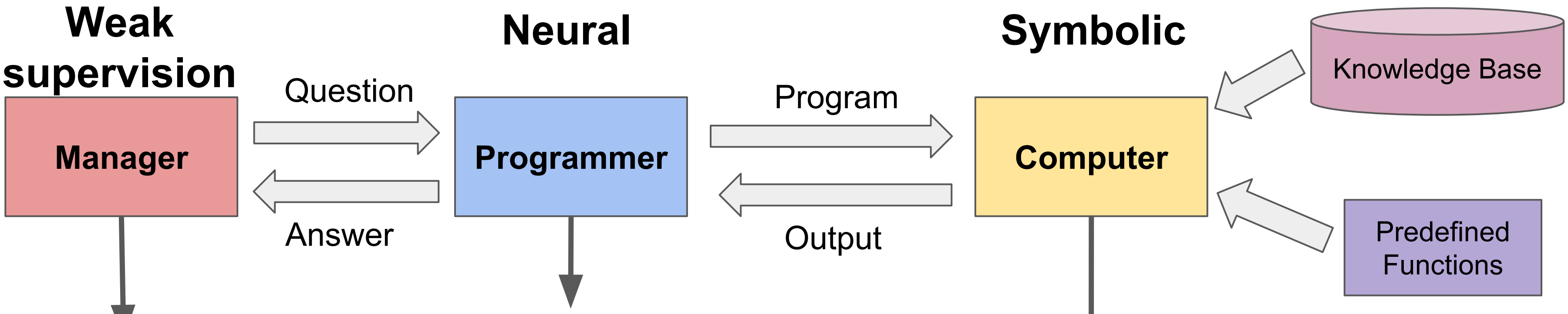
a symbolic machine

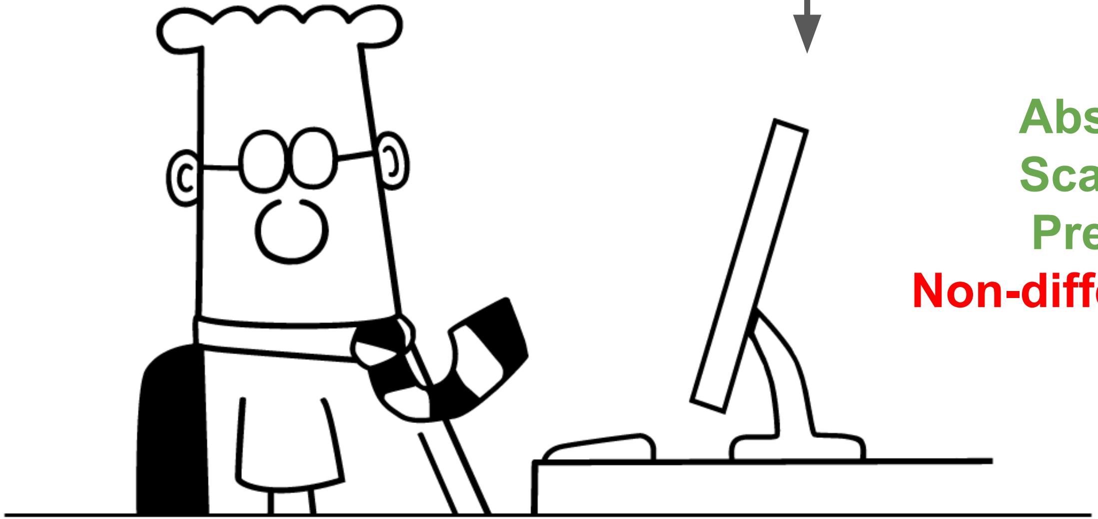
a neural controller

a sequence of symbols

Neural Symbolic Machines

[Liang+ 2017]

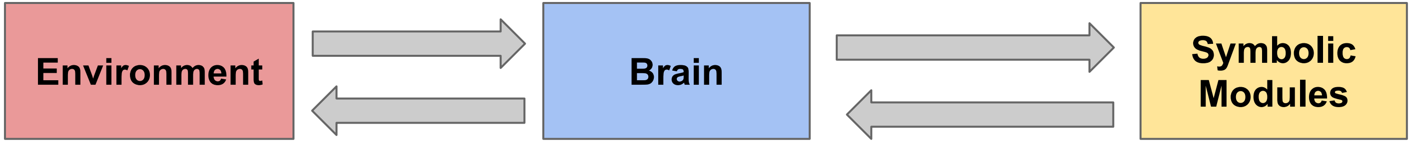


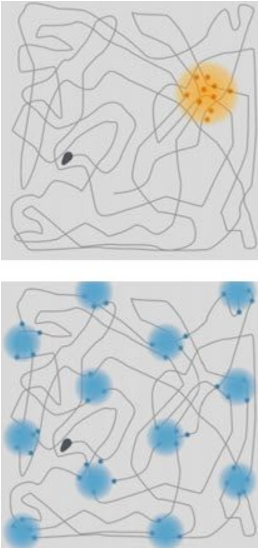


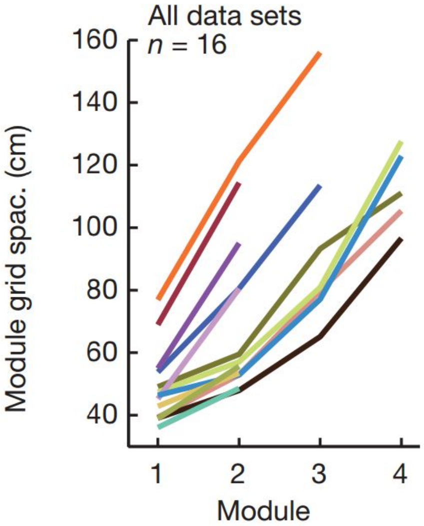
Abstract
Scalable
Precise
Non-differentiable

[Stensola+ 2012]

Symbolic Machines in Brains

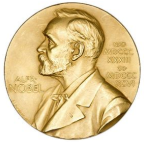






Mean grid spacing for all modules (M1–M4) in all animals (colour-coded)

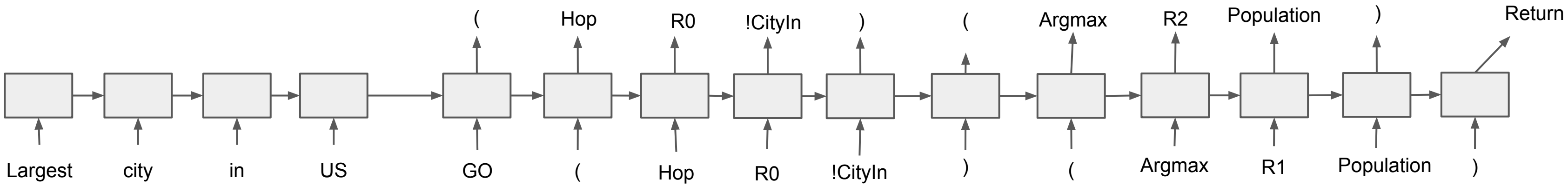
- 2014 Nobel Prize in Physiology or Medicine awarded for ‘inner GPS’ research



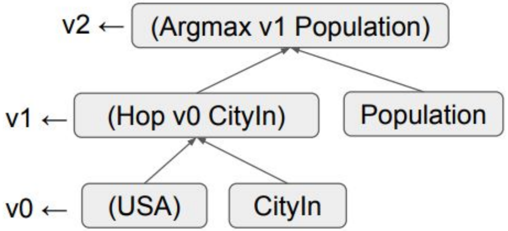
- Positions are represented as discrete numbers in animals' brains, which enable accurate and autonomous calculations

Simple Seq2Seq model is not enough

[Liang+ 2017]



1. Compositionality



2. Large Search Space

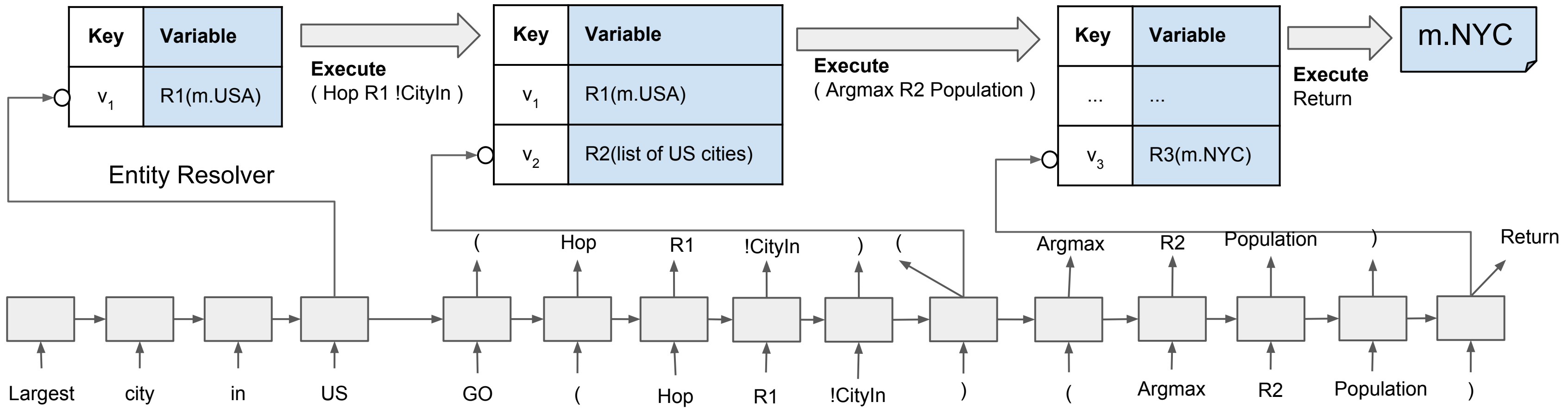
23K predicates,
82M entities,
417M triplets

1.Key-Variable Memory

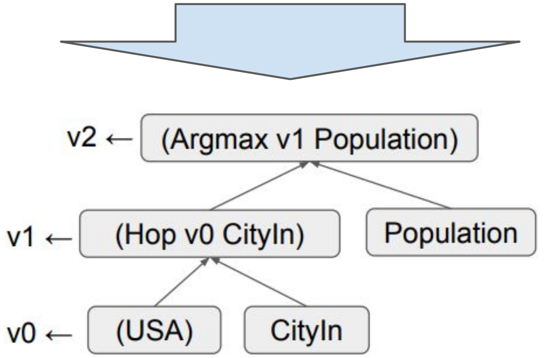
2.Code Assistance
3.Augmented REINFORCE

[Liang+ 2017]

