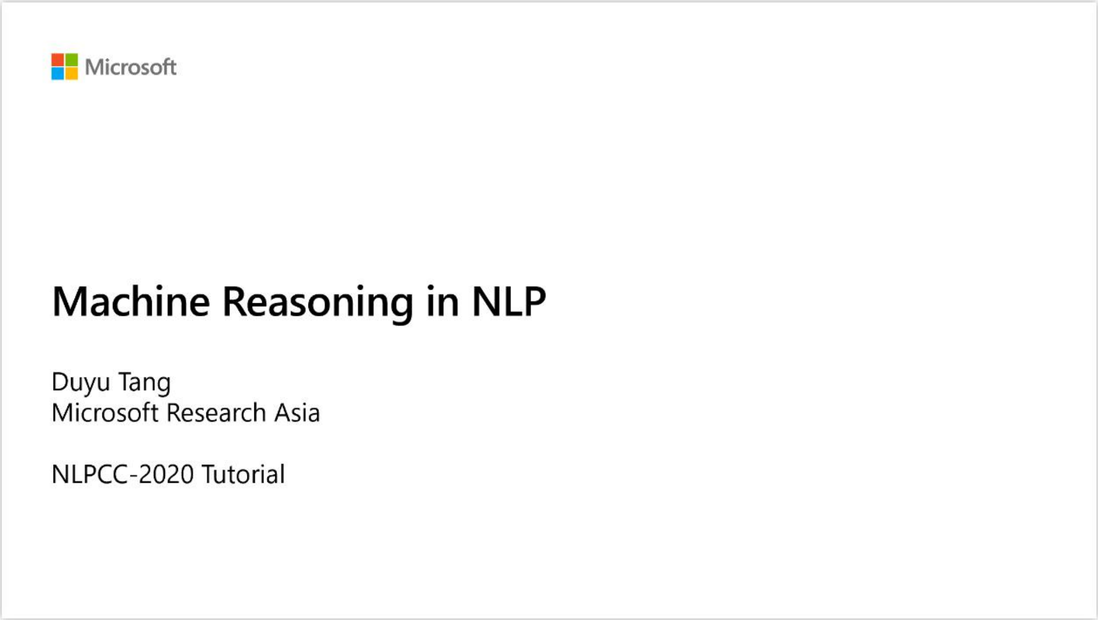
Microsoft

Machine Reasoning in NLP

Duyu Tang
Microsoft Research Asia

Past Tutorials on Machine Reasoning

- NLPCC-2020 Tutorial (3 hours)



- EMNLP-2020 Tutorial (3 hours)

Both tutorials are available at
https://tangduyu.github.io/

Microsoft AI Breakthroughs

- Gradually approaching human parity

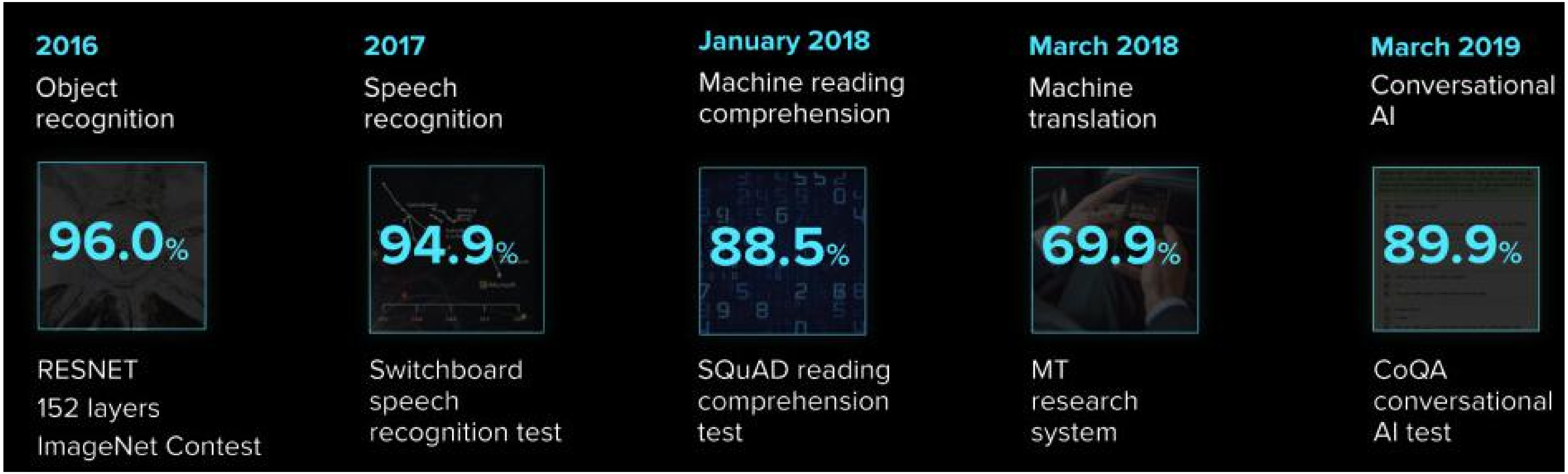
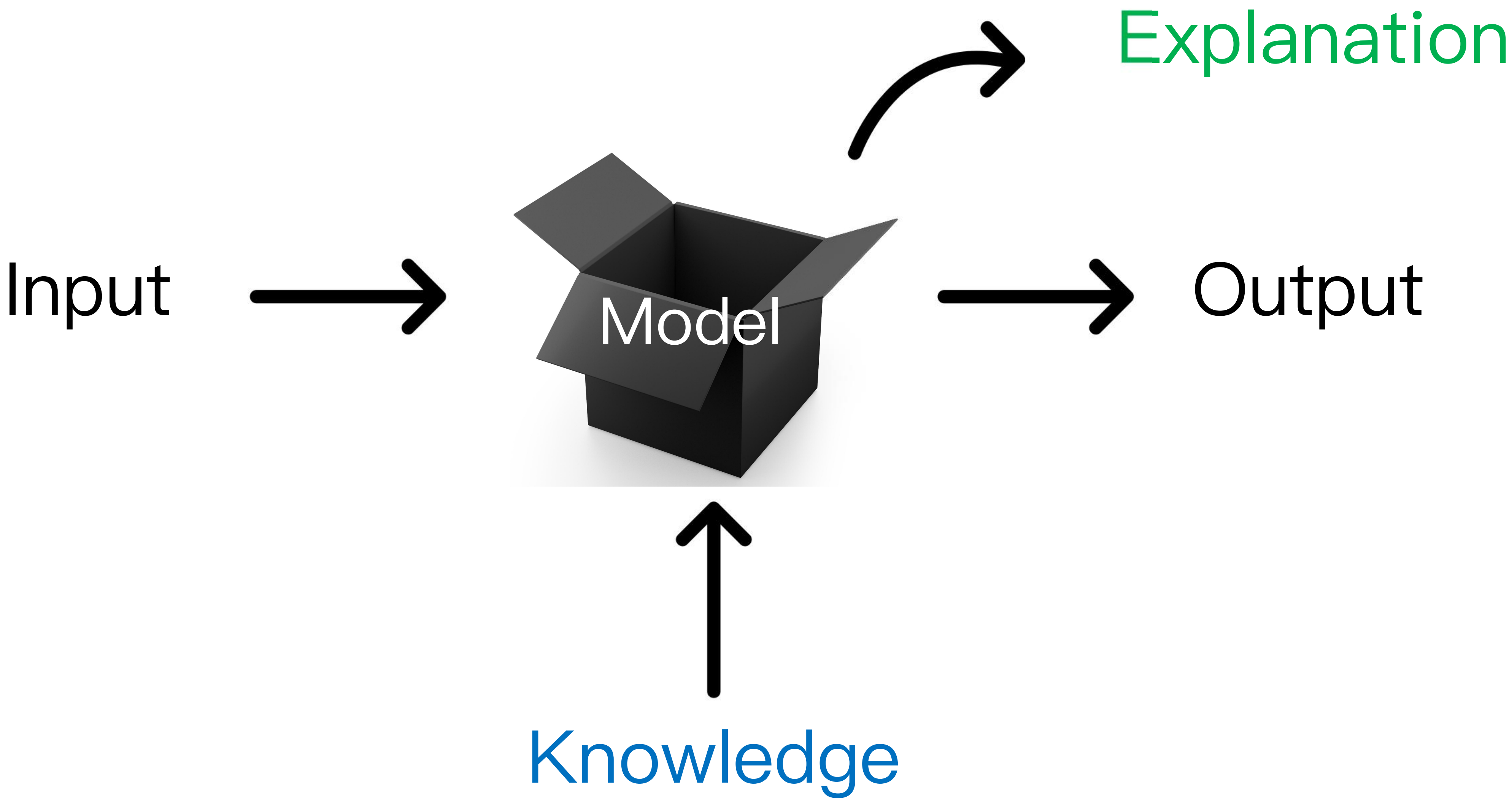


Image Courtesy: Harry Shum, 2020

Limitations:

1. Lack of transparency of the decision-making process
2. Highly rely on annotated data, ignore human/expert knowledge



Explanation
Input
Model
Output
Knowledge

Example #1: Simple Question Answering

Question Which city hosted the Summer Olympics in 2008?

Semantic Parsing

SQL SEET Chate WEB Yar = 208

Execution

Answer Beijing

Year	City	Country	Nations
1896	Athens	Greece	14
1900	Pairs	France	24
…	…	…	…
2004	Athens	Greece	201
2008	Beijing	China	204
2012	London	UK	204

Required Knowledge: Symbolic

Example #2: Multi-Turn Question Answering

Q1: **Which city hosted the Summer Olympics in 2008?**

SELECT Character WHERE Year = 2008

A1: **Beijing**

Q2: **How many nations participate that year?**

SELECT Nations WHERE Year = 2008

A2: **204**

Q3: **How about 2004?**

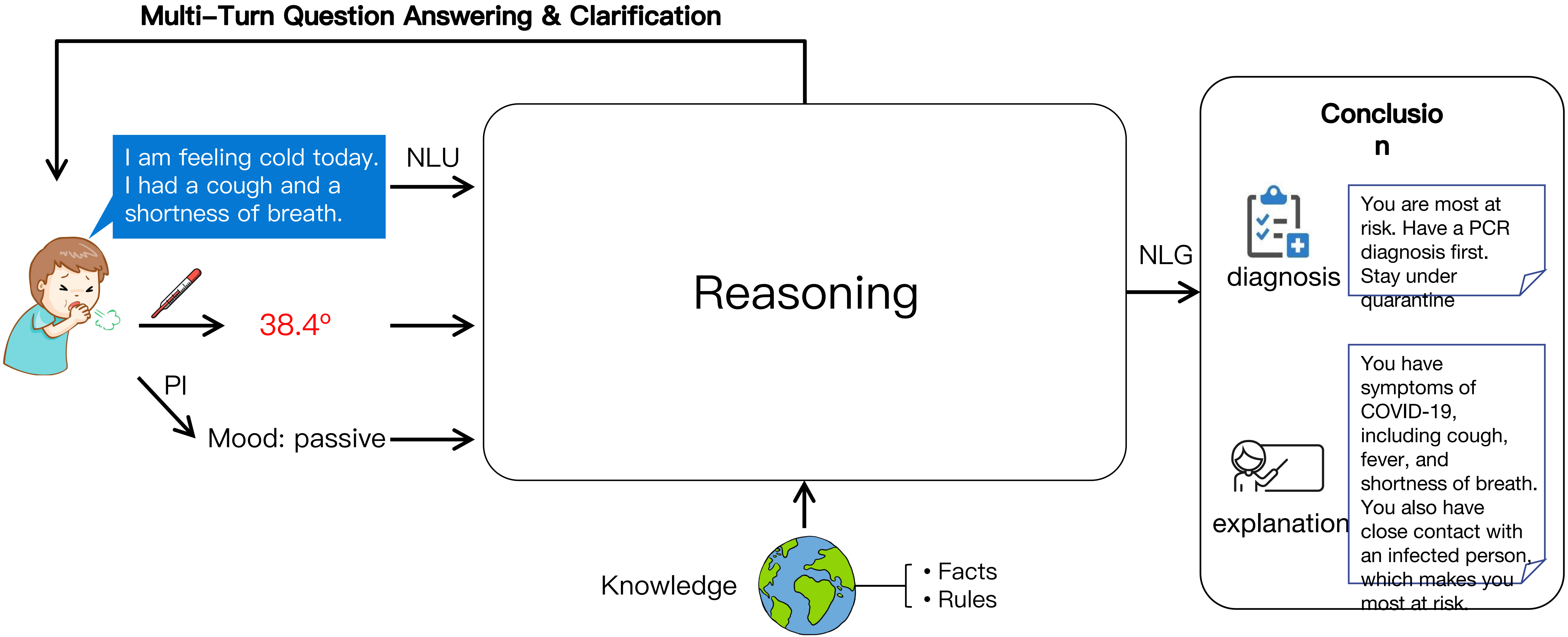
SELECT Nations WHERE Year = 2004

A3: **201**

Year	City	Country	Nations
1896	Athens	Greece	14
1900	Pairs	France	24
…	…	…	…
2004	Athens	Greece	201
2008	Beijing	China	204
2012	London	UK	204

Required Knowledge: Symbolic + **Conversation**

Example #3: Medical Diagnosis



Required Knowledge: Symbolic + Conversation + **Domain Knowledge**

Example #3: Fact

World Health Organization

Covid-19 Response Fund

Donate

WHO Coronavirus Disease (COVID-19) Dashboard

Data last updated: 2020/10/5, 3:54pm CEST

Overview | Data Table | Explore

Situation by Country, Territory & Area

Name	Cases - cumulative total	Cases - newly reported in last 24 hours	Deaths - cumulative total	Deaths - newly reported in last 24 hours	Transmission Classification
Global	**35,109,317**	**294,763**	**1,035,341**	**4,526**	
United States...	7,305,270	49,036	208,064	698	Community transmission
India	6,623,815	74,442	102,685	903	Clusters of cases
Brazil	4,906,833	26,310	145,987	599	Community transmission
Russian Fed...	1,225,889	10,888	21,475	117	Clusters of cases
Colombia	848,147	6,616	26,556	159	Community transmission
Peru	824,985	3,421	32,665	56	Community transmission
Argentina	790,818	11,129	20,795	196	Community transmission
Spain	789,932	0	32,086	0	Clusters of cases
Mexico	757,953	4,863	78,880	388	Community transmission
South Africa	681,289	1,573	16,976	38	Community transmission

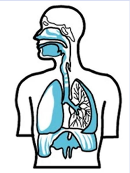
https://covid19.who.int/table

Example #3: Rule

SYMPTOM

IF <a person is infected >
THEN <he/she may have fever>

IF <a person is infected >
THEN <he/she may have cough and shortness of breath>



mild

severe

DIAGNOSIS & TREATMENT

IF < disease == COVID-19>
THEN <diagnosis(disease)=PCR
(Polymerase Chain

IF < disease == COVID-19>
THEN <treatment(disease)=none
AND vaccine(disease)=none >
(by Jan 31, 2020)

GROUPS AT RISK

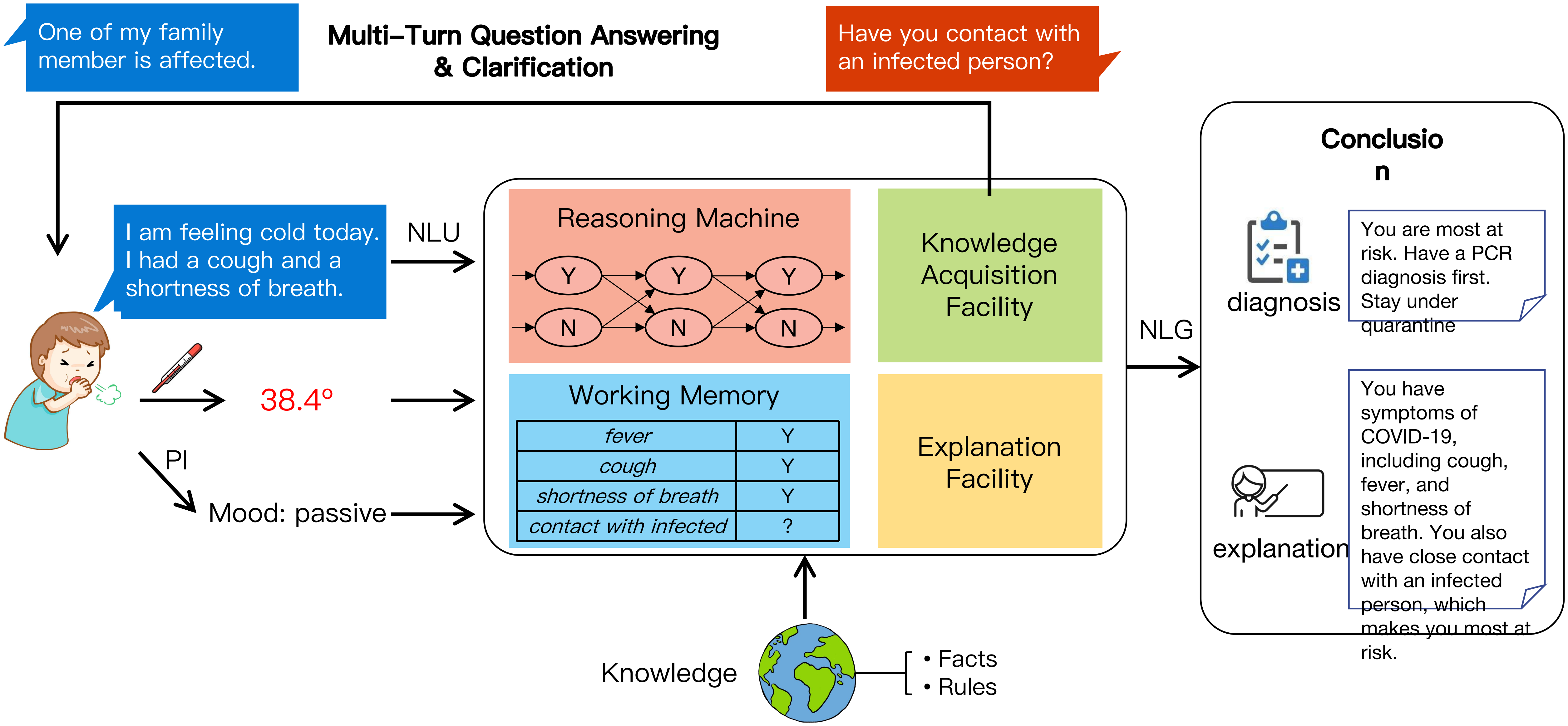
IF <a person has close contact with animals>
THEN <the person is at risk>