Key-Variable Memory for Compositionality



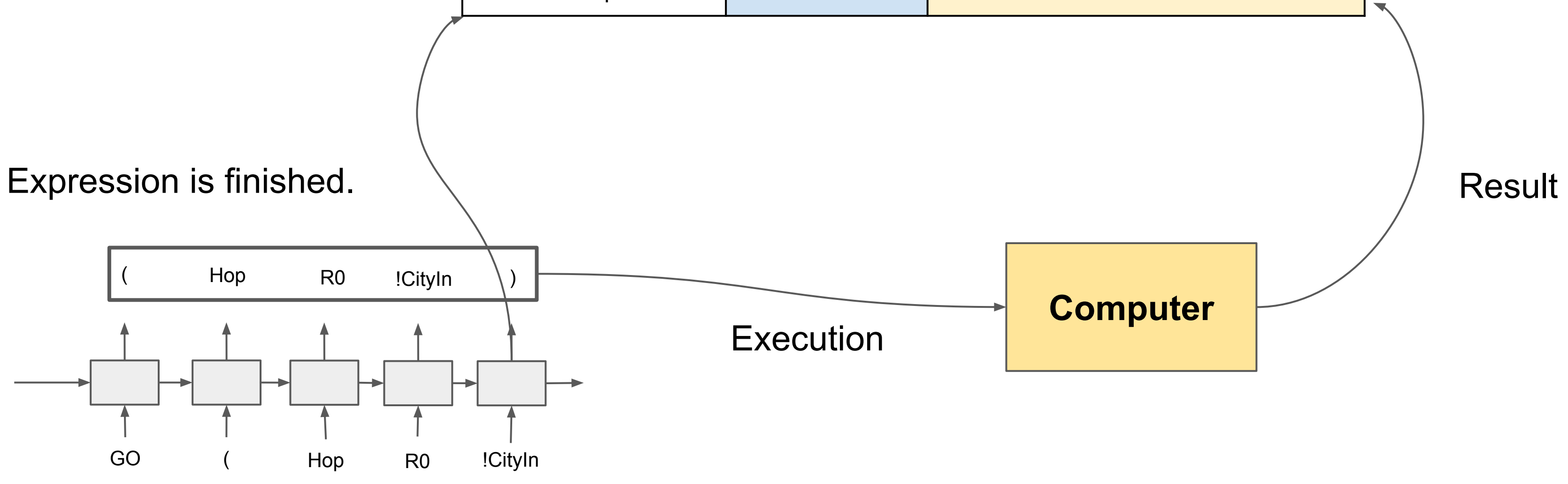
- A linearised bottom-up derivation of the recursive program.



[Liang+ 2017]

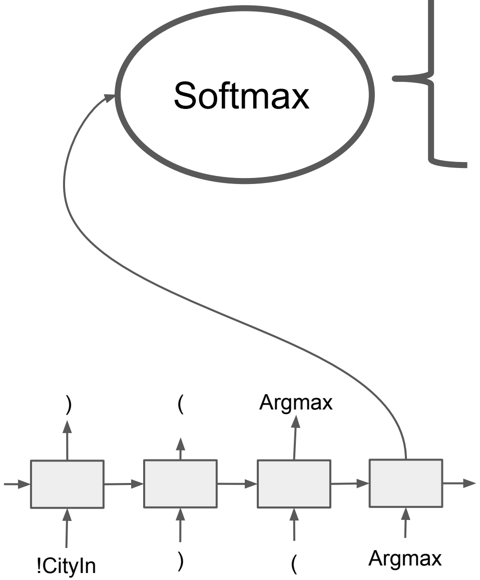
Key-Variable Memory: Save Intermediate Value

Key (Embedding)	**Variable** (Symbol)	**Value** (Data in Computer)
V_0	`R0`	`m.USA`
V_1	`R1`	`[m.SF, m.NYC, ...]`

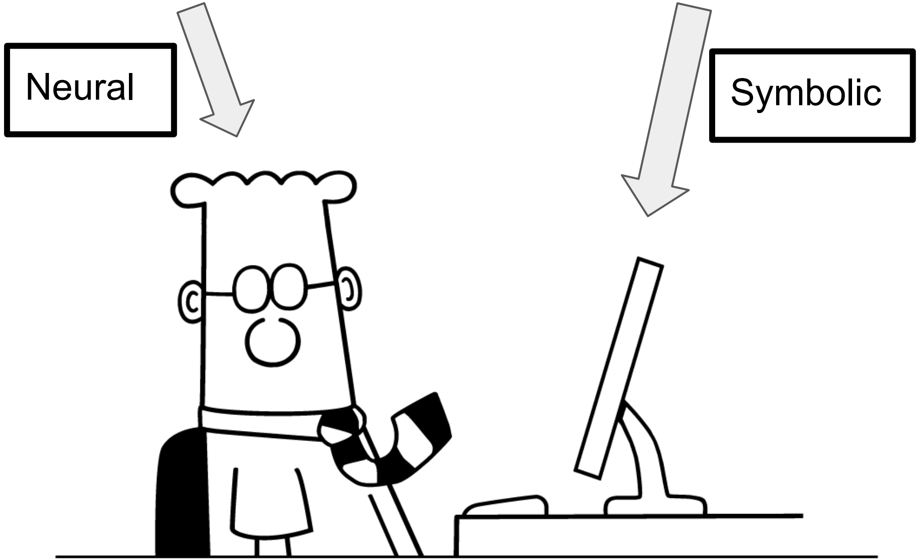


[Liang+ 2017]

Key-Variable Memory: Reuse Intermediate Value



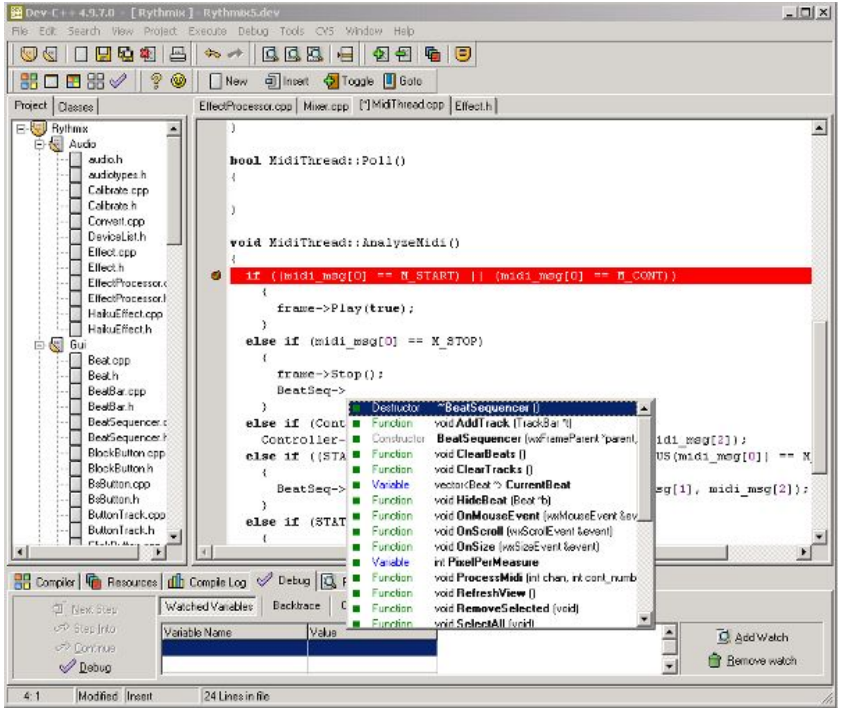
Key (Embedding)	**Variable** (Symbol)	**Value** (Data in Computer)
V_0	`R0`	`m.USA`
V_1	`R1`	`[m.SF, m.NYC, ...]`



Code Assistance: Prune Search Space

[Liang+ 2017]