IF <a person is a live animal market worker>
THEN <he/she has close contact with animals>

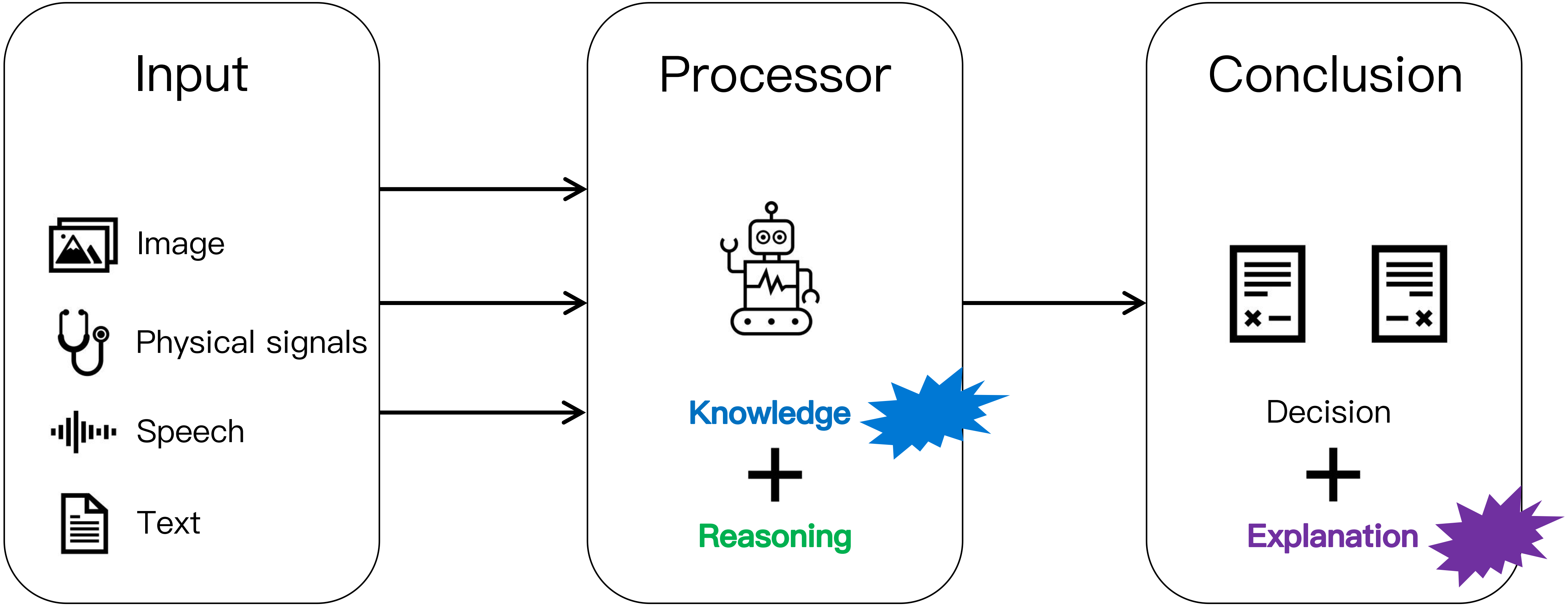
IF <a person has close contact with an infected person>
THEN <the person is most at risk>

IF <a person is a healthcare worker or a family member of infected person>
THEN <the person has close contact with infected person >

Example #3: Machine Reasoning Pipeline



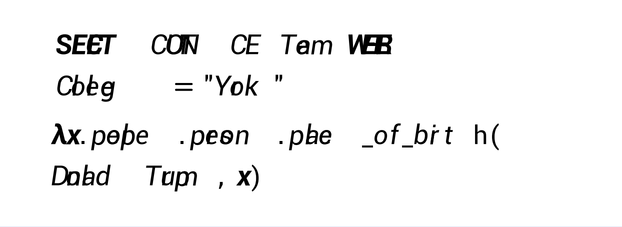
Features of Machine Reasoning



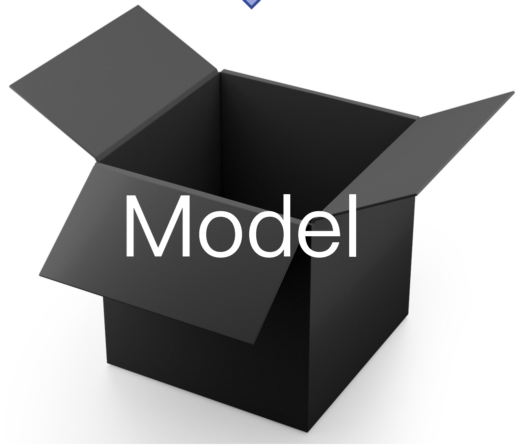
+ Propositional/First–Order Logic
+ Neuro–symbolic

$$
\begin{aligned}
(\alpha \wedge \beta) &\equiv (\beta \wedge \alpha) && \text{commutativity of } \wedge \\
(\alpha \vee \beta) &\equiv (\beta \vee \alpha) && \text{commutativity of } \vee \\
((\alpha \wedge \beta) \wedge \gamma) &\equiv (\alpha \wedge (\beta \wedge \gamma)) && \text{associativity of } \wedge \\
((\alpha \vee \beta) \vee \gamma) &\equiv (\alpha \vee (\beta \vee \gamma)) && \text{associativity of } \vee \\
\neg(\neg\alpha) &\equiv \alpha && \text{double-negation elimination} \\
(\alpha \Rightarrow \beta) &\equiv (\neg\beta \Rightarrow \neg\alpha) && \text{contraposition} \\
(\alpha \Rightarrow \beta) &\equiv (\neg\alpha \vee \beta) && \text{implication elimination}
\end{aligned}
$$

First-Order Logic



Symbolic Operations

Input → 

→ Output

~~+ Pre–trained models (e.g. ELMo, BERT, GPT)~~
+ Evidence (e.g. retrieved docs from Wikipedia/web, retrieved facts from Wikidata/ConceptNet)

BERT

WIKIPEDIA

WIKIDATA

ConceptNet
An open, multilingual knowledge graph

Agenda

Opening
Logic-based Models in NLP
Neural-Symbolic Models in NLP
Evidence-based Models in NLP
Summary

Logic-based Models in NLP

Outline

- Propositional Logic and First-Order Logic

We are Here

- Inference/Theorem Proving: Forward and Backward Chaining

- Application in NLP

Propositional Logic

- **Logical constants**: true, false
- **Propositional symbols**: P, Q, S, ... (**atomic sentences**)
- Wrapping **parentheses**: (…)
- Sentences are combined by **connectives**:

  $\wedge$ ...and [conjunction]

  $\vee$ ...or [disjunction]

  $\rightarrow$...implies [implication / conditional]

  $\leftrightarrow$..is equivalent [biconditional]

  $\neg$ ...not [negation]
- **Literal**: atomic sentence or negated atomic sentence

Attribution: Marie desJardins, Fall 2005

Propositional Logic Examples

- P means "It is hot"
- Q means "It is humid"
- R means "It is raining"

- $(P \wedge Q) \rightarrow R$
  "If it is hot and humid, then it is raining"
- $Q \rightarrow P$
  "If it is humid, then it is hot"
- Q
  "It is humid."

Attribution: Marie desJardins, Fall 2005

Propositional Logic Syntax

- Given: a set of proposition symbols $\{X_1, X_2, \ldots, X_n\}$
  - (we often add True and False for convenience)
- X_i is a sentence
- If α is a sentence then $\neg\alpha$ is a sentence
- If α and β are sentences then $\alpha \wedge \beta$ is a sentence
- If α and β are sentences then $\alpha \vee \beta$ is a sentence
- If α and β are sentences then $\alpha \Rightarrow \beta$ is a sentence
- If α and β are sentences then $\alpha \Leftrightarrow \beta$ is a sentence
- And p.s. there are no other sentences!

Logical Equivalence

$$
\begin{aligned}
(\alpha \land \beta) &\equiv (\beta \land \alpha) && \text{commutativity of } \land \\
(\alpha \lor \beta) &\equiv (\beta \lor \alpha) && \text{commutativity of } \lor \\
((\alpha \land \beta) \land \gamma) &\equiv (\alpha \land (\beta \land \gamma)) && \text{associativity of } \land \\
((\alpha \lor \beta) \lor \gamma) &\equiv (\alpha \lor (\beta \lor \gamma)) && \text{associativity of } \lor \\
\neg(\neg\alpha) &\equiv \alpha && \text{double-negation elimination} \\
(\alpha \Rightarrow \beta) &\equiv (\neg\beta \Rightarrow \neg\alpha) && \text{contraposition} \\
(\alpha \Rightarrow \beta) &\equiv (\neg\alpha \lor \beta) && \text{implication elimination} \\
(\alpha \Leftrightarrow \beta) &\equiv ((\alpha \Rightarrow \beta) \land (\beta \Rightarrow \alpha)) && \text{biconditional elimination} \\
\neg(\alpha \land \beta) &\equiv (\neg\alpha \lor \neg\beta) && \text{De Morgan} \\
\neg(\alpha \lor \beta) &\equiv (\neg\alpha \land \neg\beta) && \text{De Morgan} \\
(\alpha \land (\beta \lor \gamma)) &\equiv ((\alpha \land \beta) \lor (\alpha \land \gamma)) && \text{distributivity of } \land \text{ over } \lor \\
(\alpha \lor (\beta \land \gamma)) &\equiv ((\alpha \lor \beta) \land (\alpha \lor \gamma)) && \text{distributivity of } \lor \text{ over } \land
\end{aligned}
$$

Stuart Russell and Peter Norvig. *Artificial Intelligence. A modern approach (third edition)*. 2009.

Problems with Propositional Logic

- Hard to identify "individuals" (e.g., Mary, 3)
- Can't directly talk about properties of individuals or relations between individuals (e.g., "Bill is tall")
- Generalizations, patterns, regularities can't easily be represented (e.g., "all triangles have 3 sides")
- Lack of variables prevents stating more general rules
  - We need a set of similar rules for each cell
- First-Order Logic is expressive enough to concisely represent this kind of information
  - FOL adds relations, variables, and quantifiers, e.g.,
    - *"Every elephant is gray":* ∀ x (elephant(x) → gray(x))
    - *"There is a white alligator":* ∃ x (alligator(X) ^ white(X))

Attribution: Marie desJardins, Fall 2005

First–Order Logic

- First–order logic models the world in terms of
  - **Objects,** which are things with individual identities
  - **Properties** of objects that distinguish them from other objects
  - **Relations** that hold among sets of objects
  - **Functions,** which are a subset of relations where there is only one "value" for any given "input"
- Examples:
  - Objects: Students, lectures, companies, cars ...
  - Relations: Brother–of, bigger–than, outside, part–of, has–color, occurs–after, owns, visits, precedes, ...
  - Properties: blue, oval, even, large, ...
  - Functions: father–of, best–friend, second–half, one–more–than ...

First-Order Logic Examples

- Universal quantification
  - $(\forall x)P(x)$ means that P holds for **all** values of x in the domain associated with that variable
  - E.g., $(\forall x)$ dolphin(x) $\rightarrow$ mammal(x)
- Existential quantification
  - $(\exists\ x)P(x)$ means that P holds for **some** value of x in the domain associated with that variable
  - E.g., $(\exists\ x)$ mammal(x) $\wedge$ lays-eggs(x)
  - Permits one to make a statement about some object without naming it

Language	Propositional logic	First-order logic
Syntax	The world contains facts	The world contains objects, relations, and functions
Semantics	$\alpha \wedge \beta$ is true in a world iff α is true and β is true (etc.)	$\phi(\sigma)$ is true in a world if $\sigma = o_j$ and ϕ holds for o_j; etc.

Outline

- Propositional Logic and First-Order Logic
- Inference/Theorem Proving: Forward and Backward Chaini

  We are Here

- Application in NLP

Forward Chaining

- Start with given proposition symbols (atomic sentence)
  - e.g., A and B
- Iteratively try to infer truth of additional proposition symbols
  - e.g., $A \land B \Rightarrow C$, therefor we establish C is true
- Continue until
  - — no more inference can be carried out, or
  - — goal is reached

Attribution: Philipp Koehn, 2020

Forward Chaining Example: Proving Q

CLAUSES

- $P \Rightarrow Q$
- $L \wedge M \Rightarrow P$
- $B \wedge L \Rightarrow M$
- $A \wedge P \Rightarrow L$
- $A \wedge B \Rightarrow L$
- A
- B

COUNT

~~1~~ 0

~~2~~ ~~1~~ 0

~~2~~ ~~1~~ 0

~~2~~ ~~1~~ 0

~~2~~ ~~1~~ 0

0

0

INFERRED

A ~~false~~ ~~xxxx~~ true

B ~~false~~ ~~xxxx~~ true

L ~~false~~ ~~xxxx~~ true

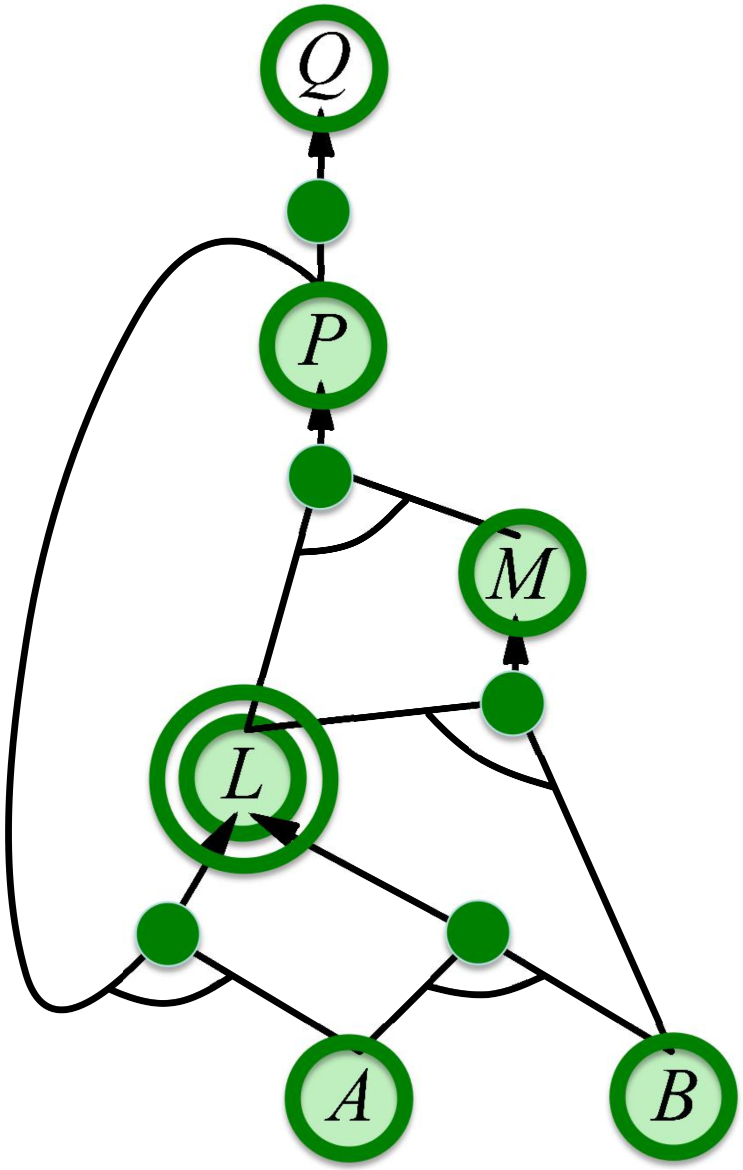
M ~~false~~ ~~xxxx~~ true

P ~~false~~ ~~xxxx~~ true

Q ~~false~~ ~~xxxx~~ true

AGENDA

~~A~~ ~~B~~ ~~L~~ ~~M~~ ~~P~~ ~~L~~ ~~Q~~



Forward Chaining Algorithm

```
function PL-FC-ENTAILS?(KB, q) returns true or false
    count ← a table, where count[c] is the number of symbols in c's premise
    inferred ← a table, where inferred[s] is initially false for all s
    agenda ← a queue of symbols, initially symbols known to be true in KB

    while agenda is not empty do
        p ← Pop(agenda)
        if p = q then return true
        if inferred[p] = false then
            inferred[p]←true
            for each clause c in KB where p is in c.premise do
                decrement count[c]
                if count[c] = 0 then add c.conclusion to agenda
    return false
```