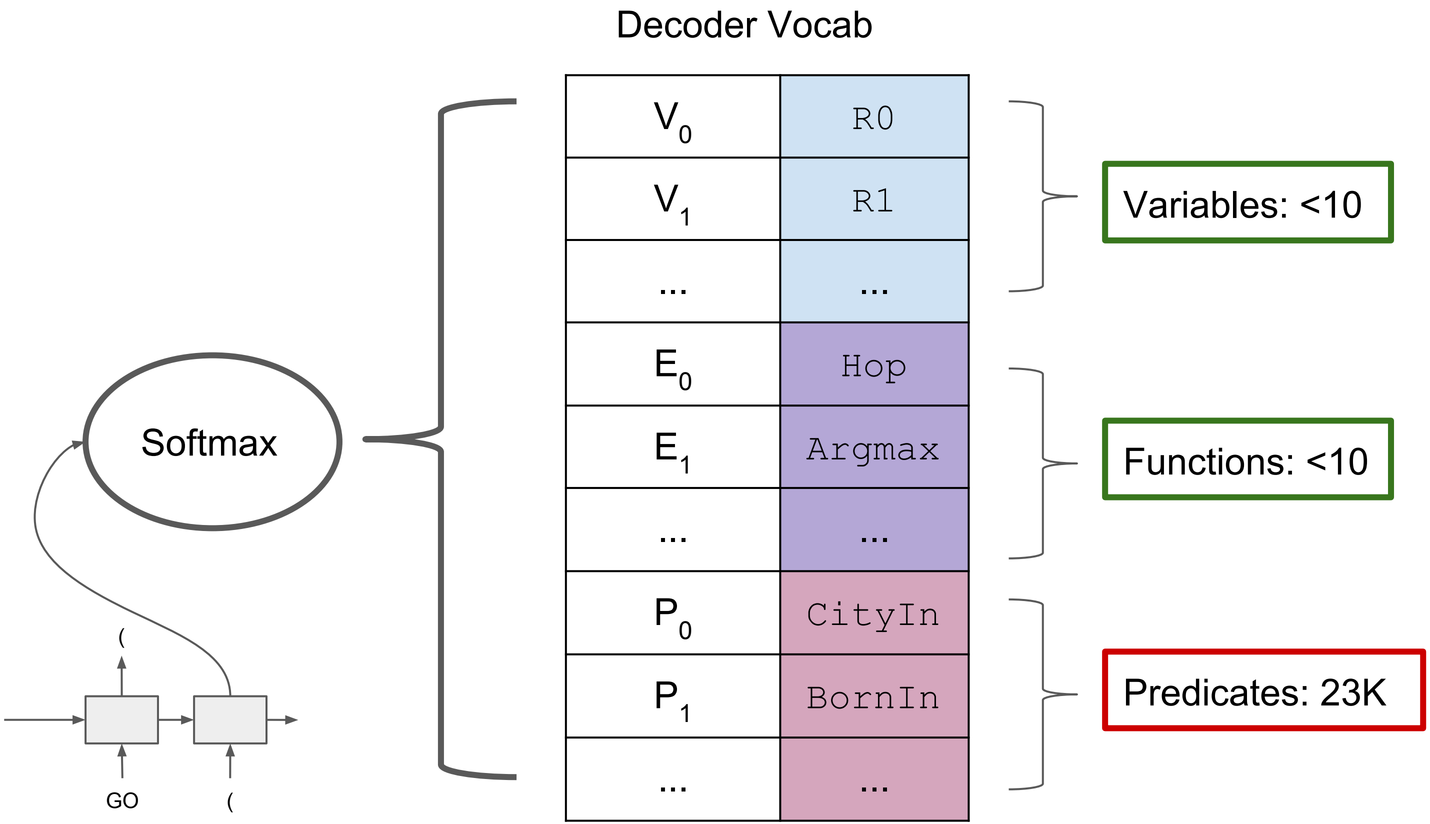
Pen and paper



IDE

[Liang+ 2017]

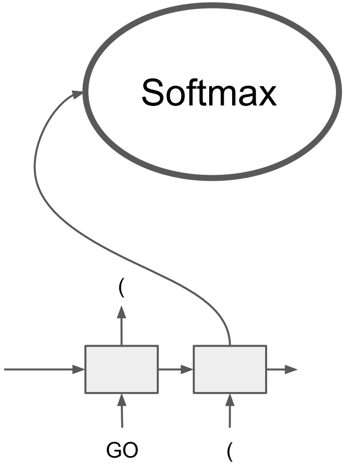
Code Assistance: Syntactic Constraint

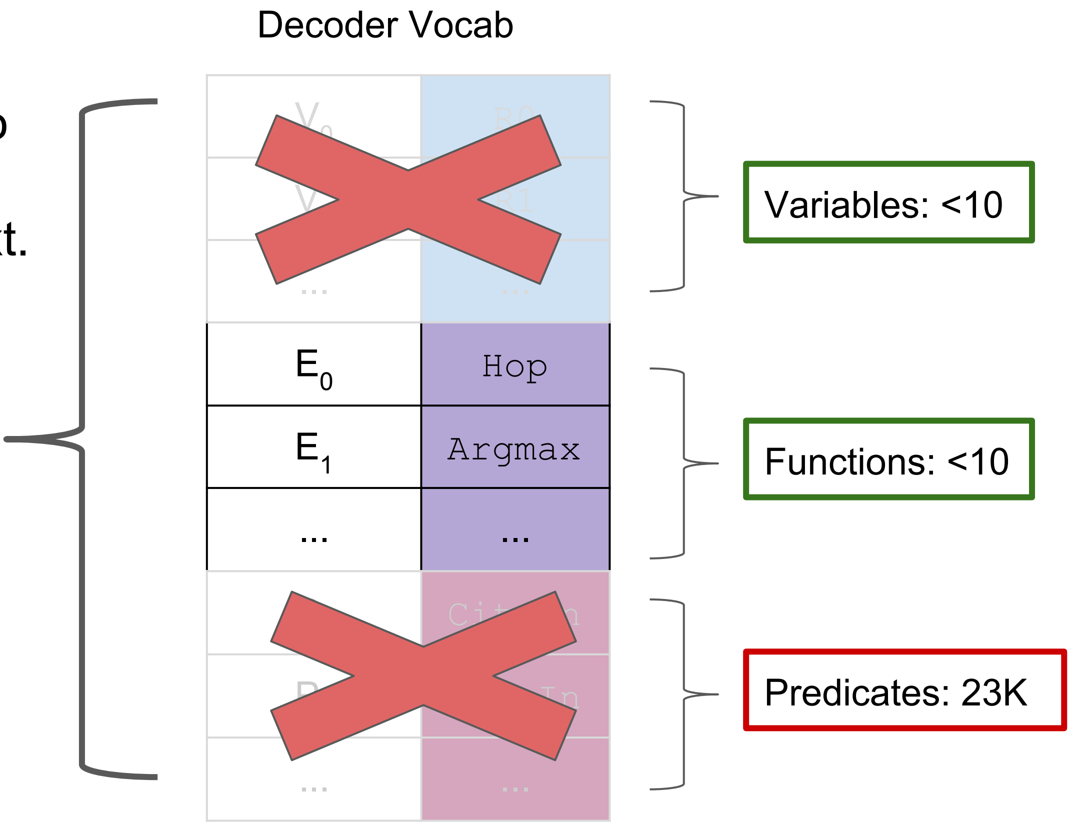


[Liang+ 2017]

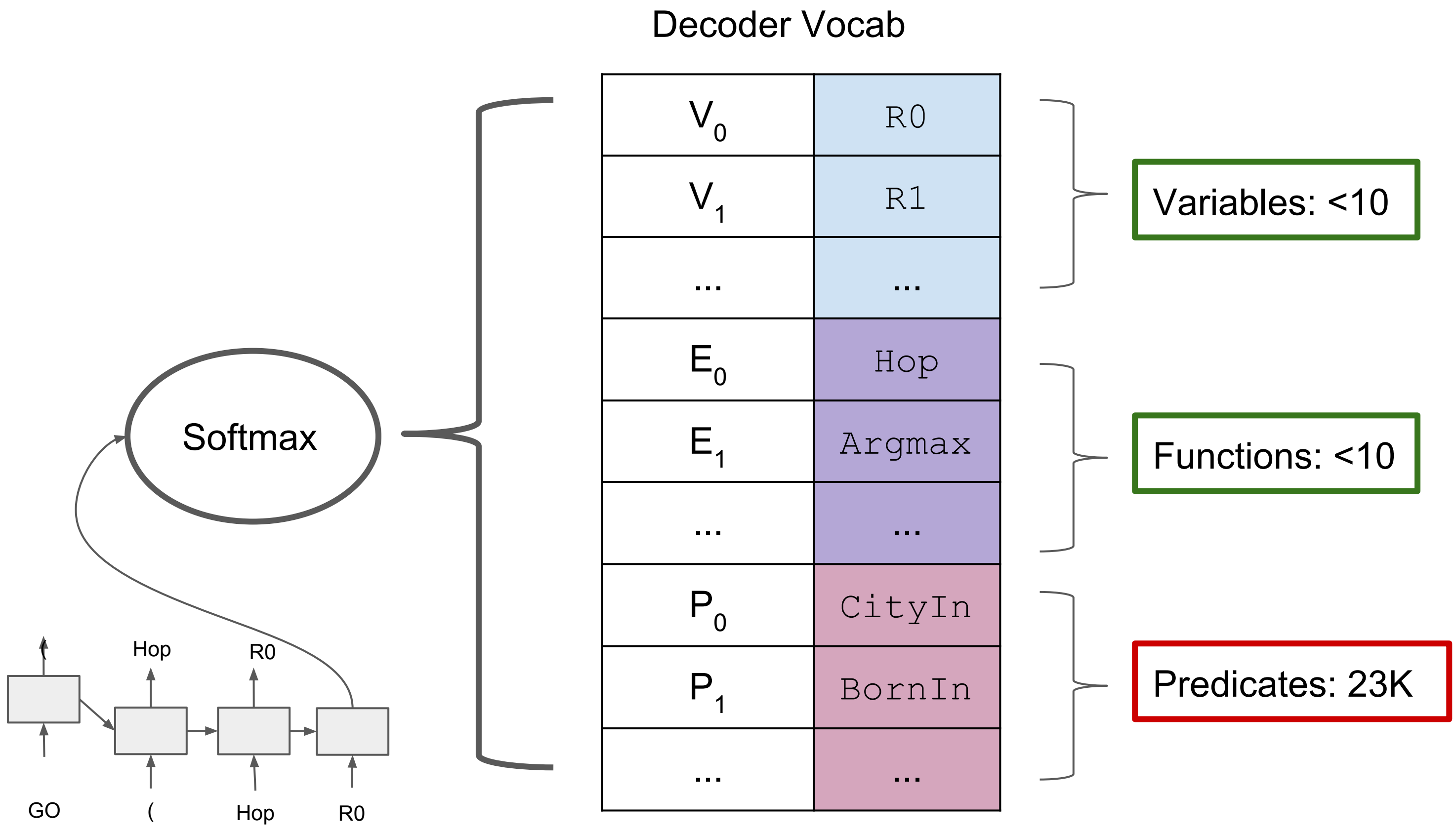
Code Assistance: Syntactic Constraint

Last token is ‘(’, so has to output a function name next.



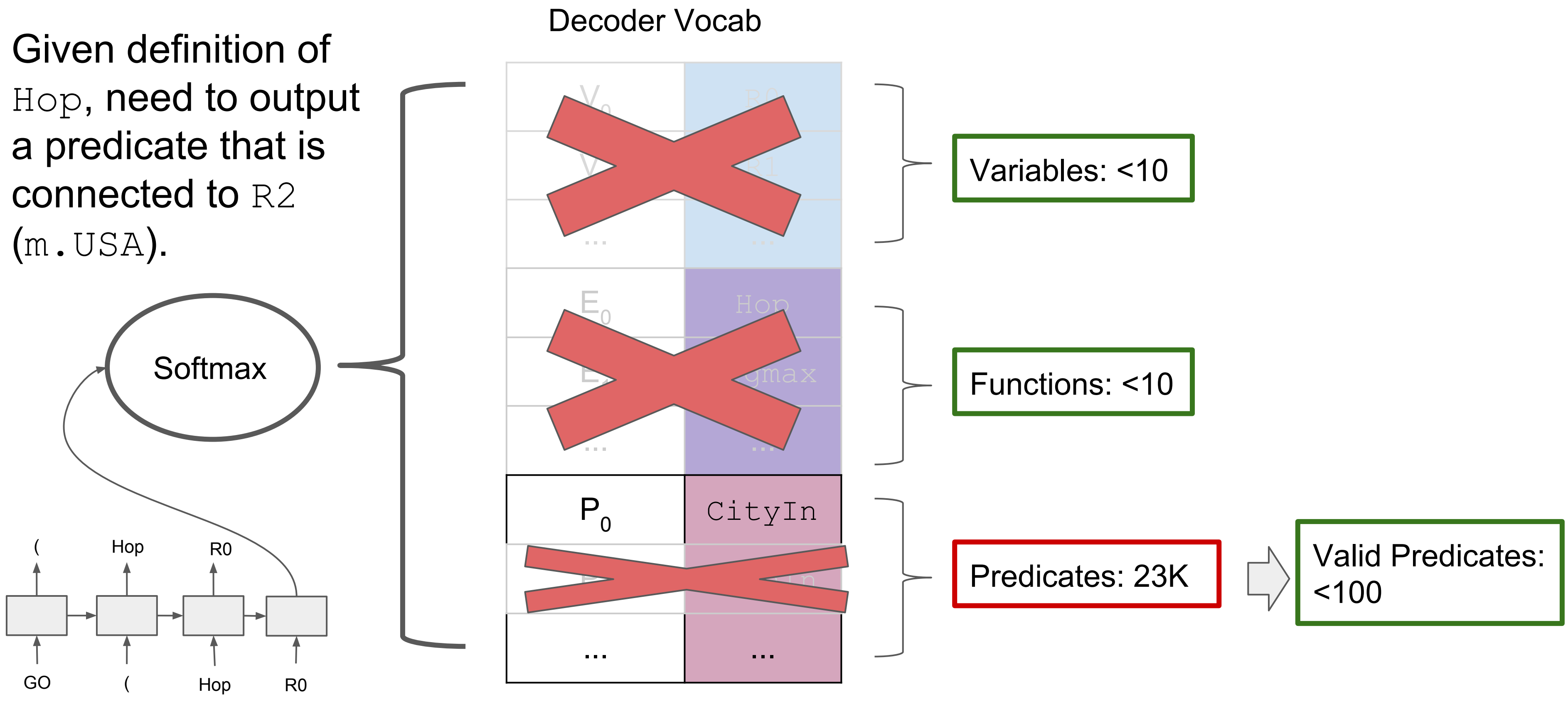


Code Assistance: Semantic Constraint



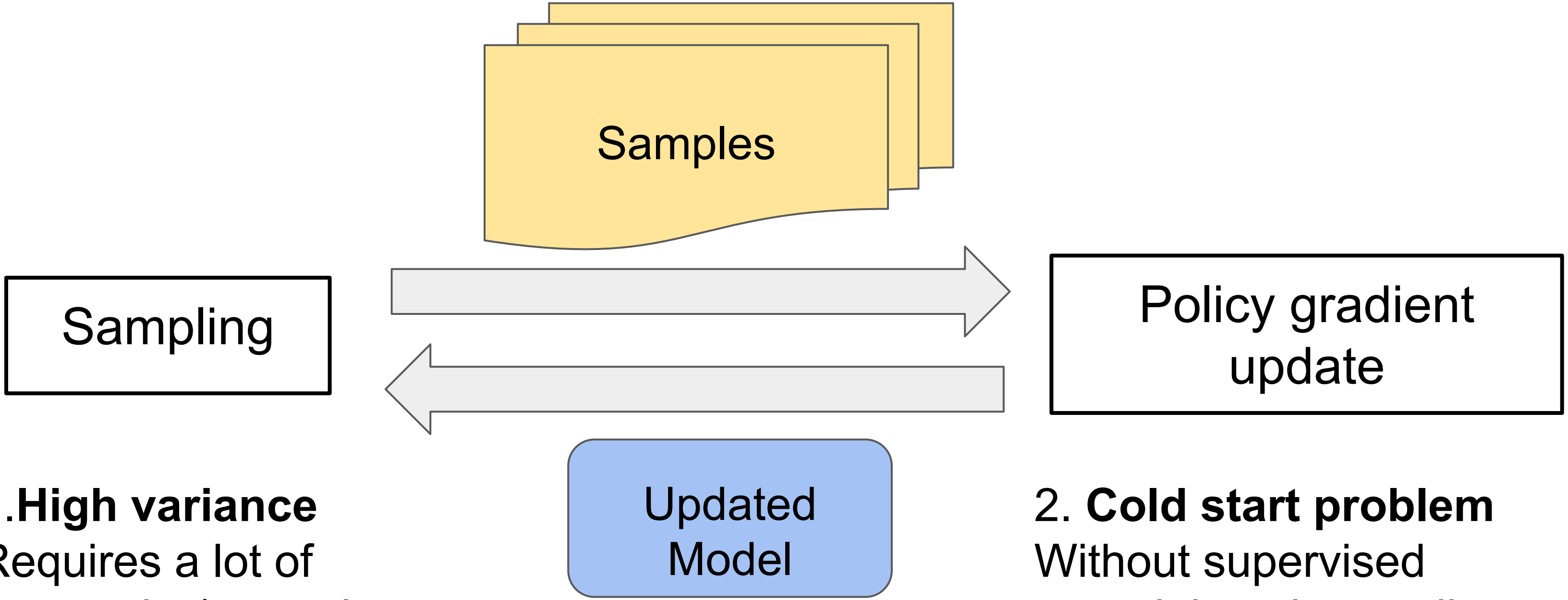
Code Assistance: Semantic Constraint

Given definition of `Hop`, need to output a predicate that is connected to `R2` (`m.USA`).



[Liang+ 2017]

REINFORCE Training



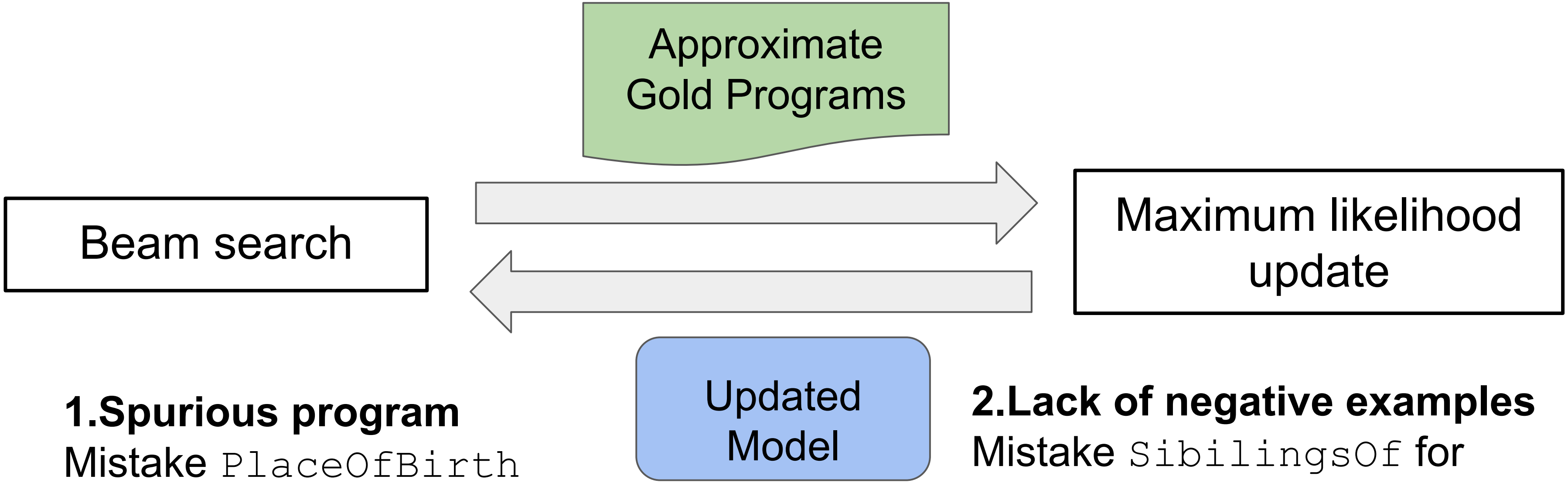
1.**High variance**
Requires a lot of (expensive) samples

2. **Cold start problem**
Without supervised pretraining, the gradients at the beginning

$$\nabla_\theta J^{RL}(\theta) = \sum_q \sum_{a_{0:T}} \boxed{P(a_{0:T}|q,\theta)}[R(q,a_{0:T}) - B(q)]\nabla_\theta \log P(a_{0:T}|q,\theta)$$

[Liang+ 2017]

Iterative Maximum Likelihood Training (Hard EM)