Backward Chaining

- Idea: work backwards from the query Q:
  - to prove Q by BC,
    - check if Q is known already, or
    - prove by BC all premises of some rule concluding q

- Avoid loops: check if new subgoal is already on the goal stack

- Avoid repeated work: check if new subgoal
  - 1. has already been proved true, or
  - 2. has already failed

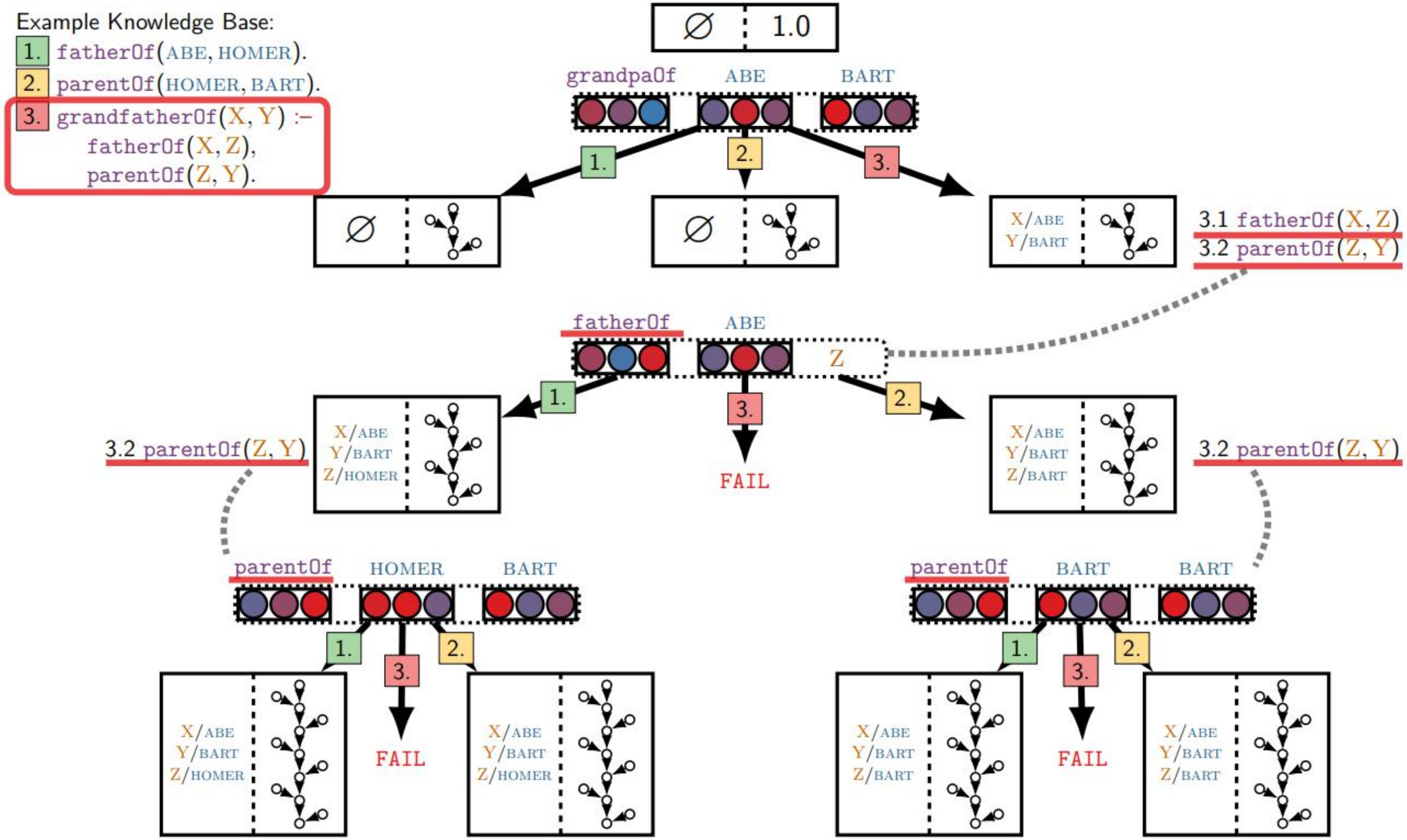
Attribution: Philipp Koehn, 2020

Outline

- Propositional Logic and First-Order Logic
- Inference/Theorem Proving: Forward and Backward Chaining
- Application in Knowledge Base Completion
  - Neural Backward Chaining
  - Logic as constraints

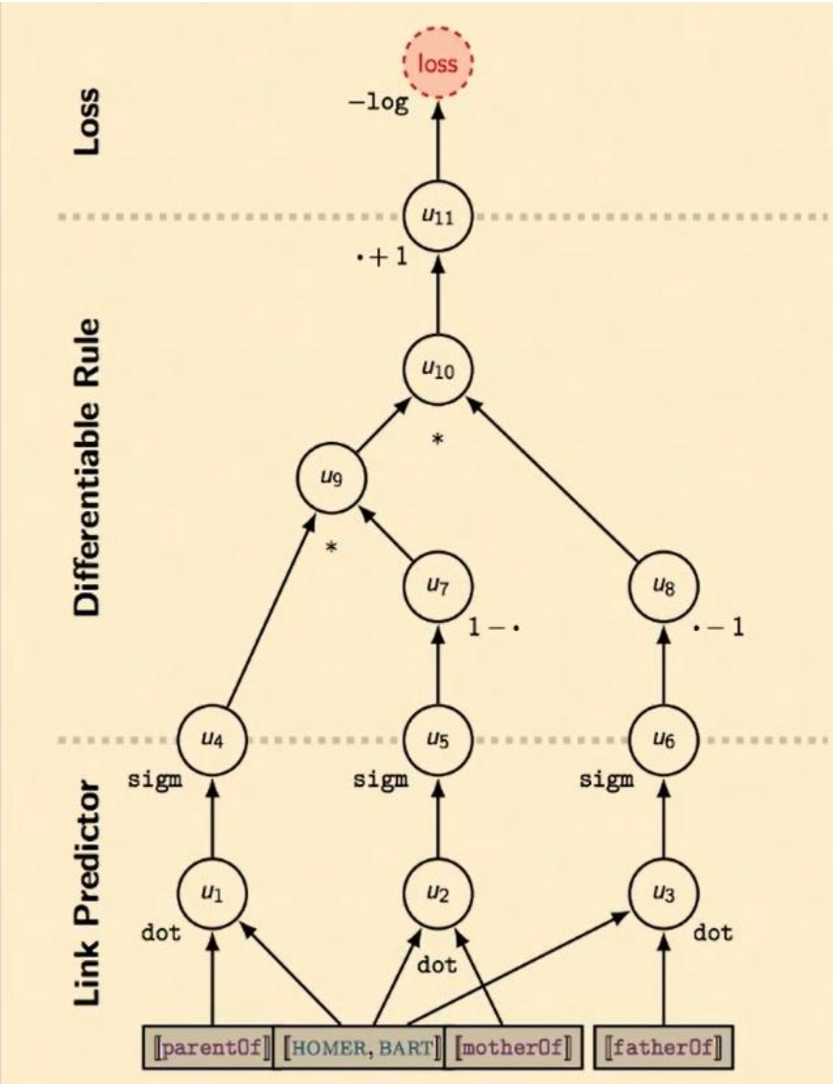
We are Here

Neural Backward Chaining



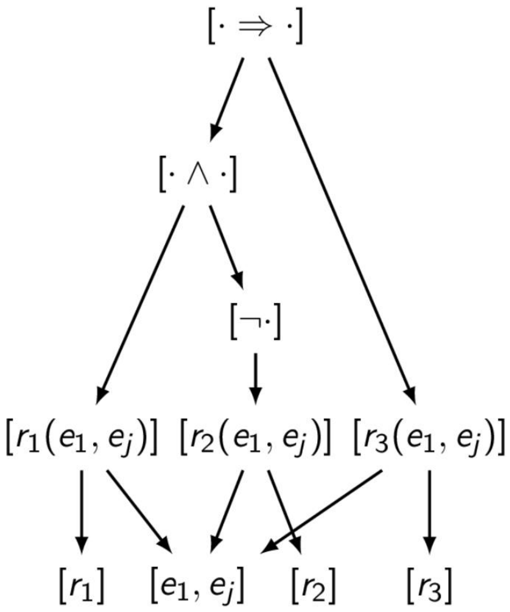
Tim Rocktaschel and Sebastian Riedel. **End-to-End Differentiable Proving**. NIPS-2017

Logic as Constraints



$$[\mathcal{F}] = \begin{cases} \sigma(\mathbf{v}_s \cdot \mathbf{v}_{ij}) & \text{if } \mathcal{F} = r_s(e_i, e_j), \text{ i.e., facts} \\ 1 - [\mathcal{A}] & \text{if } \mathcal{F} = \neg\mathcal{A} \\ [\mathcal{A}] * [\mathcal{B}] & \text{if } \mathcal{F} = \mathcal{A} \wedge \mathcal{B} \end{cases}$$

$$r_1(e_i, e_j) \wedge \neg r_2(e_i, e_j) \Rightarrow r_3(e_i, e_j)$$



Tim Rocktaschel, Sameer Singh, and Sebastian Riedel. **Injecting logical background knowledge into embeddings for relation extraction**. HLT-NAACL, 2015

Performance V.S. Interpretability

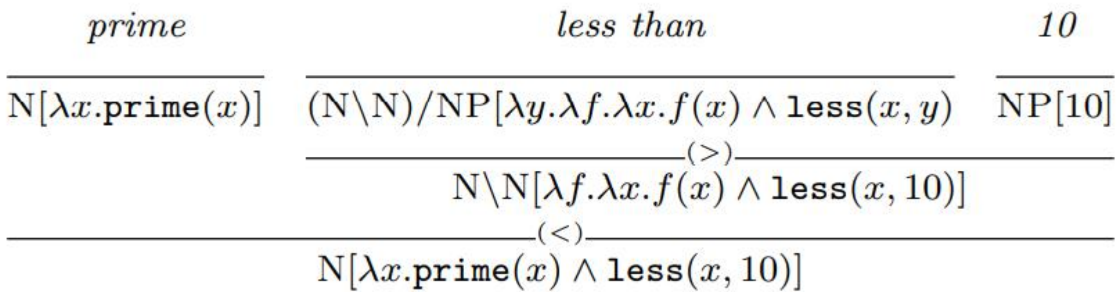
- Neural Backward Chaining
  - Good interpretability, limited scope of application (e.g., completion on structured KB)
- Regularizing Neural Models
  - Good performance with neural models as backbone, limited interpretability

Neural–Symbolic Models in NLP

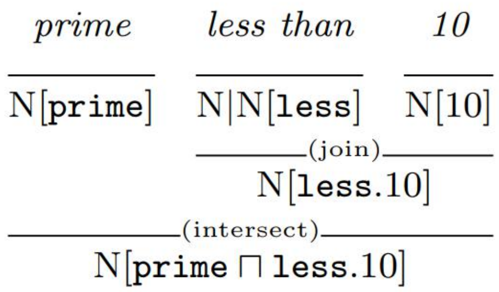
Symbolic Language

First-Order Logic	*Every prime greater than two is odd.*	∀x.prime(x) ∧ more(x, 2) → odd(x)
Lambda Calculus	*How many primes are less than 10?*	count(λx.prime(x)∧less(x, 10))
Lambda DCS	*How many primes are less than 10?*	count(prime ⊓ (less.10))

- First-order logic fails to construct a set and manipulating it.
- The λ operator can be thought of as constructing a set of all *x* that satisfy the condition; in symbols, $[\lambda x. f(x)]_c = \{x : [f(x)]_c = true\}$.



Lambda Calculus



Lambda DCS

Percy Liang. Learning Executable Semantic Parsers for Natural Language Understanding. CACM-2016

Semantic Parsing

map an utterance ***x*** in a context ***c*** to an action ***y***

x: *What is the largest prime less than 10?*

c: $\boxed{\texttt{primes} : \{2, 3, 5, 7, 11, \ldots\}}$

$\rightarrow$

y: 7

Percy Liang. Learning Executable Semantic Parsers for Natural Language Understanding. CACM–2016

A Derivation

What is the largest prime less than 10 ?

$$\frac{prime}{\mathrm{NP}[\texttt{primes}]}\,(\mathrm{R1}) \qquad \frac{10}{\mathrm{NP}[10]}\,(\mathrm{R2})$$

$$\frac{less\ than\ \ \mathrm{NP}[10]}{\mathrm{QP}[(-\infty, 10)]}\,(\mathrm{R3})$$

$$\frac{\mathrm{NP}[\texttt{primes}]\ \ \mathrm{QP}[(-\infty, 10)]}{\mathrm{NP}[\texttt{primes} \cap (-\infty, 10)]}\,(\mathrm{R4})$$

$$\frac{largest\ \ \mathrm{NP}[\texttt{primes} \cap (-\infty, 10)]}{\mathrm{NP}[\max(\texttt{primes} \cap (-\infty, 10))]}\,(\mathrm{R5})$$

$$\frac{What\ is\ the\ \ \mathrm{NP}[\max(\texttt{primes} \cap (-\infty, 10))]\ ?}{\mathrm{ROOT}[\max(\texttt{primes} \cap (-\infty, 10))]}\,(\mathrm{R7})$$

(R1)	*prime*	$\Rightarrow$	$\mathrm{NP}[\texttt{primes}]$
(R2)	*10*	$\Rightarrow$	$\mathrm{NP}[10]$
(R3)	*less than* $\mathrm{NP}[z]$	$\Rightarrow$	$\mathrm{QP}[(-\infty, z)]$
(R4)	$\mathrm{NP}[z_1]\ \mathrm{QP}[z_2]$	$\Rightarrow$	$\mathrm{NP}[z_1 \cap z_2]$
(R5)	*largest* $\mathrm{NP}[z]$	$\Rightarrow$	$\mathrm{NP}[\max(z)]$
(R6)	*largest* $\mathrm{NP}[z]$	$\Rightarrow$	$\mathrm{NP}[\min(z)]$
(R7)	*What is the* $\mathrm{NP}[z]$?	$\Rightarrow$	$\mathrm{ROOT}[z]$

Percy Liang. Learning Executable Semantic Parsers for Natural Language Understanding. CACM–2016

Semantic Parsing

- Map Natural Language into machine executable logical forms

Question: How many CFL teams are from York College?

Semantic Parsing

SQL: SEET COTN CE Tam WEB Cbleg = "Yok "

CFL Team	College
Hamilton Tiger-Cats	Wilfrid Laurier
Calgary Stampeders	York
Toronto Argonauts	York

Execution

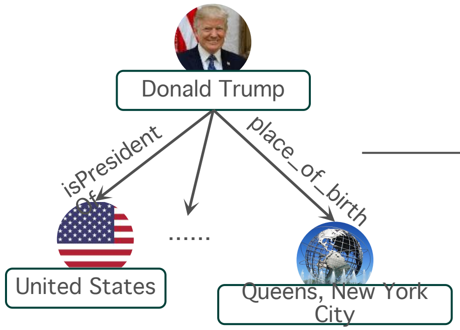
Answer: 2

Table/DB

Question: Where was Donald Trump given birth?