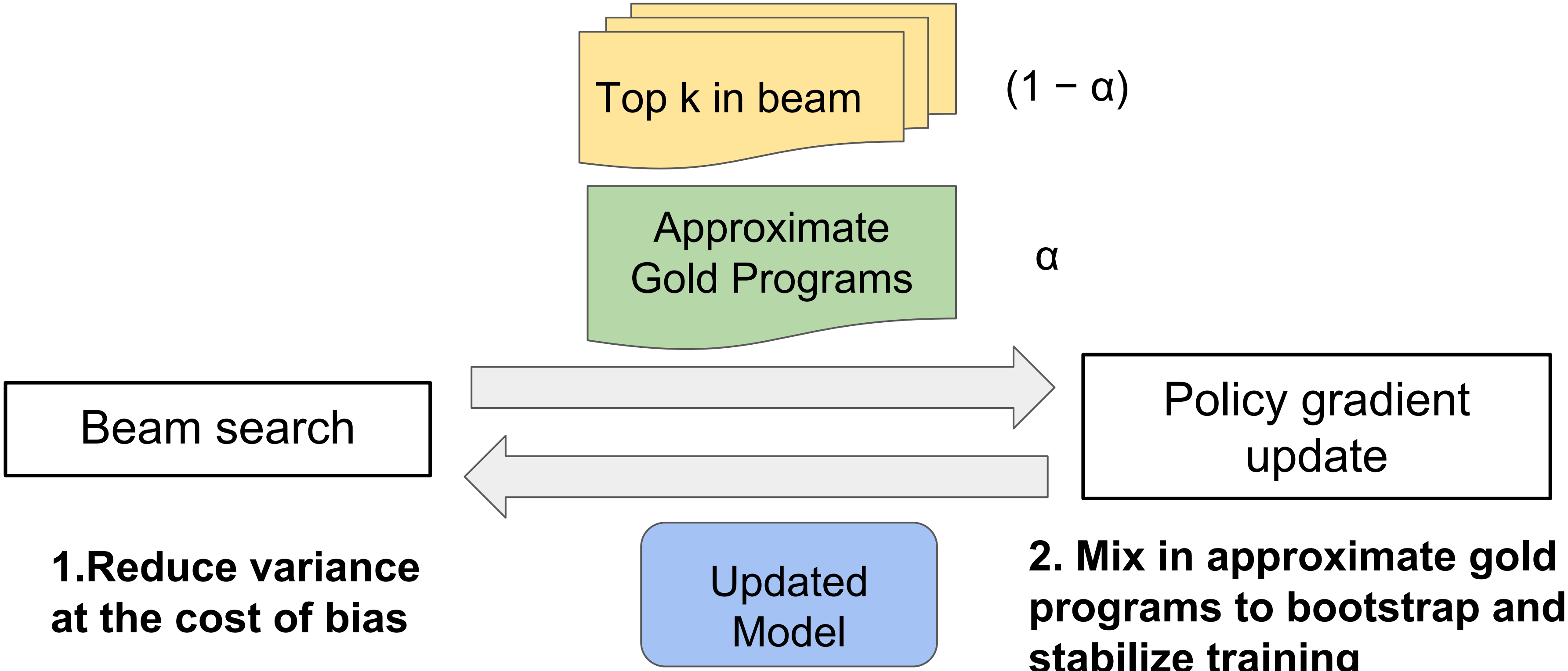


1.Spurious program
Mistake `PlaceOfBirth` for `PlaceOfDeath`.

2.Lack of negative examples
Mistake `SiblingsOf` for `ParentsOf`.

$$J^{ML}(\theta) = \sum_{q} \log P(a_{0:T}^{best}(q)|q,\theta)$$

Augmented REINFORCE



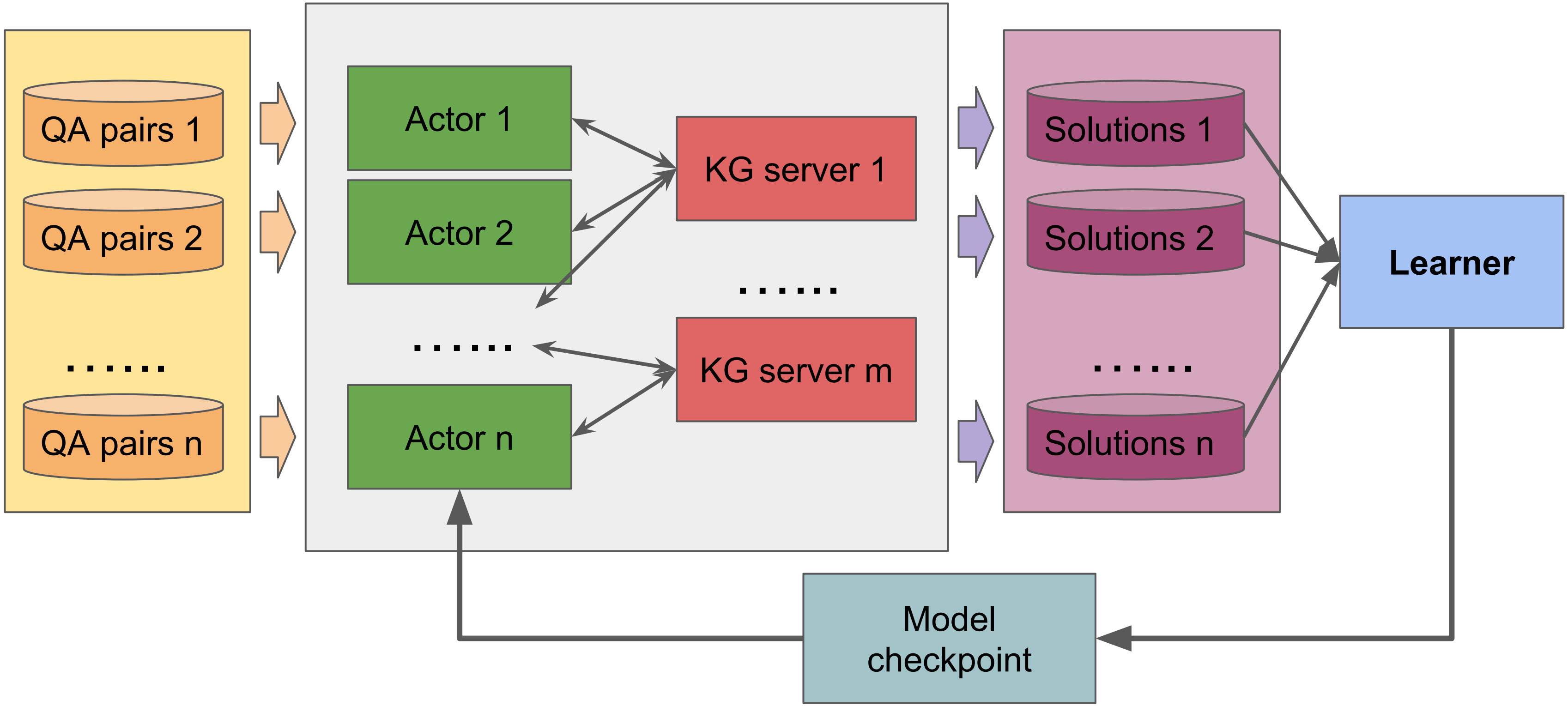
1.Reduce variance at the cost of bias

2. Mix in approximate gold programs to bootstrap and stabilize training

Distributed Architecture

[Liang+ 2017]

- 200 actors, 1 learner, 50 Knowledge Graph servers



[Liang+ 2017]

Generated Programs

- **Question**: “what college did russell wilson go to?”
- **Generated program**:

```
(hop v1 /people/person/education)
(hop v2 /education/education/institution)
(filter v3 v0 /common/topic/notable_types )
<EOP>
```

In which

```
v0 = “College/University” (m.01y2hnl)
v1 = “Russell Wilson” (m.05c10yf)
```

- **Distribution of the length of generated programs**

#Expressions	0	1	2	3
Percentage	0.4%	62.9%	29.8%	6.9%
F1	0.0	73.5	59.9	70.3

[Liang+ 2017]

New State-of-the-Art on ***WebQuestionsSP***

- First end-to-end neural network to achieve SOTA on semantic parsing with weak supervision over large knowledge base
- The performance is approaching SOTA with full supervision

Model	Avg. Prec.@1	Avg. Rec.@1	Avg. F1@1	Acc.@1
STAGG	67.3	73.1	66.8	58.8
NSM – our model	70.8	76.0	**69.0**	59.5
STAGG (full supervision)	70.9	80.3	71.7	63.9

[Liang+ 2017]

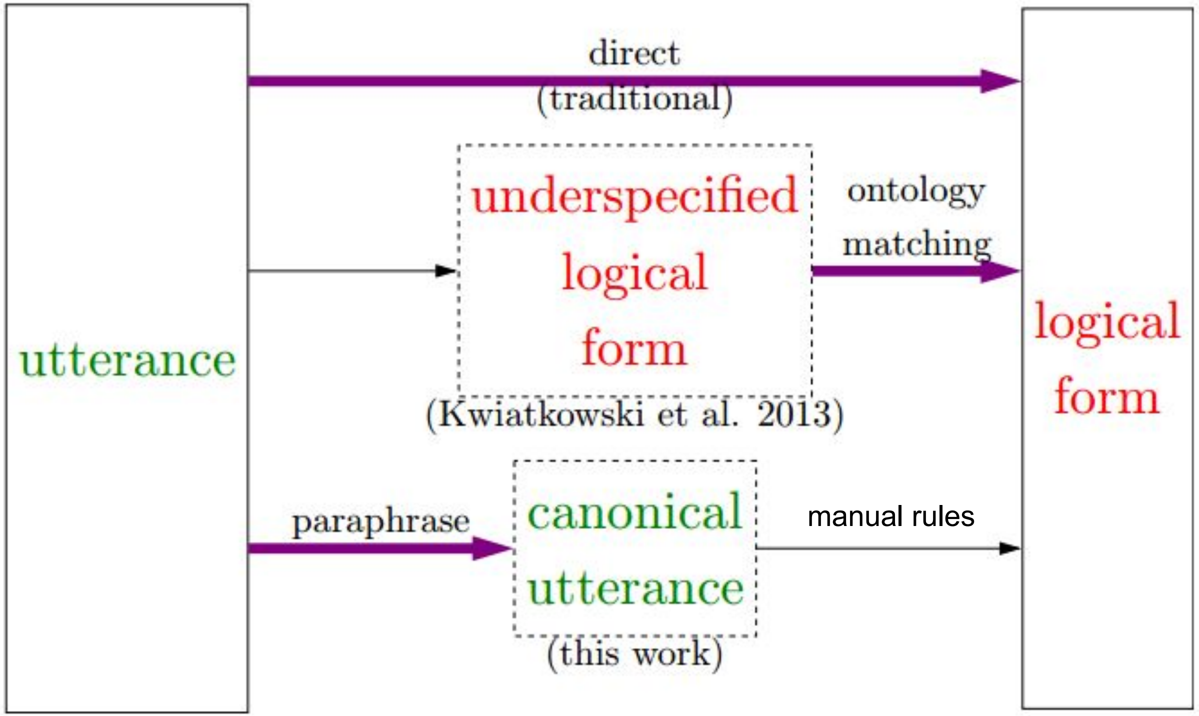
Augmented REINFORCE

- REINFORCE get stuck at local maxima
- Iterative ML training is not directly optimizing the F1 score
- Augmented REINFORCE obtains the best performances

Settings	Train Avg. F1@1	Valid Avg. F1@1
iterative ML only	68.6	60.1
REINFORCE only	55.1	47.8
Augmented REINFORCE	83.0	**67.2**