Semantic Parsing

LF: λx. pepe .peon .pke _of_bir t h(Dobd Tupn , x)

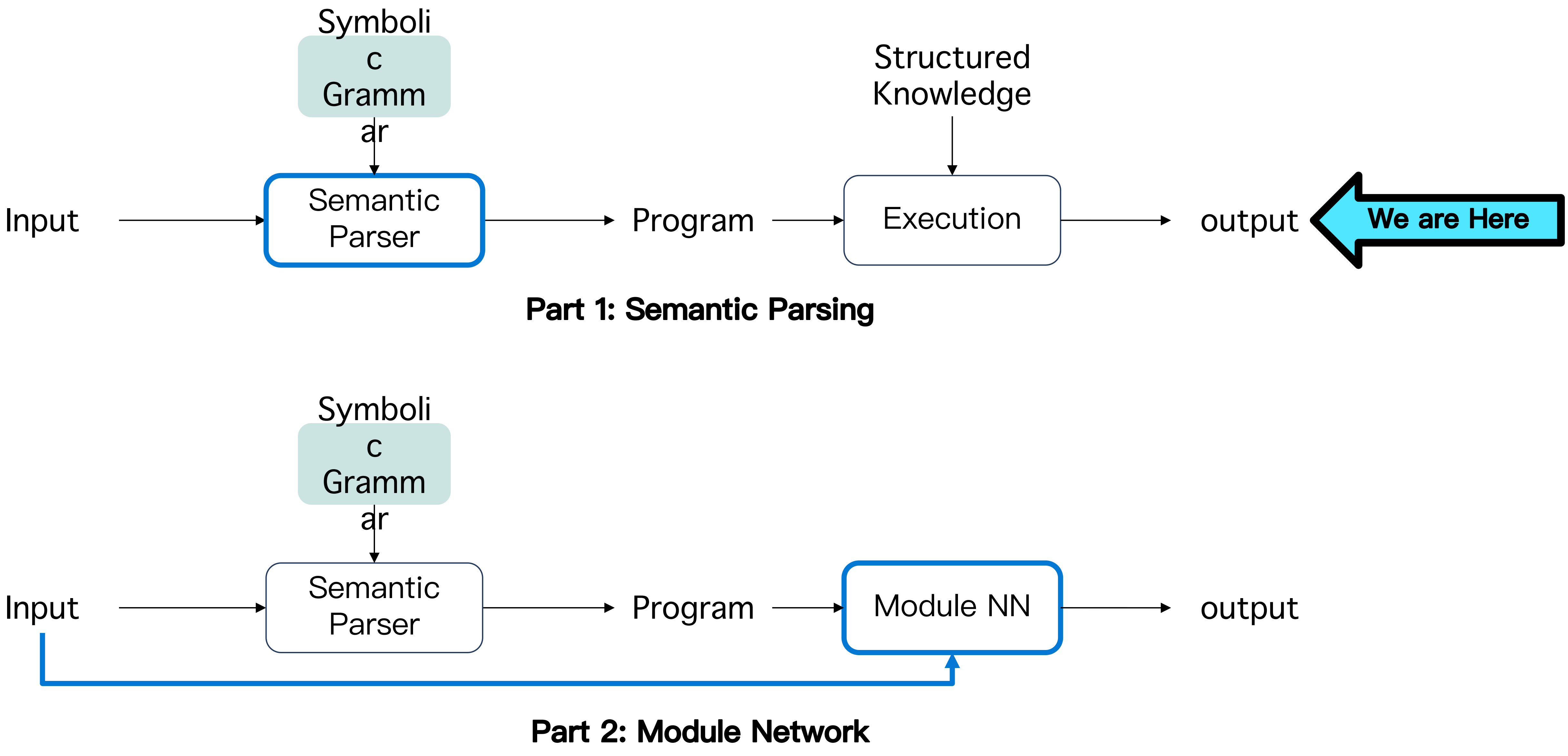


Execution

Answer: Queens, New York City

Knowledge Graph

Outline



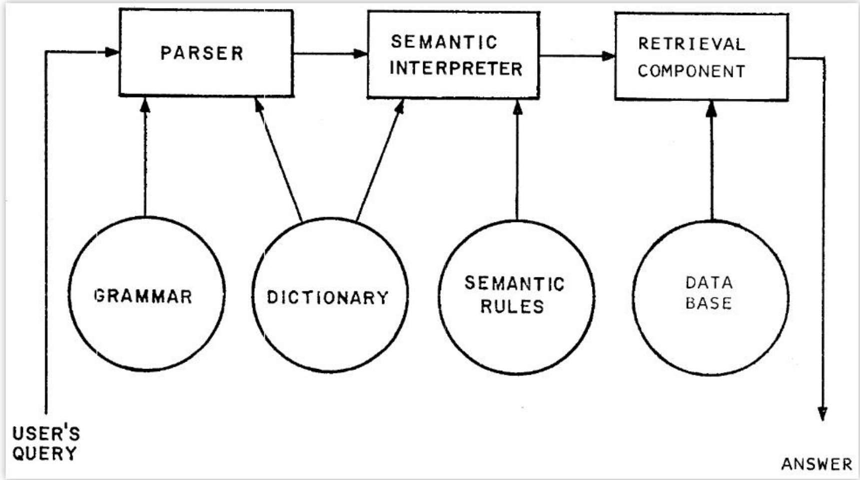
Symbolic Grammar
Structured Knowledge
Input
Semantic Parser
Program
Execution
output
We are Here
Part 1: Semantic Parsing
Symbolic Grammar
Input
Semantic Parser
Program
Module NN
output
Part 2: Module Network

LSNLIS (Lunar Sciences Natural Language Information System)

A question–answering system to enable a lunar geologist to conveniently access, compare, and evaluate the chemical analysis data on lunar rock and soil composition that is accumulating as a result of the Apollo moon missions.

Two DB files. One is a 13,000 line table of chemical and age analysis of the Apollo 11 samples extracted from the reports of a the First Annual Lunar Science Conference, and the second is a keyphrase index to those reports.

A dictionary about 3500 words.

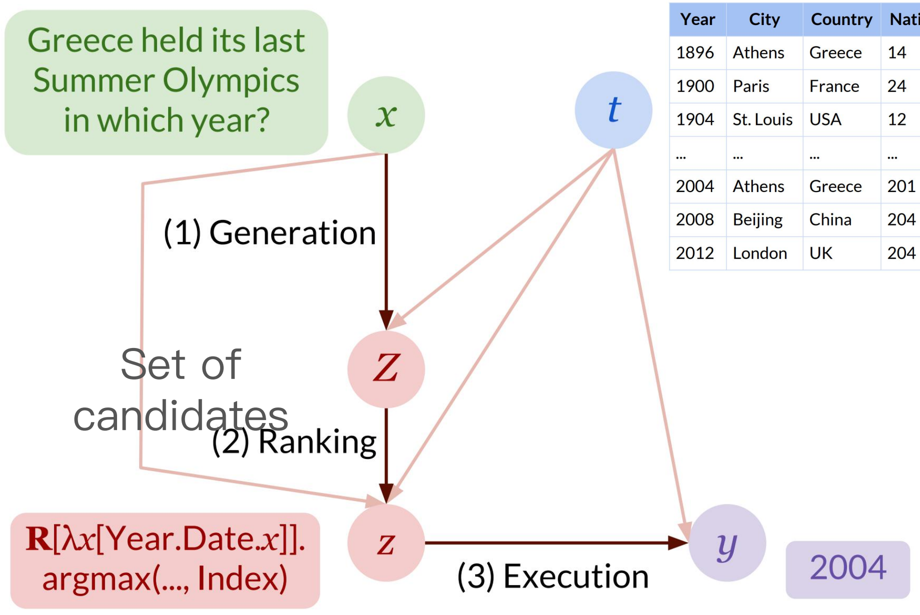


1. List the rocks which contain chromite and ulvospinel.
2. Give me all references on fayalitic olivine.
3. What minerals have been identified in the lunar samples?
4. What analyses of olivine are there?
5. What is the average analysis of Ir in rock S10055?
6. List the modes for all low Rb rocks.
7. Give me the K / Rb ratios for all lunar samples.
8. Has the mineral analcite been identified in any lunar sample?
9. What is the concentration of La in rock S10034?
10. Identify all samples in which glass was found.
11. Give me all modal analyses of lunar fines.
12. In what samples has apatite been identified?

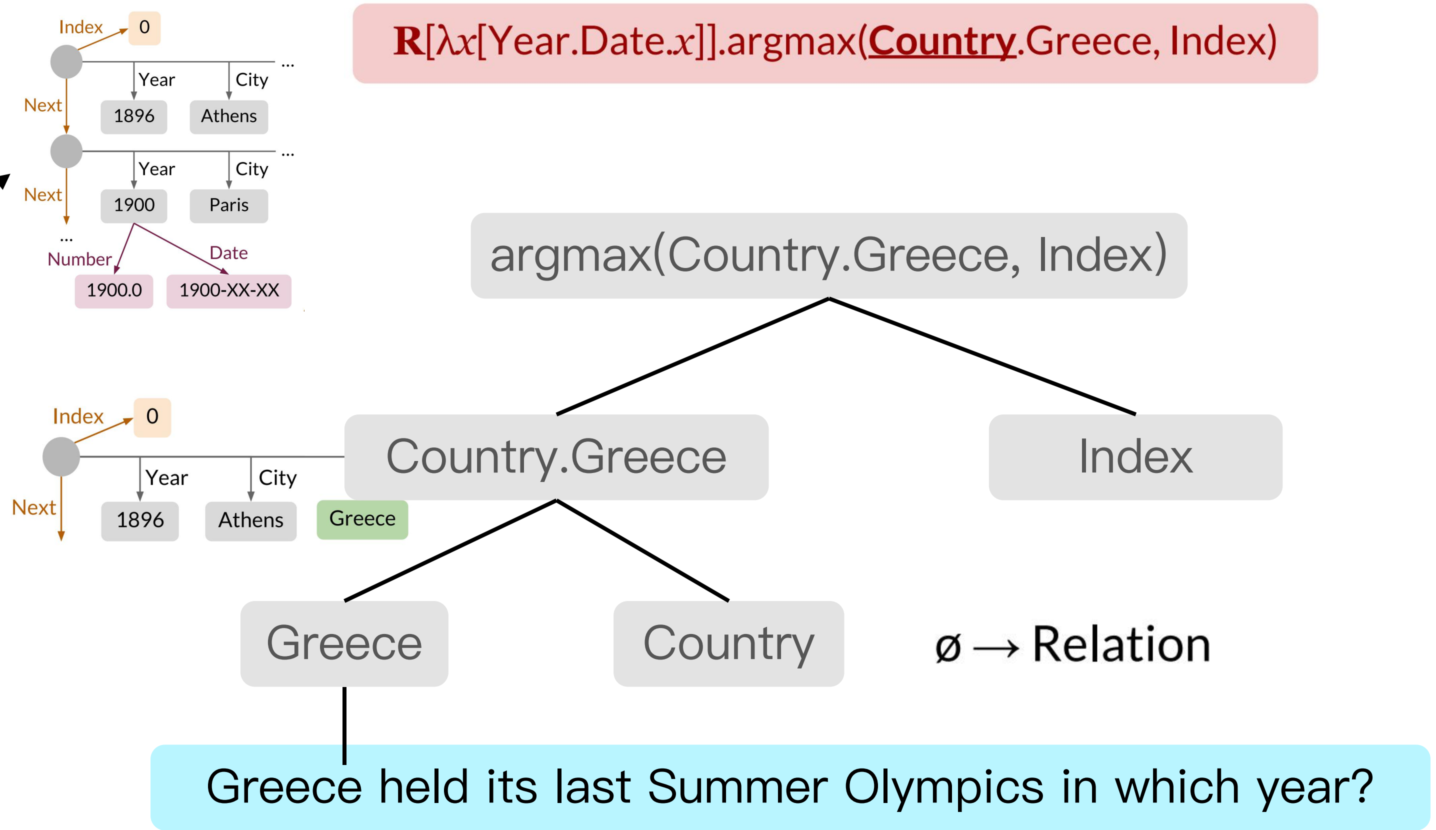
W. A. Woods, R. M. Kaplan, and B. N. Webber. The lunar sciences natural language information system: Final report. Technical report, BBN Report 2378, Bolt Beranek and Newman Inc., 1972.

Floating Parser

- Build formulas bottom–up according to a set of deduction rules
- Allow formulas to be created from nothing ("floating")



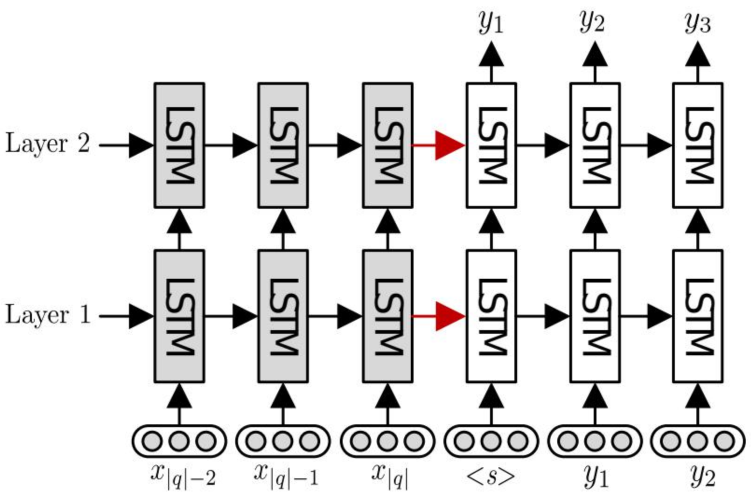
Year	City	Country	Nations
1896	Athens	Greece	14
1900	Paris	France	24
1904	St. Louis	USA	12
...	...	...	...
2004	Athens	Greece	201
2008	Beijing	China	204
2012	London	UK	204



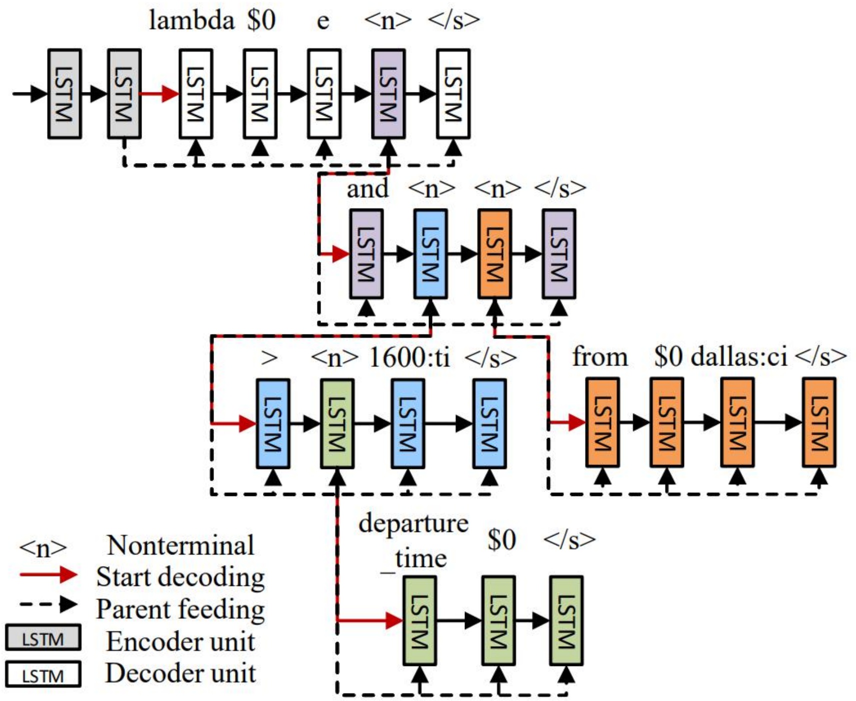
- **Entities** are anchored to token spans
- **Relations** and **Operations** are not

Panupong Pasupat and Percy Liang. "*Compositional Semantic Parsing on Semi-Structured Tables.*" ACL-2015.

Seq2Seq and Seq2Tree



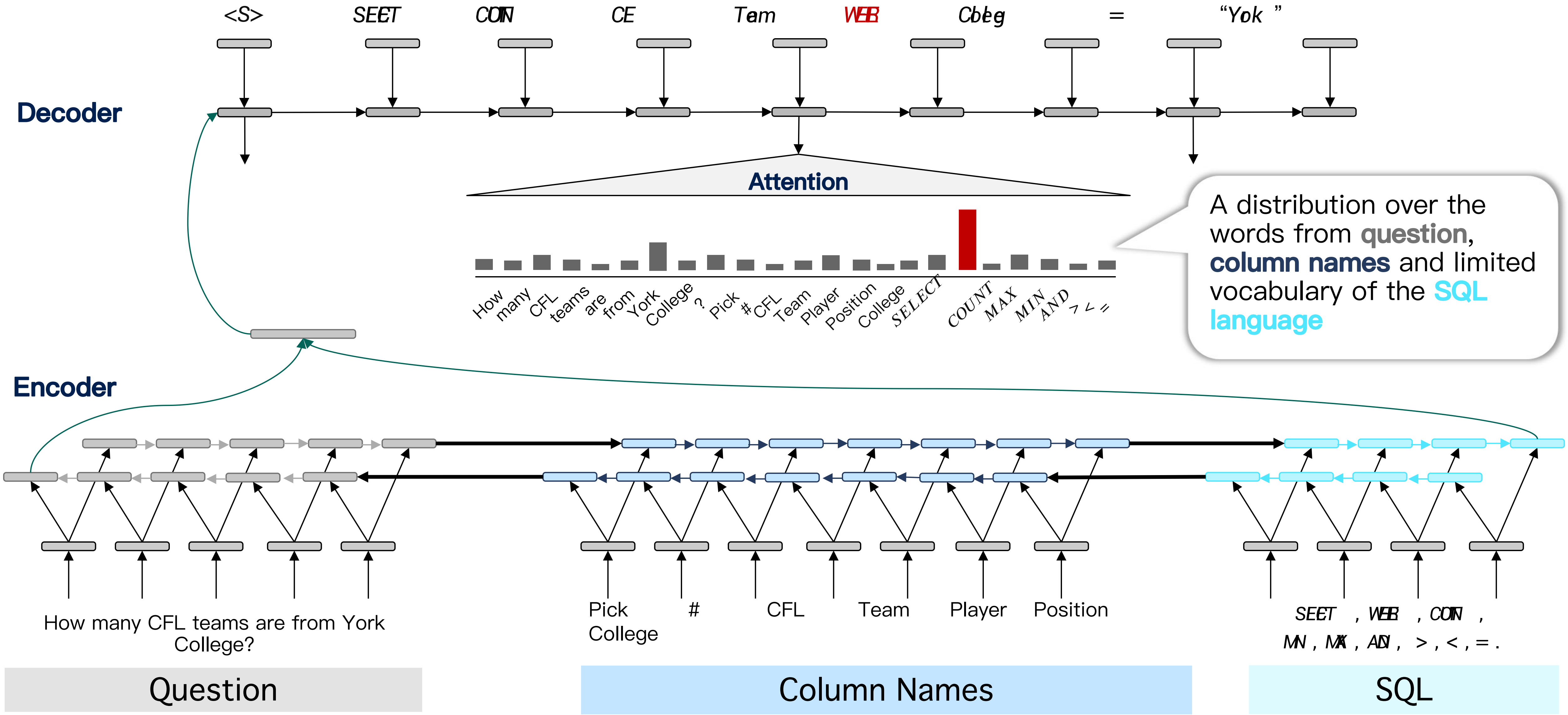
Seq2Seq with 2-layer RNN



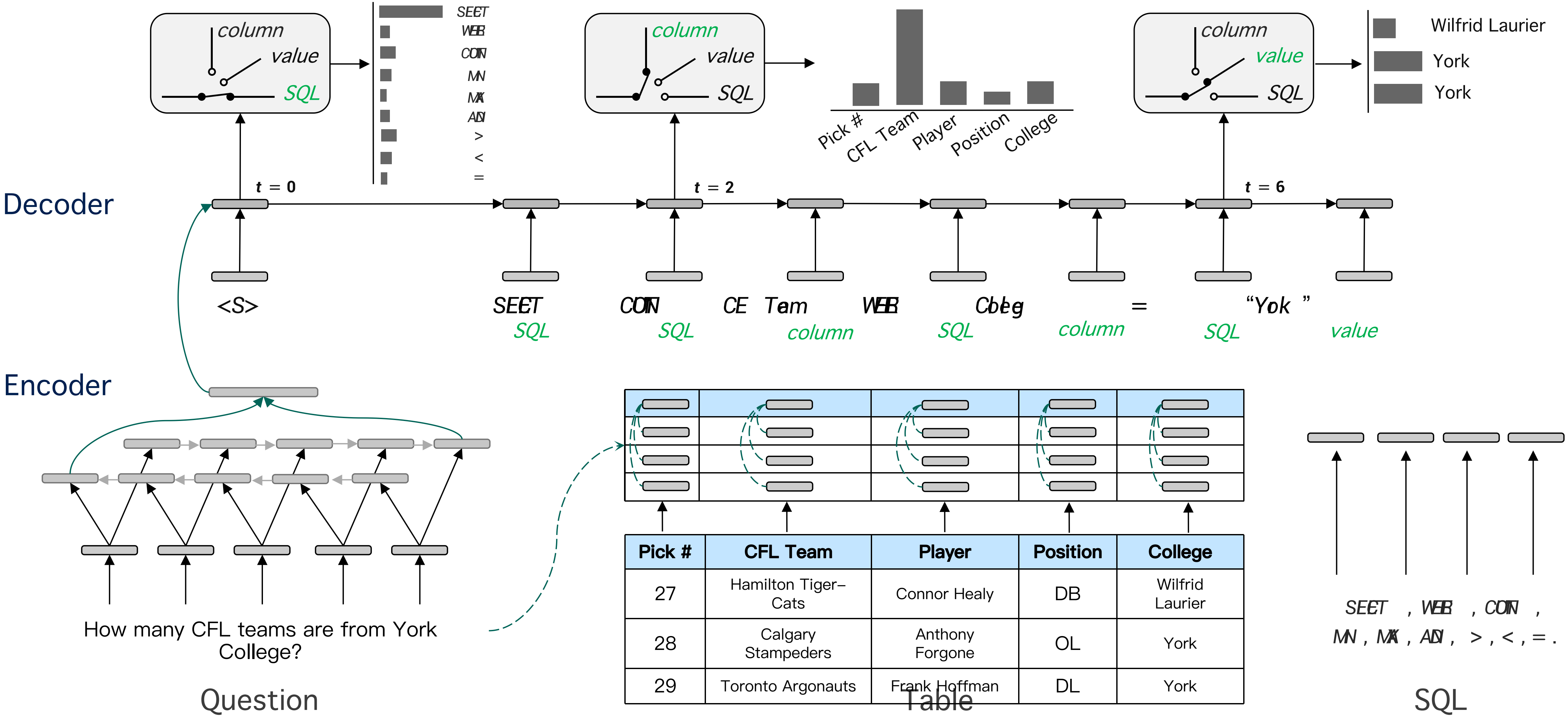
Seq2Tree with a hierarchical tree decode

Li Dong and Mirella Lapata. "*Language to Logical Form with Neural Attention.*" ACL-2016.

Seq-to-Seq with Pointer Network



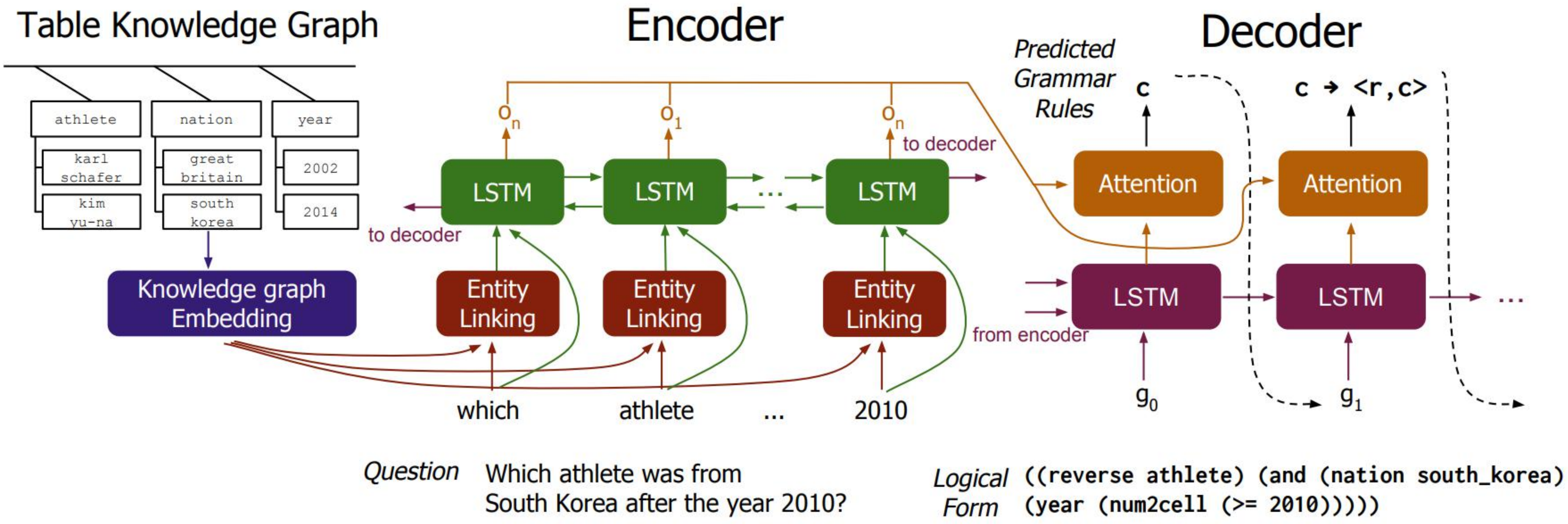
Seq–to–Seq with Structural Decoding



Pick #	CFL Team	Player	Position	College
27	Hamilton Tiger–Cats	Connor Healy	DB	Wilfrid Laurier
28	Calgary Stampeders	Anthony Forgone	OL	York
29	Toronto Argonauts	Frank Hoffman	DL	York

Yibo Sun, Duyu Tang, Nan Duan, Jianshu Ji, Guihong Cao, Xiaocheng Feng, Bing Qin, Ting Liu and Ming Zhou. "*Semantic Parsing with Syntax- and Table-Aware SQL Generation.*" ACL-2018

Seq-to-Seq with Typed Constrained Decoding



Jayant Krishnamurthy, Pradeep Dasigi, and Matt Gardner. Neural semantic parsing with type constraints for semi-structured tables. EMNLP-2017.

Action and Module

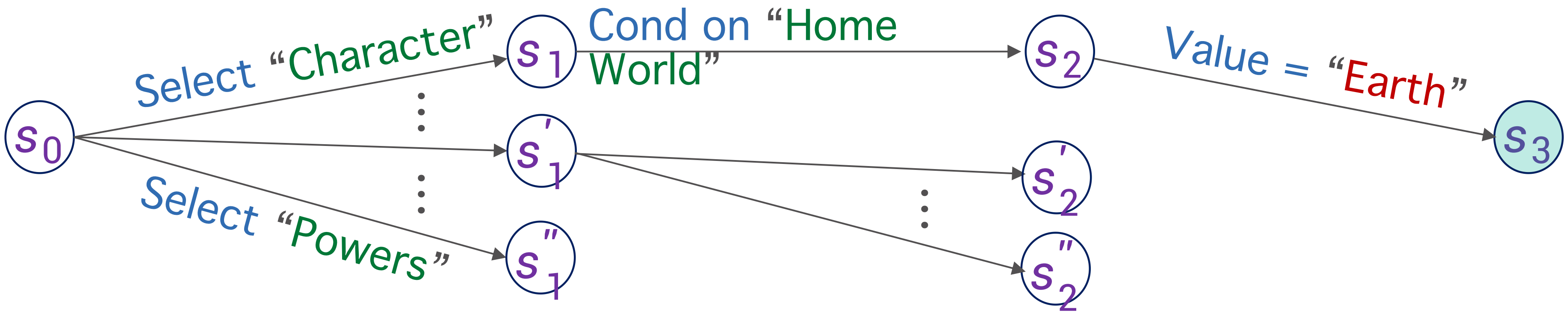
➢ *Which super heroes came from Earth and first appeared after 2009?*

```
SELECT    Character
WHERE {Home World = Earth} ∧
      {First Appeared > 2009}
```

Legion of Super Heroes Post-*Infinite Crisis*

Character	*First Appeared*	*Home World*	*Powers*
Night Girl	2007	Kathoon	Super strength
Dragonwing	2010	Earth	Fire breath
Gates	2009	Vyrga	Teleporting
XS	2009	Aarok	Super speed
Harmonia	2011	Earth	Elemental

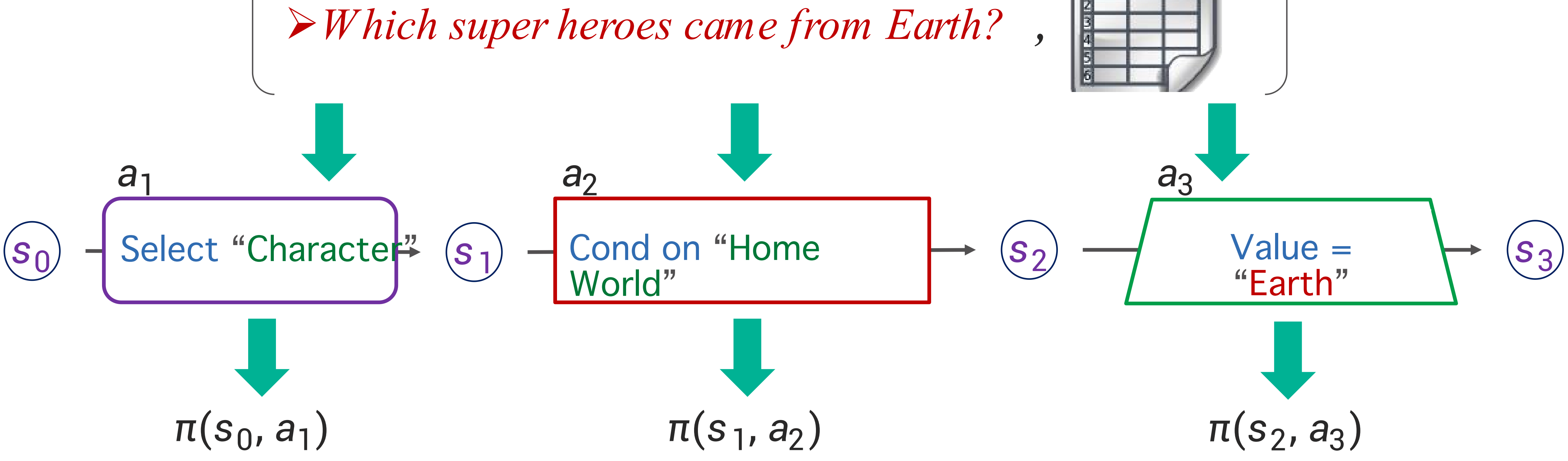
$Q =$ *"Which super heroes came from Earth?"*, $A^* = \{Dragonwing, Harmonia\}$



Mohit Iyyer, Wen-tau Yih, Ming-Wei Chang. Search-based Neural Structured Learning for Sequential Question Answering. ACL-2017.

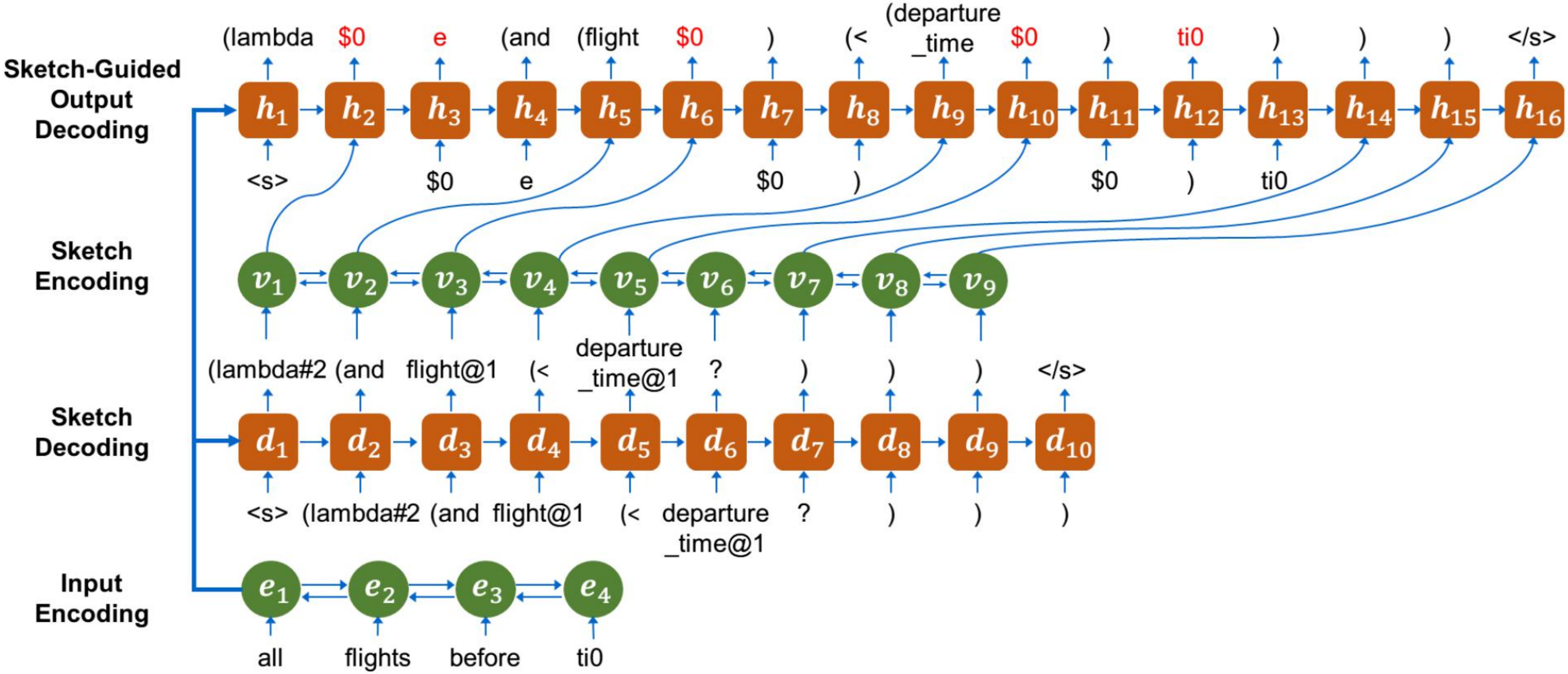
Action and Module

- The goodness of a state: $V(s_t) = V(s_{t-1}) + \pi(s_{t-1}, a_t)$, $V(s_0) = 0$
- Value of $\pi(s, a)$ is determined by a neural-network model
- Actions of the same type (e.g., `select-column`) share the same neural-network module



Mohit Iyyer, Wen-tau Yih, Ming-Wei Chang. Search-based Neural Structured Learning for Sequential Question Answering. ACL-2017.