[Berant & Liang 2014]

From open IE to matching problems



"what is Italy money"
"latest meeting"

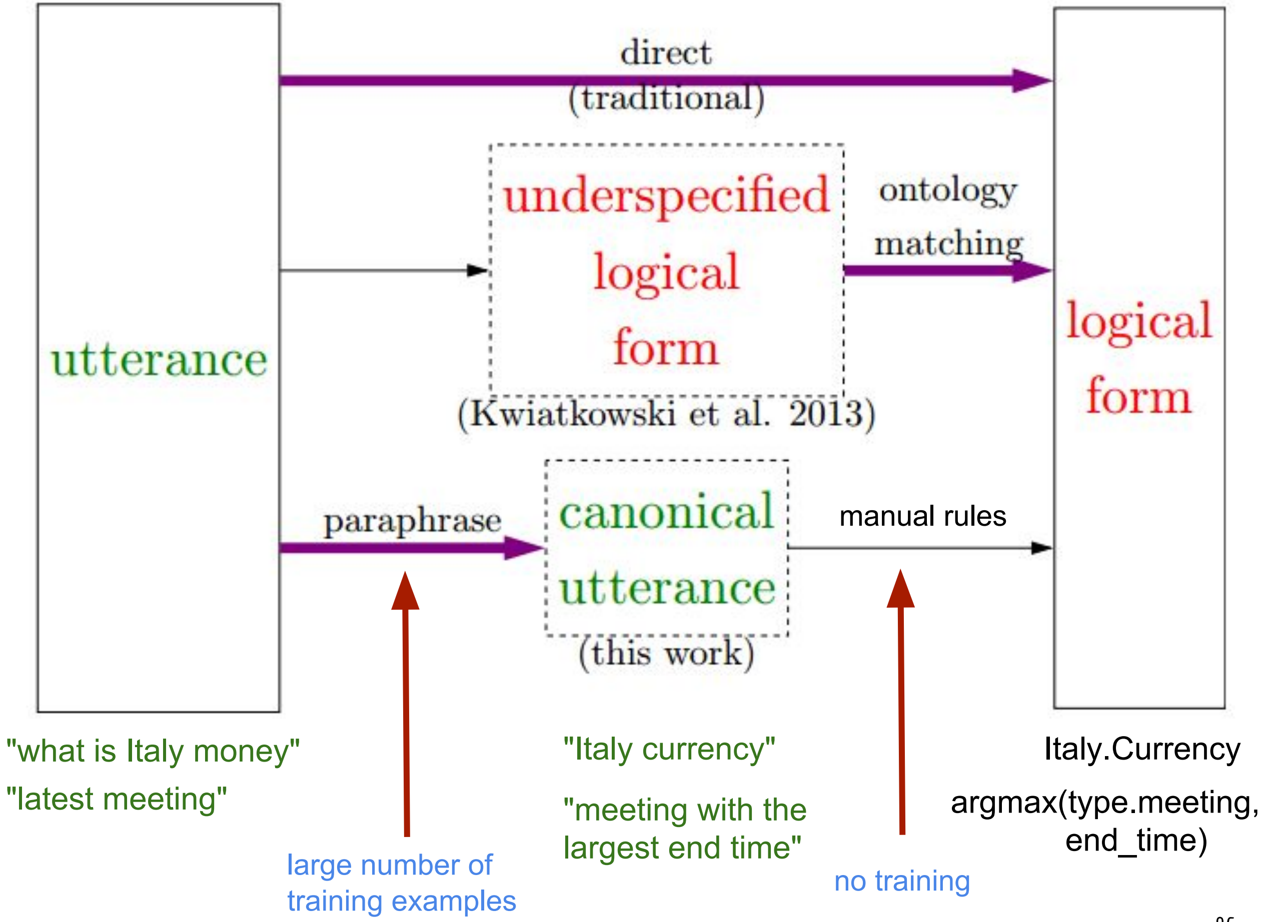
"Italy currency"
"meeting with the largest end time"

Italy.Currency
argmax(type.meeting, end_time)

[Berant & Liang 2014]

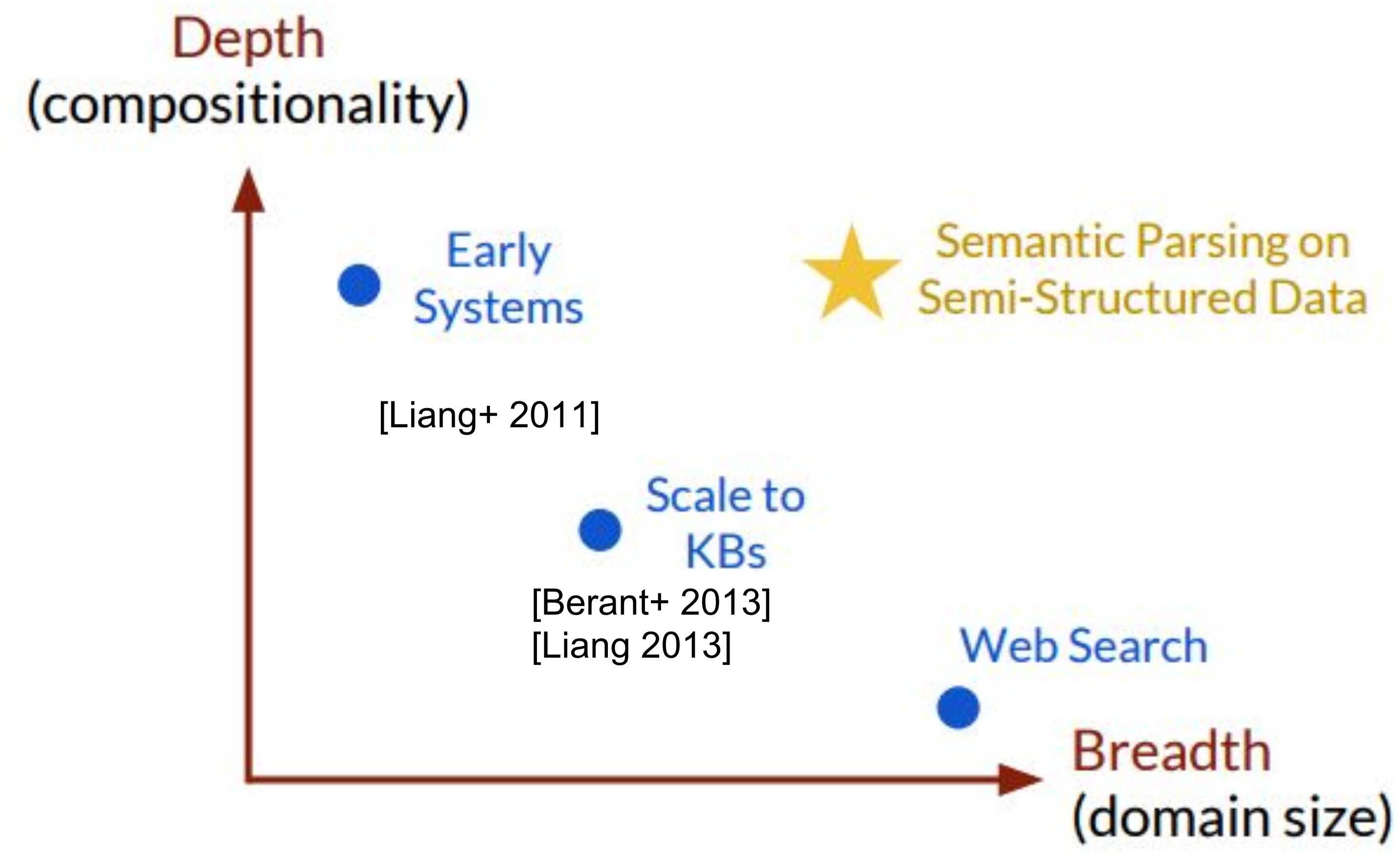
From open IE to matching problems

The beauty of the proposed approach



The Web as a KB

[Pasupat & Liang 2015]



[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 / ...]

[Pasupat & Liang 2015]

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 / ...]

[Pasupat & Liang 2015]

The Web as a KB

QA on semi-structured data

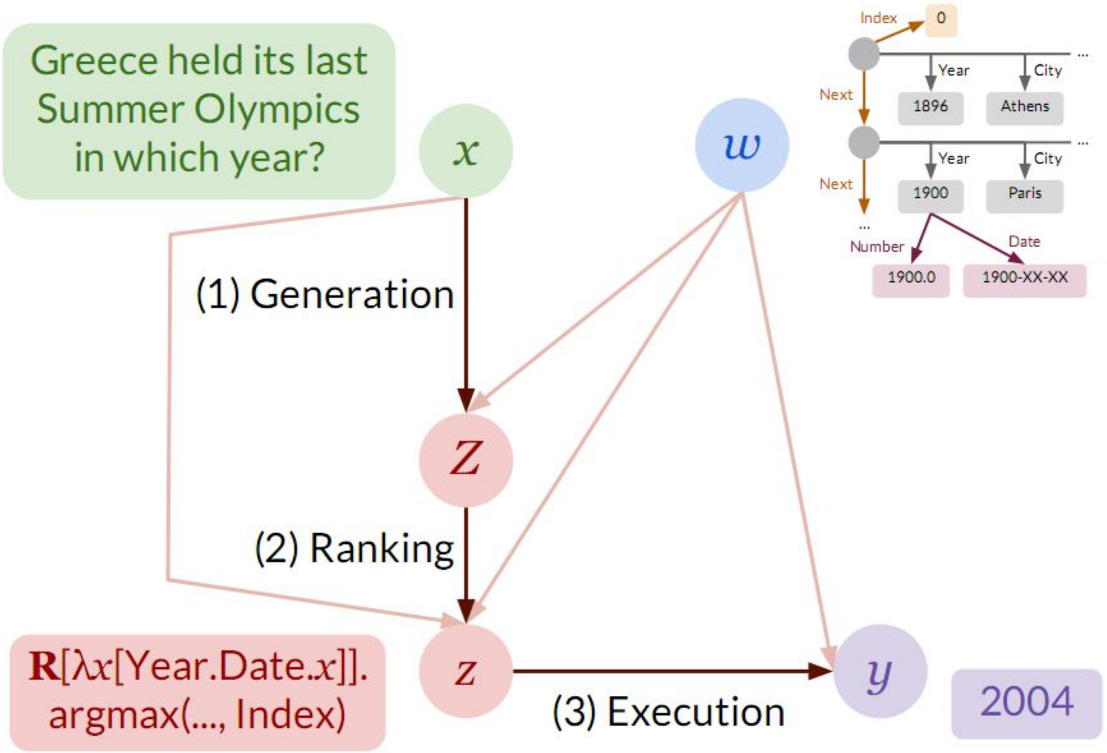
Input: utterance x and HTML table t
Output: answer y
Training data: list of (x, t, y) — no logical form
WikiTableQuestions dataset:

- 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?