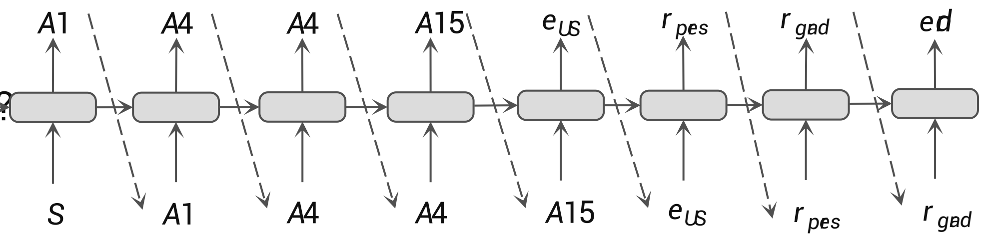
Coarse-to-Fine Decoding

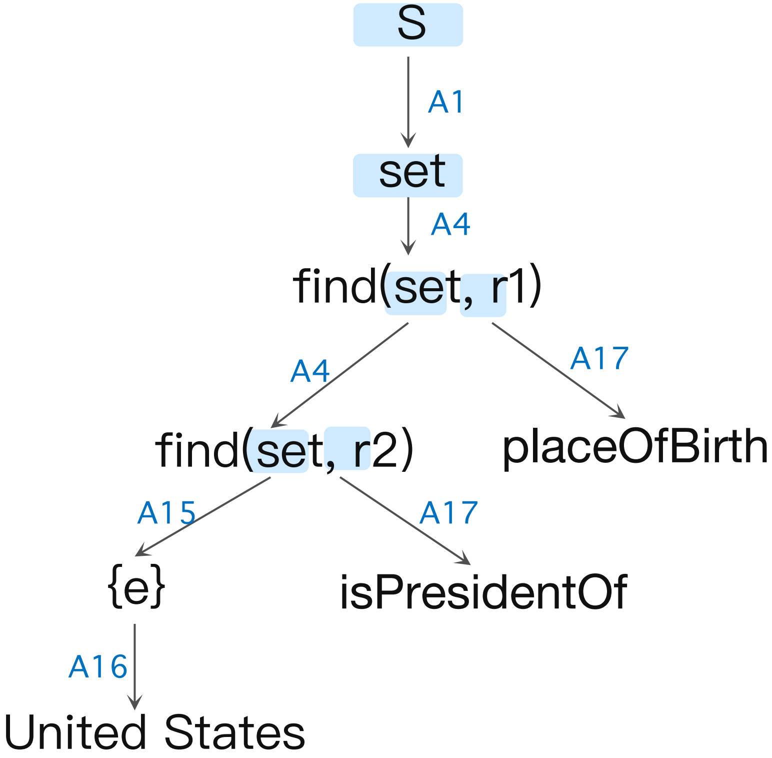


Li Dong, Mirella Lapata. Coarse-to-Fine Decoding for Neural Semantic Parsing. ACL-2018.

KBQA with Semantic Parsing (single–turn)

Where was the president of the United States born?





A1: $S \rightarrow set$

A4: $set \rightarrow find(set, r1)$

A4: $set \rightarrow find(set, r2)$

A15: $set \rightarrow \{e\}$

A16: $e \rightarrow United\ States$

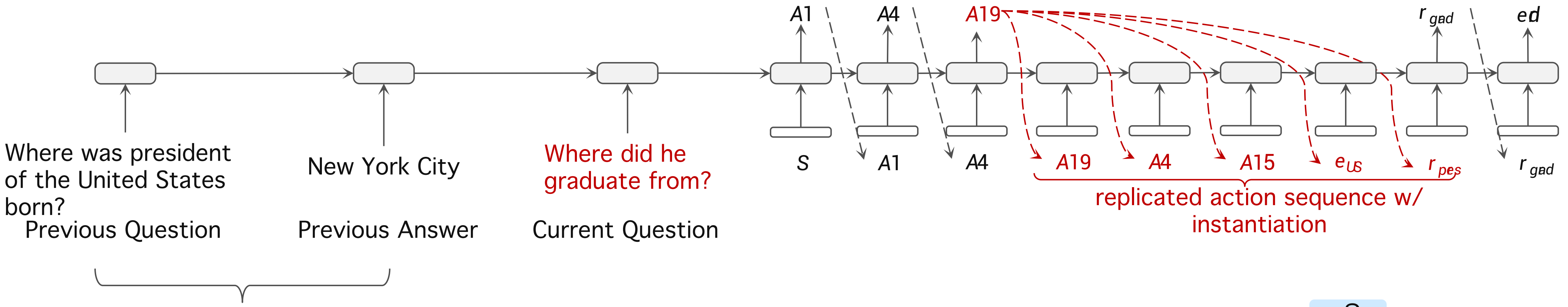
A17: $r2 \rightarrow isPresidentOf$

A17: $r1 \rightarrow placeOfBirth$

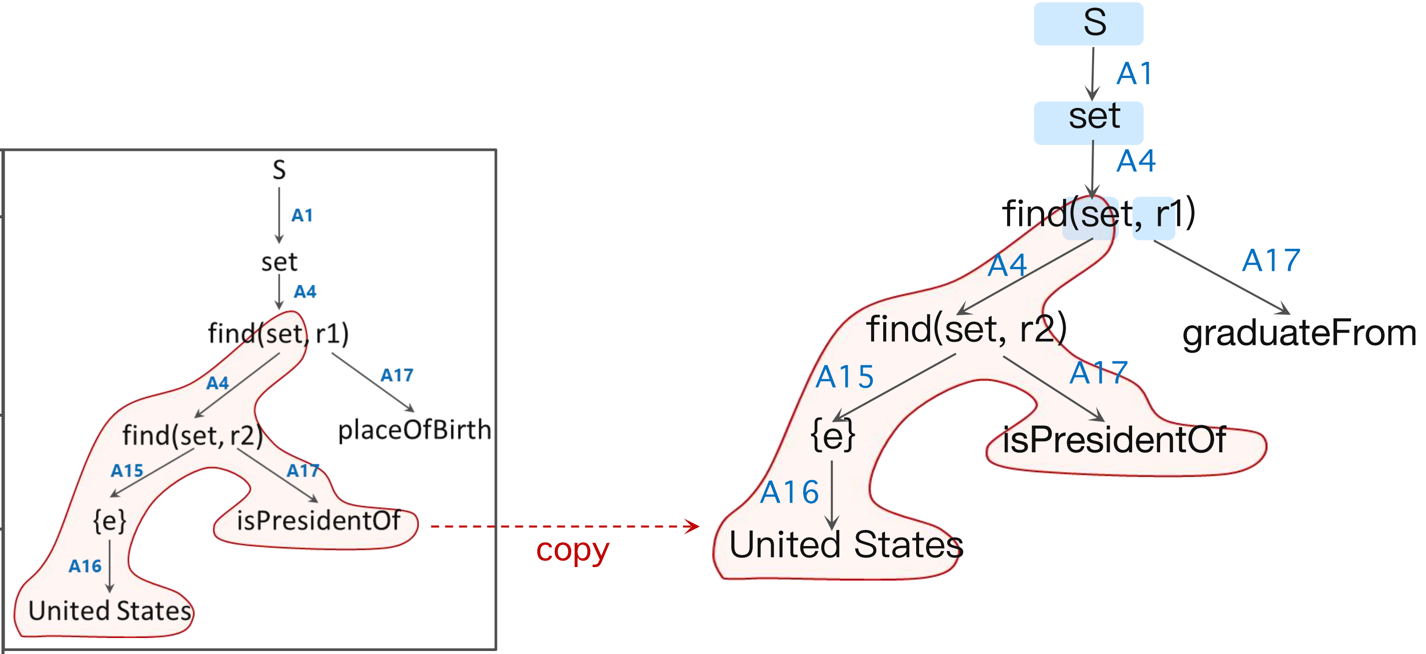
Action	Operation	Description
A1-A3	S—>Set | Num | Bool	S is start symbol
A4	Set—>Find(R, E)	Set of entities with a r edge to e
A5	Num—>Count(Set)	Total number of set
A6	Bool—>(∈, E, Set)	Whether E∈Set
A7	Set—>Set ∪ Set	Union of Sets
A8	Set—>Set ∩ Set	Intersection of Sets
A9	Set—>Set —Set	Difference of Sets
A10	Set->larger(set, r, num)	Entity from set linking to more than num entities with relation r
A11	Set->less(set, r, num)	Entity from set linking to less than num entities with relation r
A12	Set->equal(set, r, num)	Entity from set linking to num entities with relation r
A13	Set->arxmax(set, r, num)	Entity from set linking to most entities with relation r
A14	Set->argmin(set, r, num)	Entity from set linking to least entities with relation r
A15	Set->{e}	
A16-A18	e | r | num -> constant	instantiation for entity e, predicate r or number num
A19-A21	Set | Num | Bool ->action(i-1)	Replicate previous operation sequence

Daya Guo, Duyu Tang, Nan Duan, Ming Zhou and Jian Yin. "***Dialog-to-Action: Conversational Question Answering Over a Large-Scale Knowledge Base.***" NeurIPS-2018.

KBQA with Semantic Parsing (multi-turn)

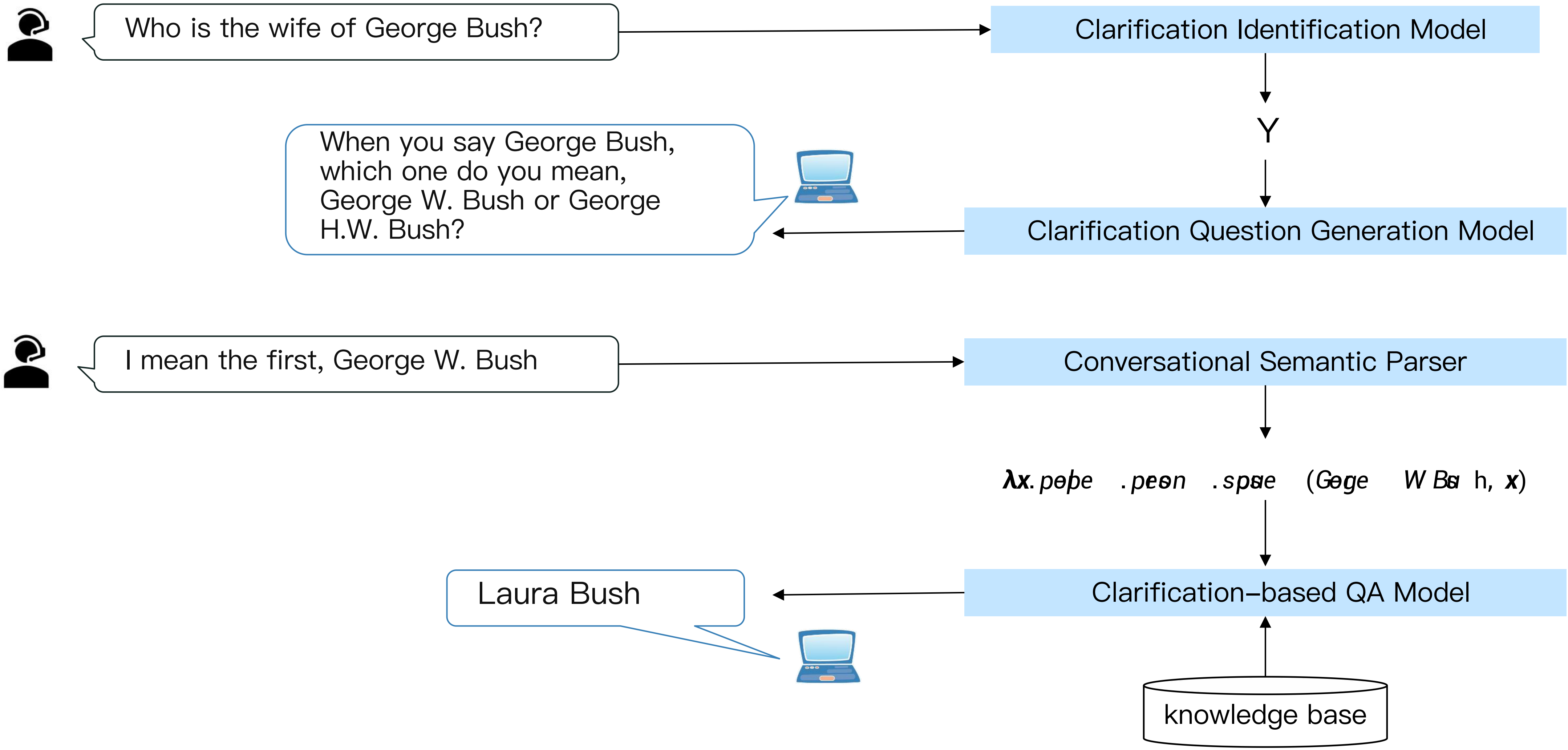


Dialog Memory	
Entity	{United States, tag=utterance} {New York City, tag=answer}
Predicate	{isPresidentOf} {placeOfBirth}
Action Subsequence	set → A4 A15 e_{US} r_{pes} set → A4 A15 set → A4 A4 A15 e_{US} r_{pes} r_{bth} set → A4 A4 A15



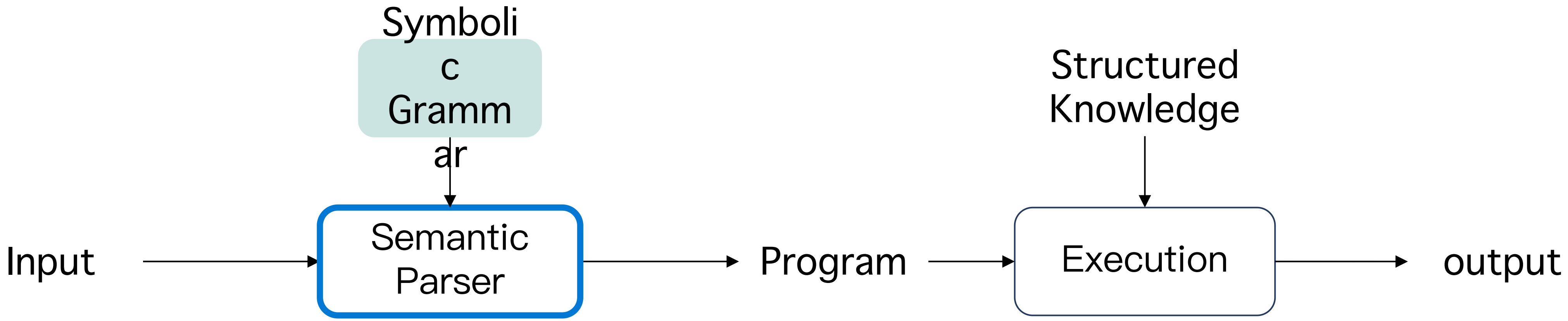
Daya Guo, Duyu Tang, Nan Duan, Ming Zhou and Jian Yin. "***Dialog-to-Action: Conversational Question Answering Over a Large-Scale Knowledge Base.***" NeurIPS-2018.

Conversational Question Answering

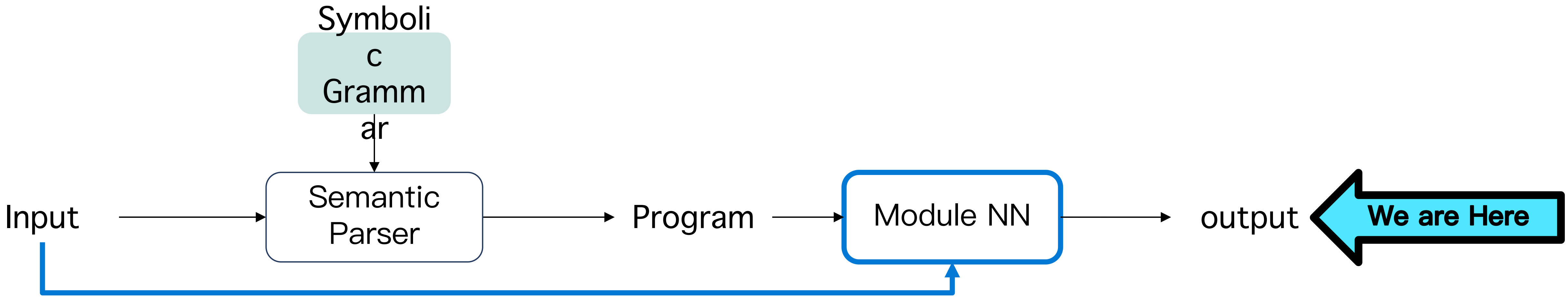


Jingjing Xu, Yuechen Wang, Duyu Tang, Nan Duan, Pengcheng Yang, Qi Zeng, Ming Zhou, Xu Sun. "*Asking Clarification Questions in Knowledge–Based Question Answering.*" EMNLP–2019.

Outline



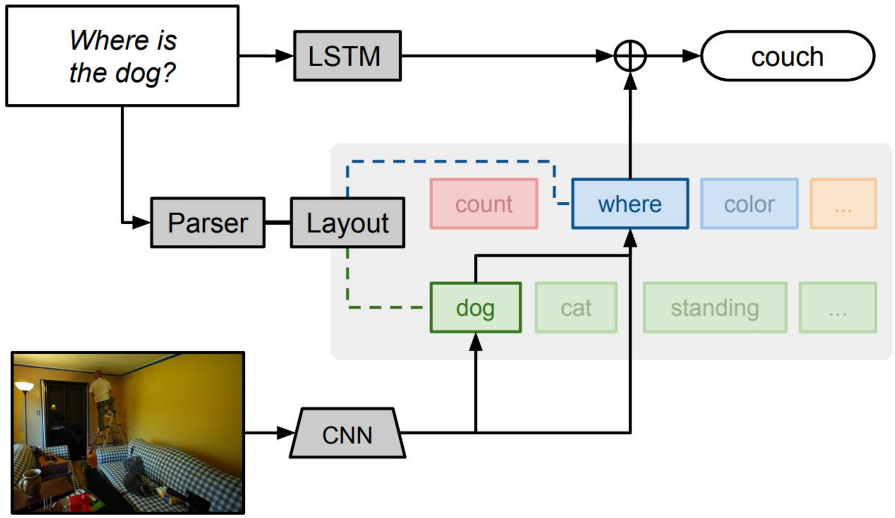
Part 1: Semantic Parsing



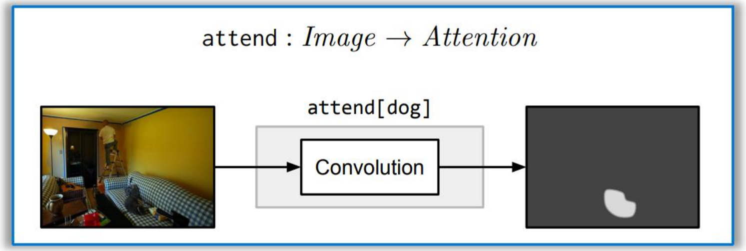
Part 2: Module Network

Module Network

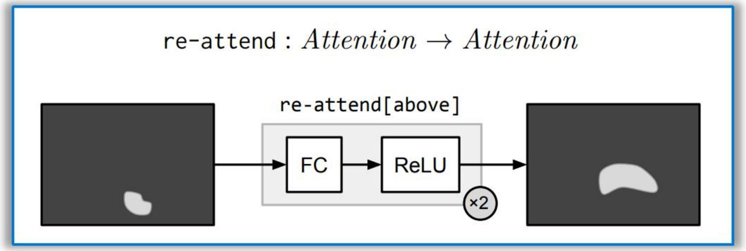
Reusable neural modules with different architectures



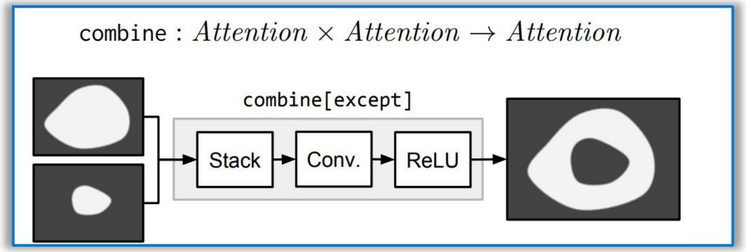
Attention



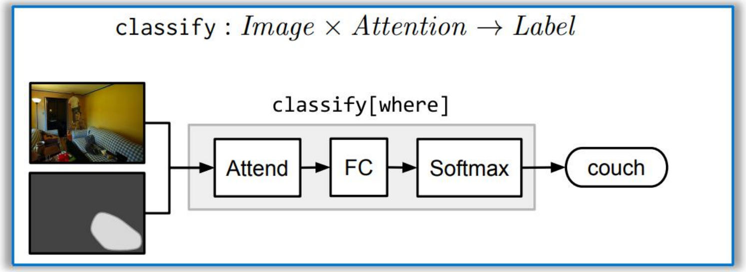
Re–Attention



Combination



Classification



Jacob Andreas, Marcus Rohrbach, Trevor Darrell, and Dan Klein.
Neural module networks. CVPR-2016.

Performance V.S. Interpretability

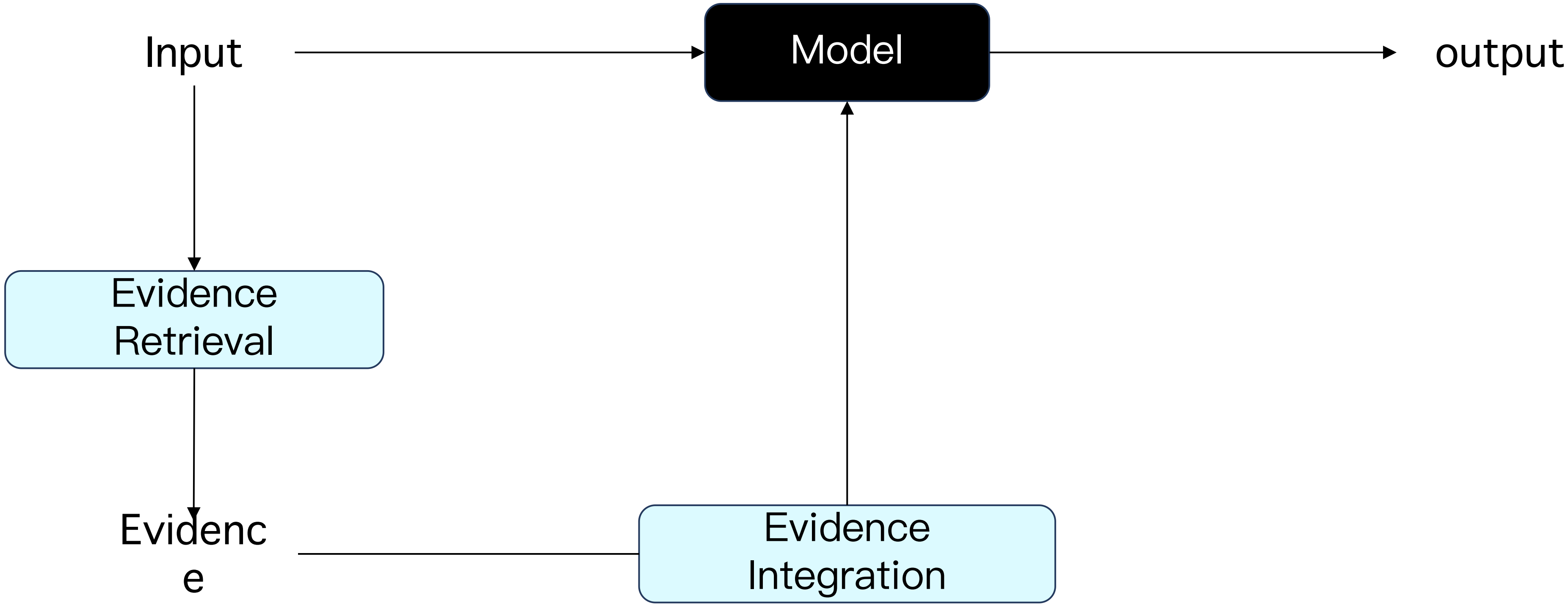
- Semantic Parser
  - Good interpretability, good performance on limited applications
  - The extension of grammar to open domain is challenging

- Module Network
  - Good performance with neural models as backbone, limited interpretability
  - Moderate interpretability, compared with semantic parser.
  - The definition of grammar is typically task-specific, and manually designed by experts

Evidence-based Models in NLP

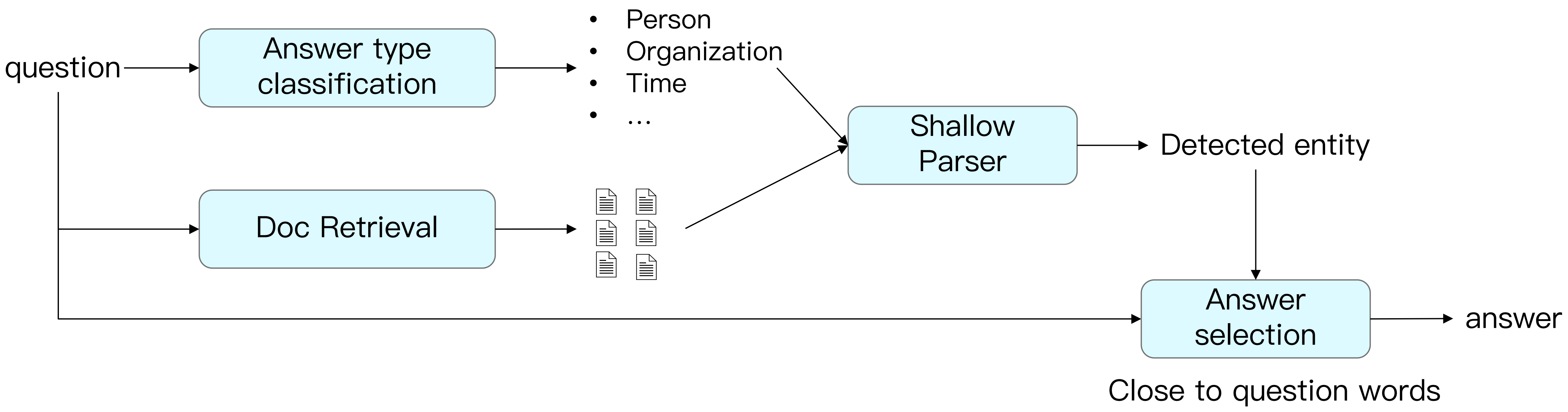
General Framework

- Consider evidence as an additional input of the model



Open Question Answering

- First large-scale evaluation of domain-independent QA systems.
- Participants were given 200 fact-based, short-answer questions
- Each question was guaranteed to have at least one document in the collection that explicitly answered the question.
- Participants returned a ranked list of [document-id, answer-string] pairs per question such that each answer string was believed to contain an answer to the question.

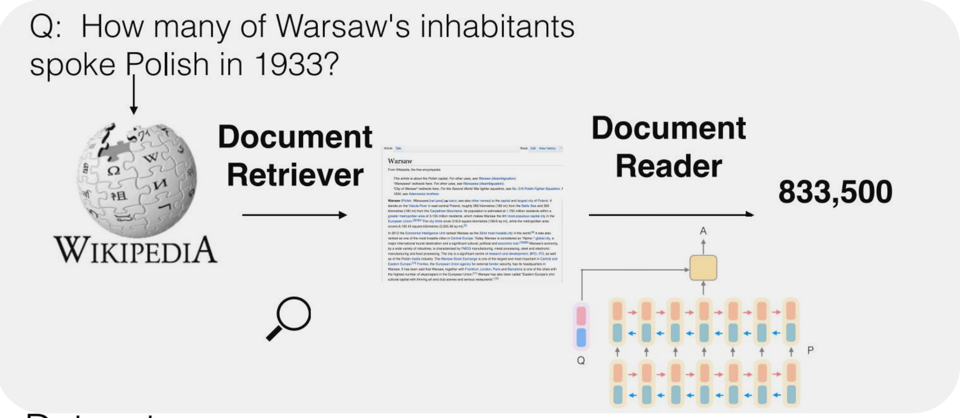


Ellen M Voorhees. **The TREC-8 Question Answering Track Report**. TREC, 1999

Sparse Retrieval Model (DrQA)

Document Retriever + Document Reader

- Document retriever: finding relevant articles from 5 million Wikipedia articles
- Document reader (reading comprehension system): identifying the answer spans from those articles



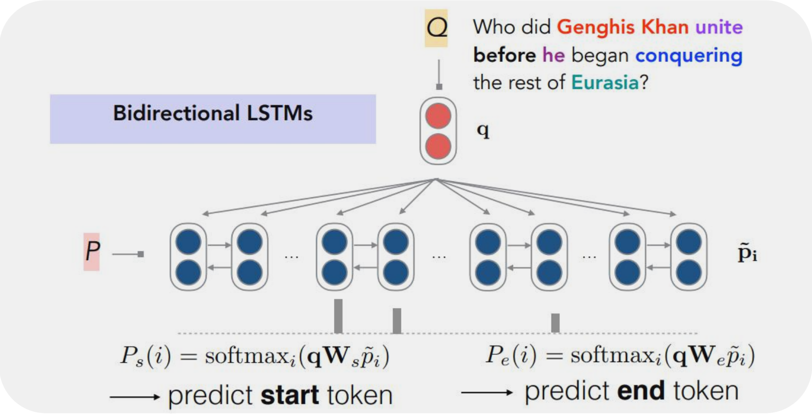
- Datasets:
  - SQuAD (Rajpurkar et al, 2016)
  - TREC (Baudiš and Šedivý, 2005)
  - WebQuestions Freebase (Berant et al, 2013)
  - WikiMovies (Miller et al, 2016)

Document Retriever

TF-IDF bag-of-words vectors + efficient bigram hashing (Weinberger et al., 2009)

Document Reader

Task: given paragraph P and question Q, the goal is to find a span A in the paragraph which answers the question.
Model: similar to AttentiveReader (Hermann et al, 2015; Chen et al, 2016). We aim to keep it **simple**!



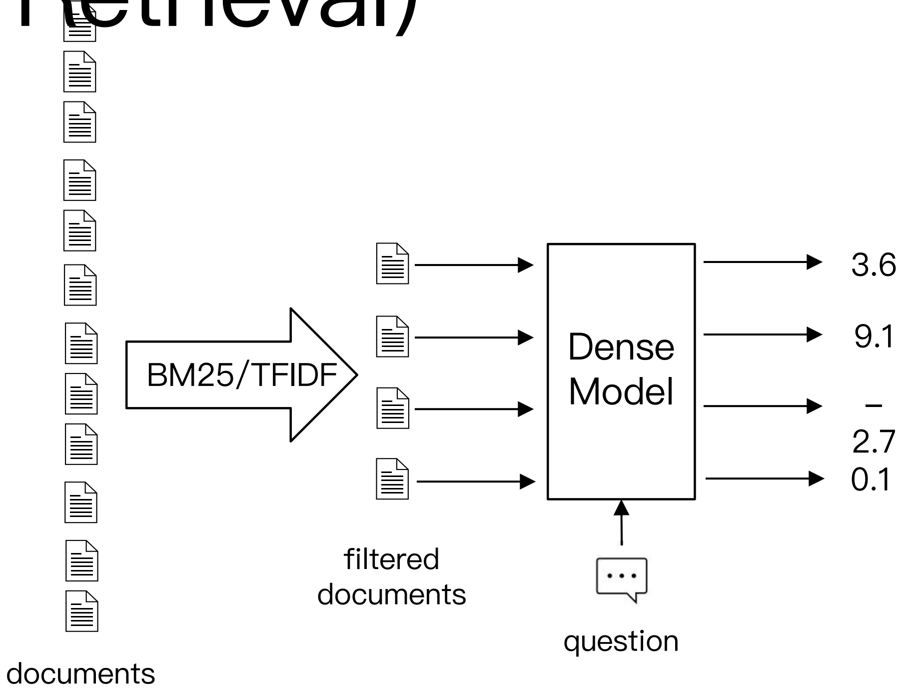
The input vectors consist of:
- Word embeddings
- Exact match features: whether the word appears in question
- Token features: POS, NER, term frequency
- Aligned question embedding

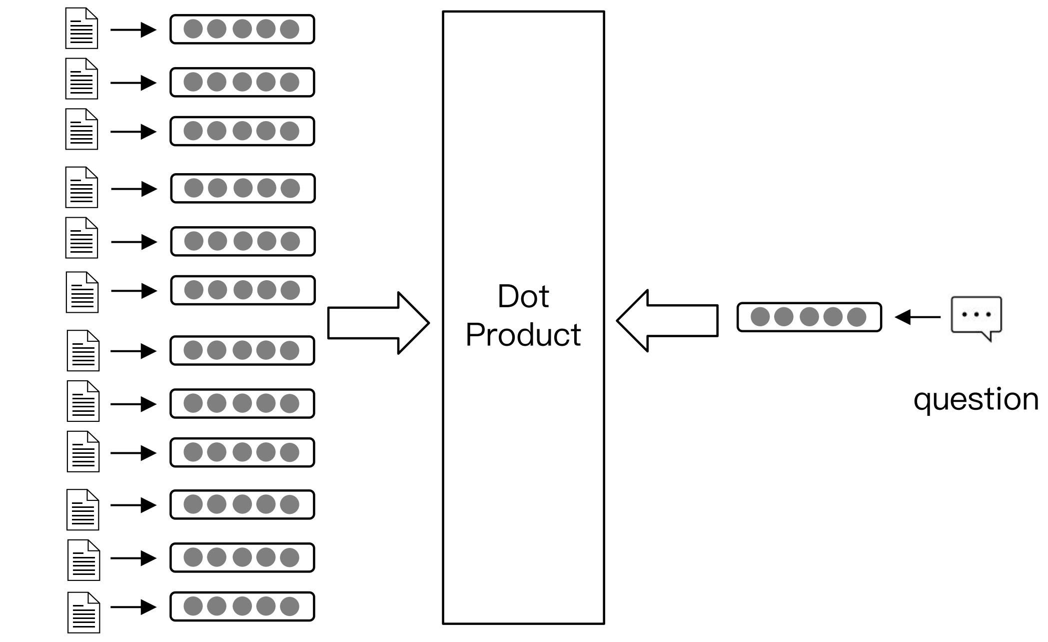
Data: SQuAD + **Distantly Supervised** Data

(Q, A) $\longrightarrow$ (P, Q, A) if P is retrieved and A can be found in P

Danqi Chen, Adam Fisch, Jason Weston, Antoine Bordes. "*Reading Wikipedia to Answer Open-Domain Questions.*" ACL-2017.