[Pasupat & Liang 2015]

The Web as a KB



Tables are represented as graphs

[Iyer+ UW]

Learning a Neural Semantic Parser from User Feedback

- neural sequence models to map utterances directly to SQL, bypassing any intermediate meaning representations

- These models are immediately deployed online to solicit feedback from real users to flag incorrect queries.

Most recent papers of Michael I. Jordan

```
SELECT paper.paperId, paper.year
FROM paper, writes, author
WHERE paper.paperId = writes.paperId
  AND writes.authorId = author.authorId
  AND author.authorName = "michael i. jordan"
  AND paper.year =
    (SELECT max(paper.year)
     FROM paper, writes, author
     WHERE paper.paperId = writes.paperId
      AND writes.authorId = author.authorId
      AND author.authorName = "michael i. jordan");
```

I'd like to book a flight from San Diego to Toronto

```
SELECT DISTINCT f1.flight_id
FROM flight f1, airport_service a1, city c1,
  airport_service a2, city c2
WHERE f1.from_airport = a1.airport_code
  AND a1.city_code = c1.city_code
  AND c1.city_name = 'san diego'
  AND f1.to_airport = a2.airport_code
  AND a2.city_code = c2.city_code
  AND c2.city_name = 'toronto';
```

[Cheng+ edinburgh]

Learning Structured Natural Language Representations for Semantic Parsing

Sentence: *which states do not border texas*
Non-terminal symbols in buffer: *which, states, do, not, border*
Terminal symbols in buffer: *texas*

Stack	Action	NT choice	TER choice
	NT	*answer*	
answer (	NT	*exclude*	
answer (*exclude* (	NT	*states*	
answer (*exclude* (*states* (	TER		*all*
answer (*exclude* (*states* (*all*	RED		
answer (*exclude* (*states* (*all*)	NT	*border*	
answer (*exclude* (*states* (*all*) , *border* (	TER		*texas*
answer (*exclude* (*states* (*all*) , *border* (*texas*	RED		
answer (*exclude* (*states* (*all*) , *border* (*texas*)	RED		
answer (*exclude* (*states* (*all*) , *border* (*texas*))	RED		
answer (*exclude* (*states* (*all*) , *border* (*texas*)))			

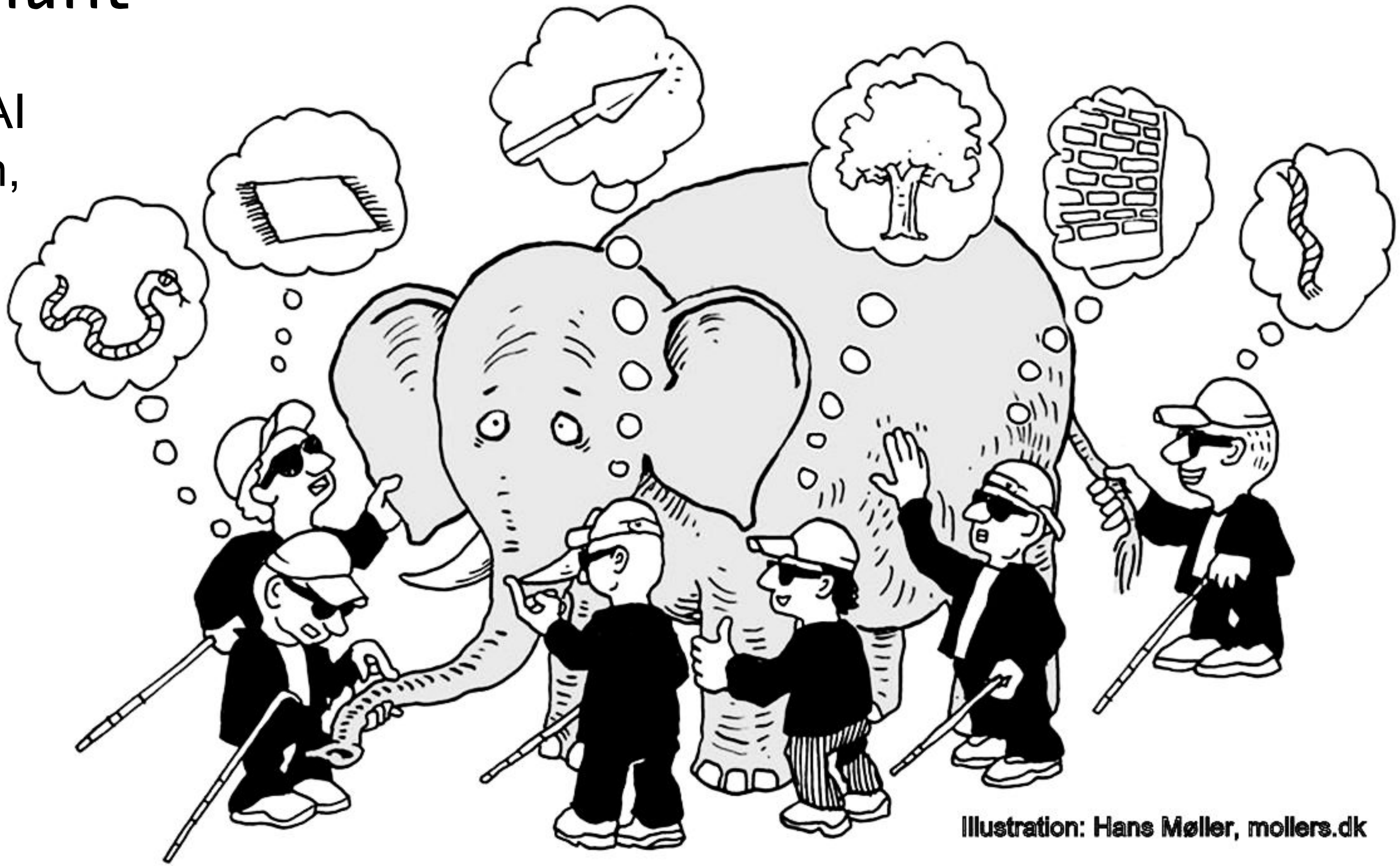
Table 2: Actions taken by the transition system for generating the ungrounded meaning representation of the example utterance. Symbols in red indicate domain-general predicates.

Plan

- Information extraction
- Semantic parsing
- **Semantic representation**

The AI Elephant

- Each subfield of AI holds certain truth, but not all of it



Putting things together

DL

function approximation

RL

correct training

NLP

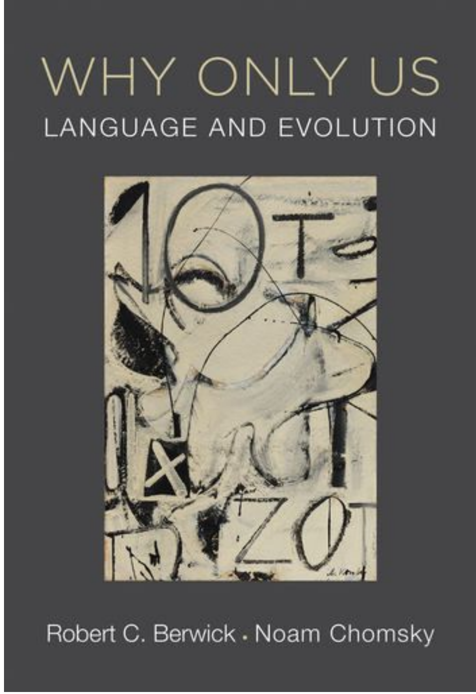
structural bias

AlphaGo

?

Language & reasoning

- Language was primarily invented for reasoning
- Communication comes later



[Lapata ACL talk 2017]

Lapata's scream

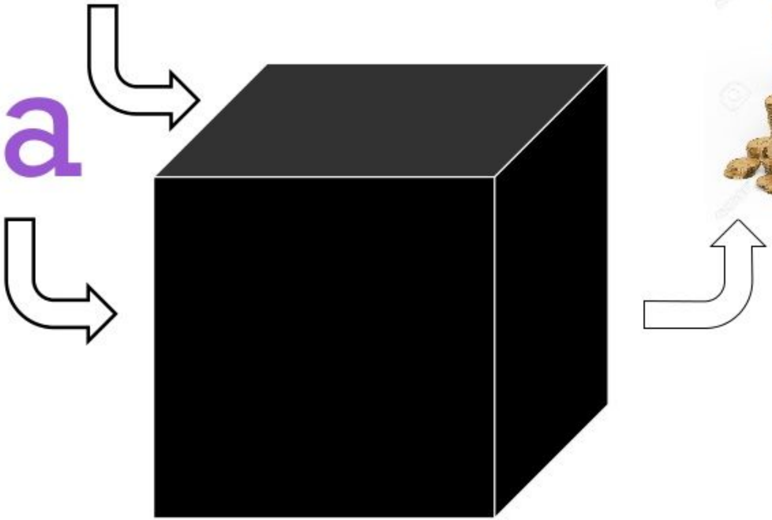
[Smith ACL talk 2017]

Noah’s Bias

- Parsing sentences into predicate-argument structures
  - Fillmore frames
  - Semantic dependency graphs
- Language models that dynamically track entities

(structural) bias

data



Why Relation Extraction Worked

- In very restricted domains

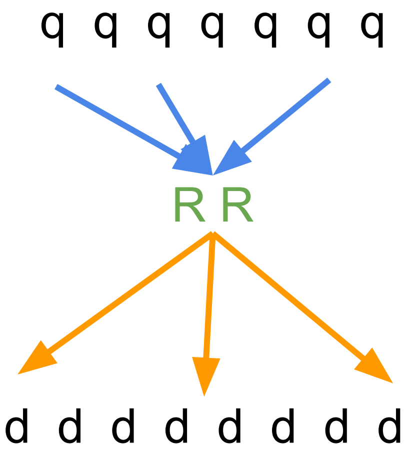
Closed domain queries

Semantic parser

KB relations

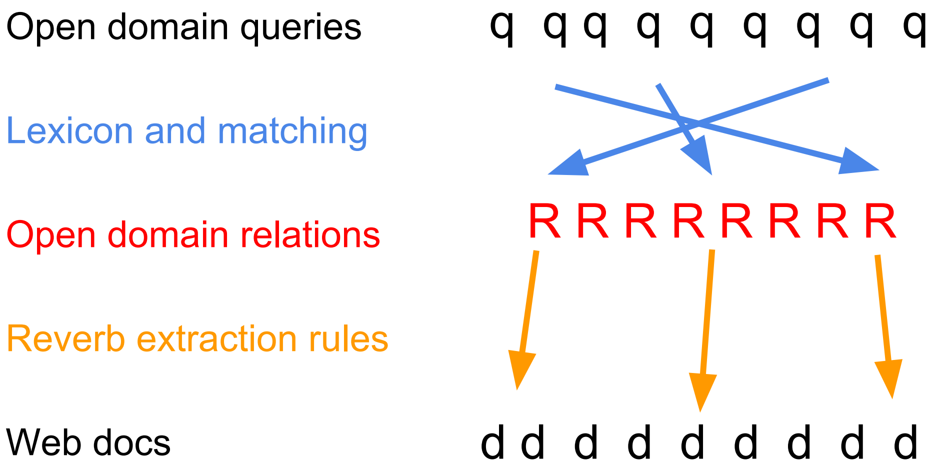
Text or html patterns

Web docs



Why Open Domain Relation Extraction Is Hard

- Open domain schemas are not compact enough



how to define a schema for OIE?

[Konstas+ 2017, UW]

Neural AMR: Sequence-to-Sequence Models for Parsing and Generation

US officials held an expert group meeting in **January** **2002** in **New York**.

```
(h / hold-04
  :ARG0 (p2 / person
    :ARG0-of (h2 / have-org-role-91
          :ARG1 (c2 / country
            :name (n3 / name
              :op1 "United" op2: "States"))
          :ARG2 (o / official)))
  :ARG1 (m / meet-03
    :ARG0 (p / person
          :ARG1-of (e / expert-01)
            :ARG2-of (g / group-01)))
  :time (d2 / date-entity :year 2002 :month 1)
  :location (c / city
    :name (n / name :op1 "New" :op2 "York")))
```

(a) **US** officials held an expert group meeting in **January** **2002** in **New York**.

```
hold
  :ARG0 person :ARG0-of have-org-role :ARG1 country :name name :op1
United :op2 States :ARG2 official
  :ARG1 meet :ARG0 person :ARG1-of expert :ARG2-of group
  :time date-entity :year 2002 :month 1
  :location city :name name :op1 New :op2 York
```

(b) **country_0** officials held an expert group meeting in **month_0** **year_0** in **city_1**.

```
hold
  :ARG0 person :ARG0-of have-org-role :ARG1 country_0 :ARG2 official
  :ARG1 meet :ARG0 person :ARG1-of expert :ARG2-of group
  :time date-entity year_0 month_0
  :location city_1
```

(c) **loc_0** officials held an expert group meeting in **month_0** **year_0** in **loc_1**.

```
hold
  :ARG0 person :ARG0-of have-org-role :ARG1 loc_0 :ARG2 official
  :ARG1 meet :ARG0 person :ARG1-of expert :ARG2-of group
  :time date-entity year_0 month_0
  :location loc_1
```

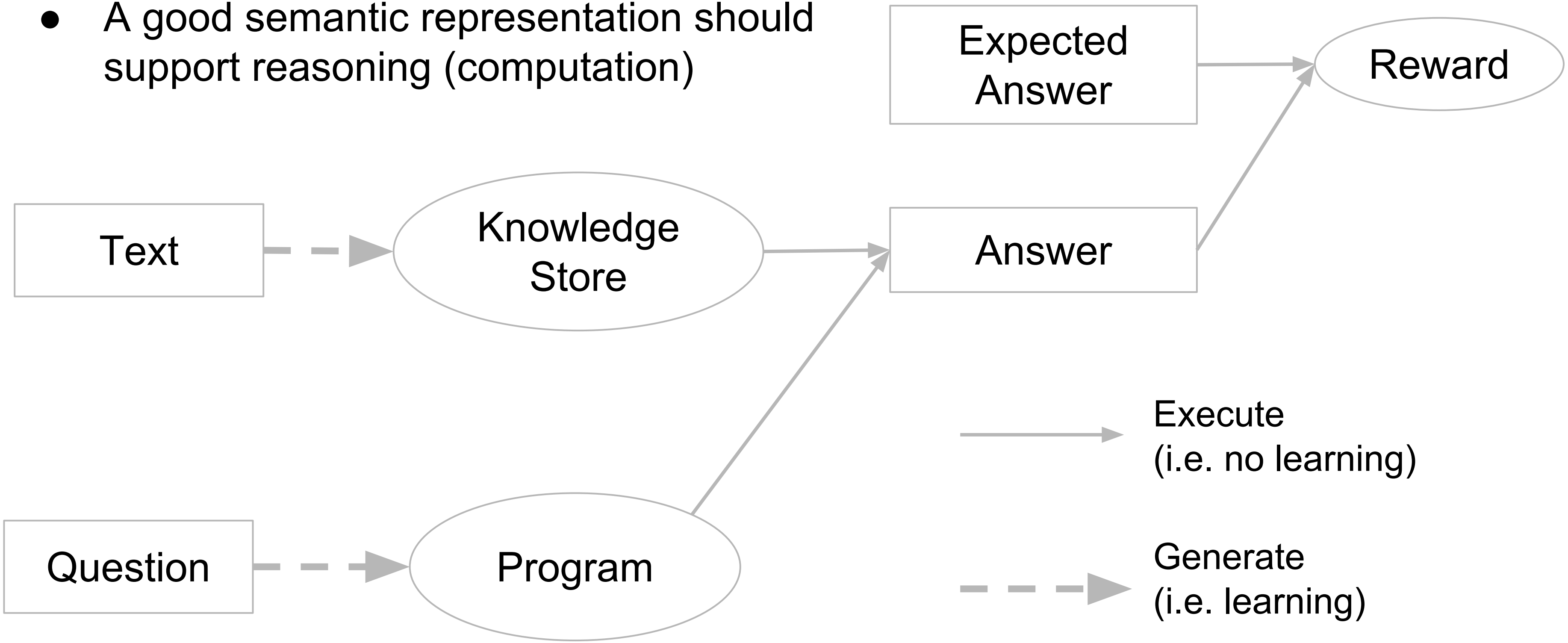
(d) **loc_0** officials held an expert group meeting in **month_0** **year_0** in **loc_1**.

```
hold
  :ARG0 ( person :ARG0-of ( have-org-role :ARG1 loc_0 :ARG2 official ) )
  :ARG1 ( meet :ARG0 ( person :ARG1-of expert :ARG2-of group ) )
  :time ( date-entity year_0 month_0 )
  :location loc_1
```

Figure 2: Preprocessing methods applied to sentence (top row) - AMR graph (left column) pairs. Sentence-graph pairs after (a) graph simplification, (b) named entity anonymization, (c) named entity clustering, and (d) insertion of scope markers.

Question answering as a simple test bed

- A good semantic representation should support reasoning (computation)

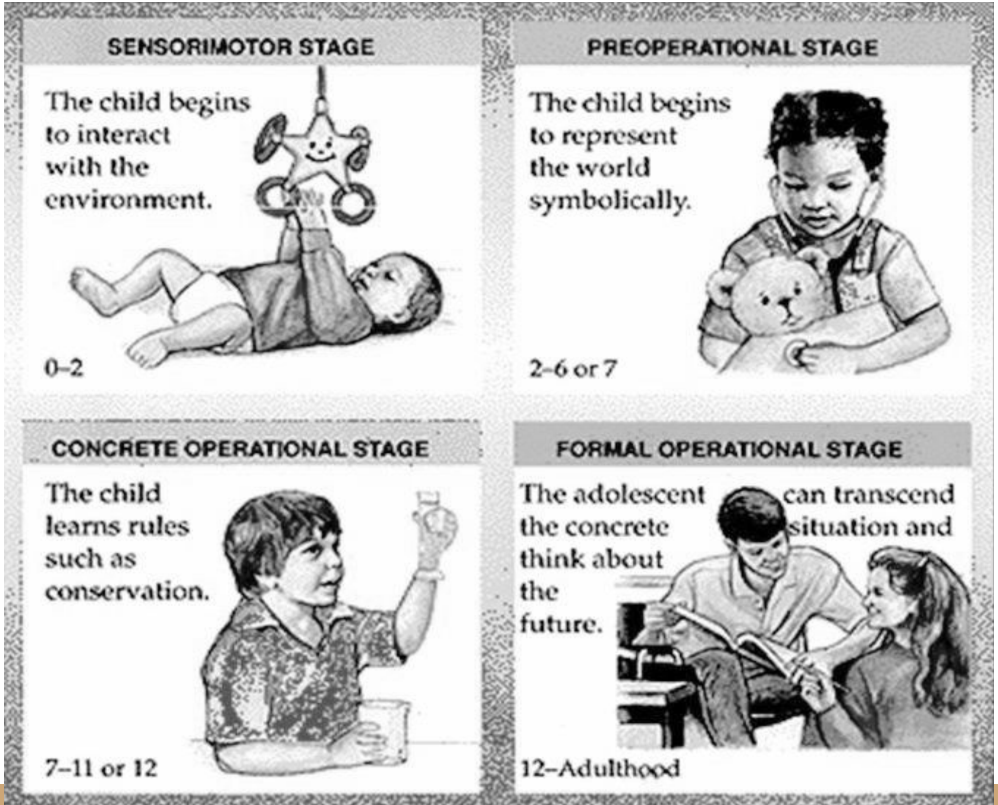


Thanks

"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

Theory of cognitive development

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



Combine KB completion models with relation extractions

[Dong+ 2014]