Dense Retrieval Model (DPR, Dense Passage Retrieval)





$$\mathrm{sim}(q,p) = E_Q(q)^{\intercal} E_P(p)$$

$$L(q_i, p_i^+, p_{i,1}^-, \cdots, p_{i,n}^-) = -\log \frac{e^{\mathrm{sim}(q_i,p_i^+)}}{e^{\mathrm{sim}(q_i,p_i^+)} + \sum_{j=1}^{n} e^{\mathrm{sim}(q_i,p_{i,j}^-)}}$$

Who is the bad guy in lord of the rings?

Sala Baker is an actor and stuntman from New Zealand. He is best known for portraying the villain Sauron in the Lord of the Rings trilogy..

Vladimir Karpukhin, Barlas Oğuz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, Wen-tau Yih. **Dense Passage Retrieval for Open-Domain Question Answering**. Arxiv-2020

DPR Results

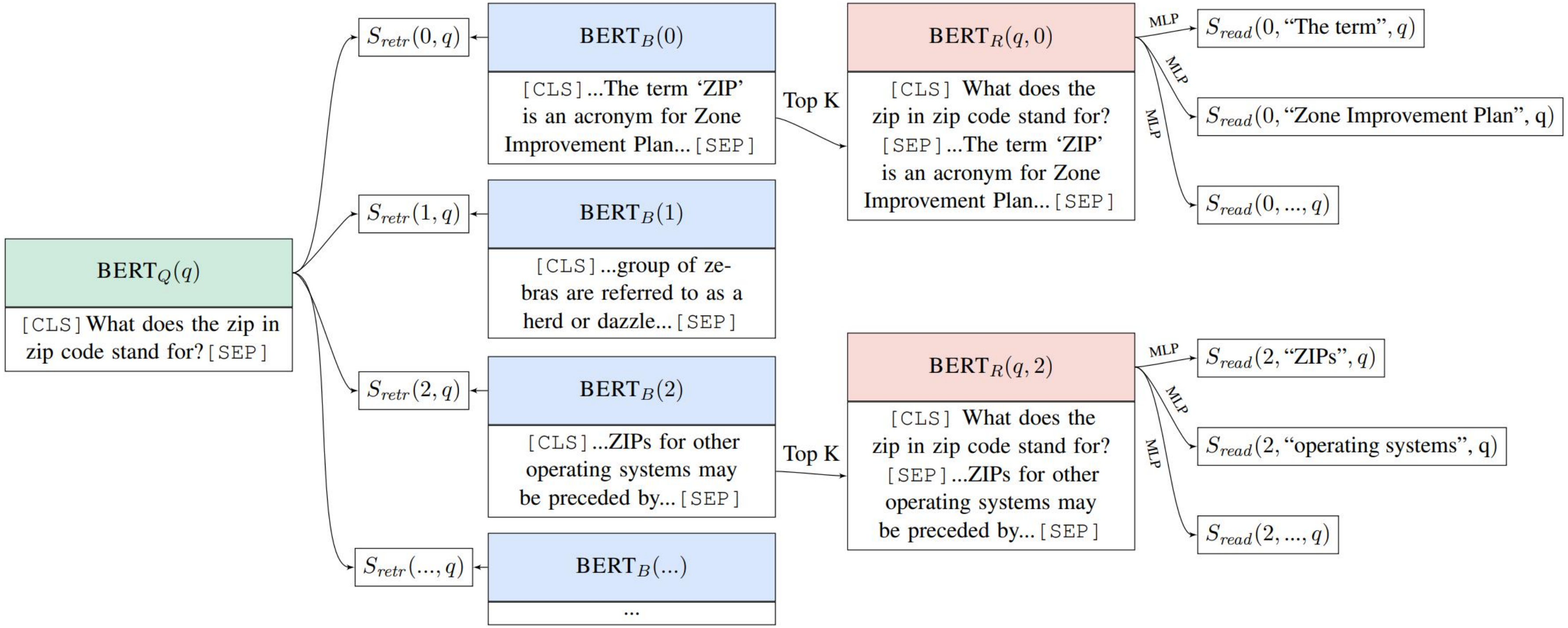
Training	Model	NQ	TriviaQA	WQ	TREC	SQuAD
Single	BM25+BERT (Lee et al., 2019)	26.5	47.1	17.7	21.3	33.2
Single	ORQA (Lee et al., 2019)	33.3	45.0	36.4	30.1	20.2
Single	HardEM (Min et al., 2019a)	28.1	50.9	-	-	-
Single	GraphRetriever (Min et al., 2019b)	34.5	56.0	36.4	-	-
Single	PathRetriever (Asai et al., 2020)	32.6	-	-	-	**56.5**
Single	$\text{REALM}_{\text{Wiki}}$ (Guu et al., 2020)	39.2	-	40.2	46.8	-
Single	$\text{REALM}_{\text{News}}$ (Guu et al., 2020)	40.4	-	40.7	42.9	-
Single	BM25	32.6	52.4	29.9	24.9	38.1
	DPR	**41.5**	56.8	34.6	25.9	29.8
	BM25+DPR	39.0	57.0	35.2	28.0	36.7
Multi	DPR	**41.5**	56.8	**42.4**	49.4	24.1
	BM25+DPR	38.8	**57.9**	41.1	**50.6**	35.8

Table 4: End-to-end QA (Exact Match) Accuracy. The first block of results are copied from their cited papers. $\text{REALM}_{\text{Wiki}}$ and $\text{REALM}_{\text{News}}$ are the same model but pretrained on Wikipedia and CC-News, respectively. *Single* and *Multi* denote that our Dense Passage Retriever (DPR) is trained using individual or combined training datasets (all except SQuAD). For WQ and TREC in the *Multi* setting, we fine-tune the reader trained on NQ.

Vladimir Karpukhin, Barlas Oğuz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, Wen-tau Yih.
Dense Passage Retrieval for Open-Domain Question Answering. Arxiv-2020

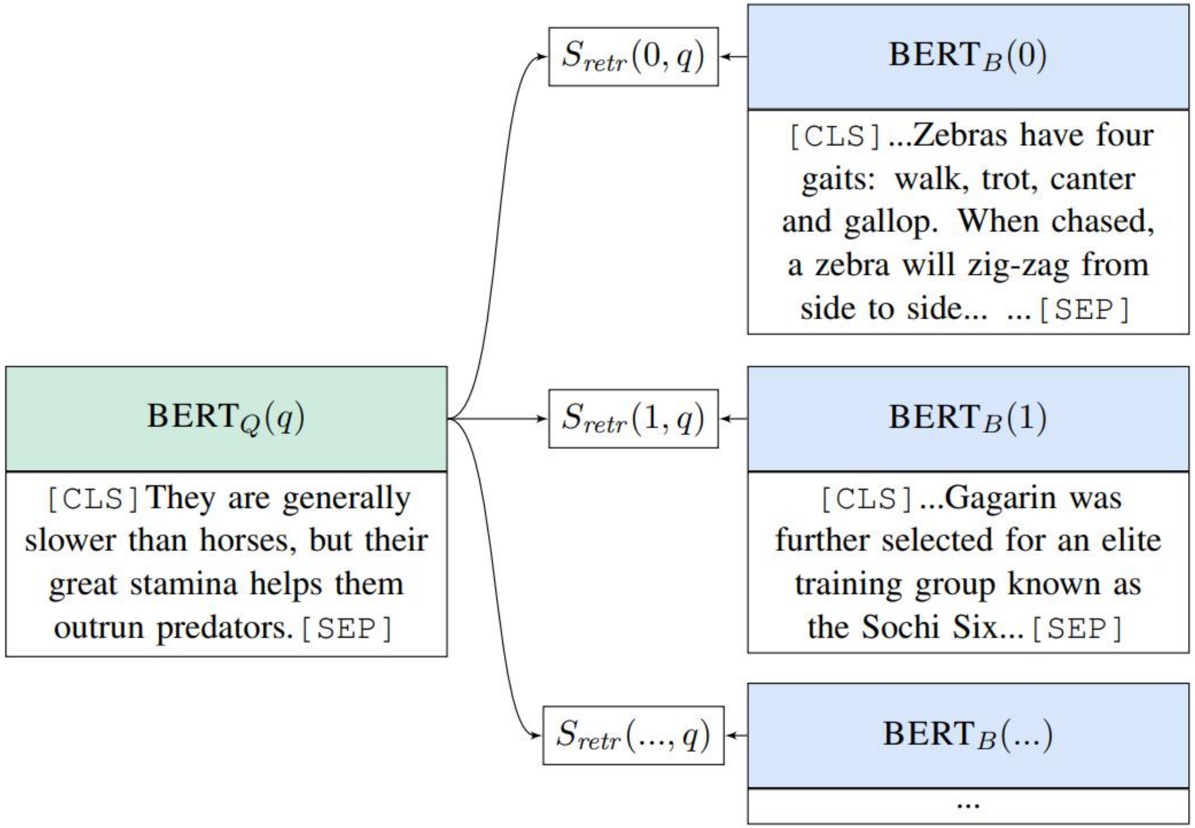
Joint Retrieval and Reader

- ORQA: Open-Retriever Question Answering
  - jointly learn the retriever and reader from question-answer string pairs
  - pre-train the retriever with an Inverse Cloze Task.



Kenton Lee, Ming-Wei Chang, Kristina Toutanova. "*Latent Retrieval for Weakly Supervised Open Domain Question Answering.*" ACL-2019

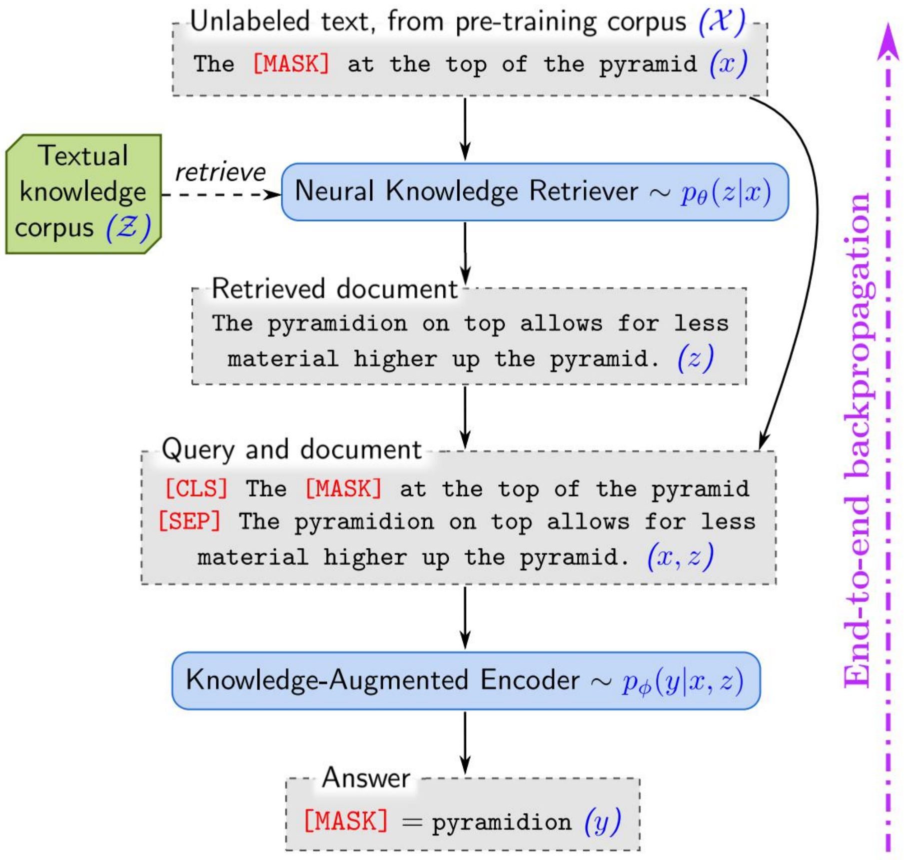
Pre-train Retrieval Model with Inverse Cloze Task



- In ICT, a sentence is treated as a pseudo-question, and its context is treated as pseudo-evidence.
- Given a pseudo-question, ICT requires selecting the corresponding pseudo-evidence out of the candidates in a batch.

Kenton Lee, Ming-Wei Chang, Kristina Toutanova. "*Latent Retrieval for Weakly Supervised Open Domain Question Answering.*" ACL-2019

Pre-train Retrieval Model with REALM



Knowledge Retriever: dense inner product model

$$p(z \mid x) = \frac{\exp f(x, z)}{\sum_{z'} \exp f(x, z')},$$

$$f(x, z) = \texttt{Embed}_{\texttt{input}}(x)^\top \texttt{Embed}_{\texttt{doc}}(z),$$

$$\texttt{Embed}_{\texttt{input}}(x) = \mathbf{W}_{\texttt{input}} \mathrm{BERT}_{\texttt{CLS}}(\texttt{join}_{\texttt{BERT}}(x))$$

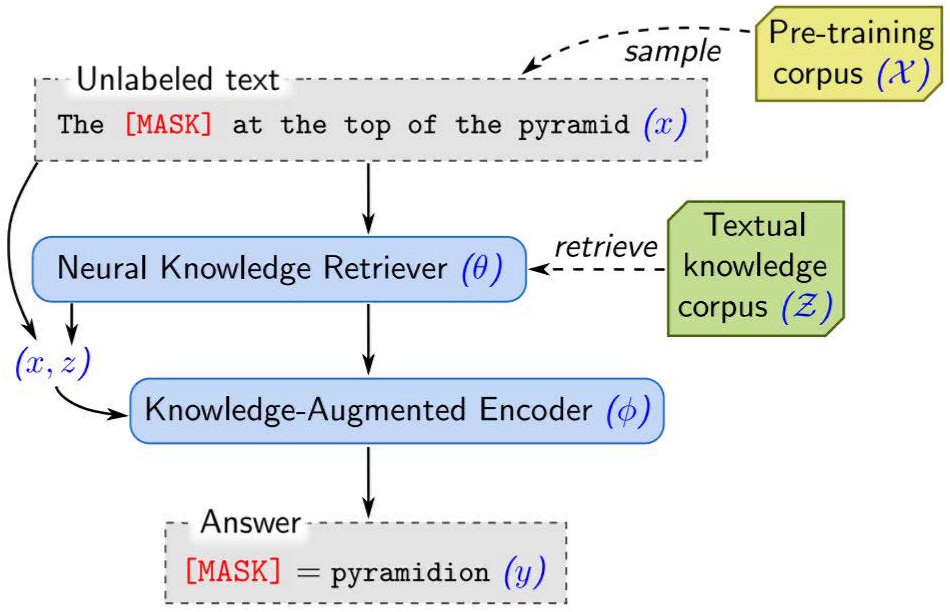
$$\texttt{Embed}_{\texttt{doc}}(z) = \mathbf{W}_{\texttt{doc}} \mathrm{BERT}_{\texttt{CLS}}(\texttt{join}_{\texttt{BERT}}(z_{\text{title}}, z_{\text{body}}))$$

$$\texttt{join}_{\texttt{BERT}}(x) = \texttt{[CLS]}\, x\, \texttt{[SEP]}$$

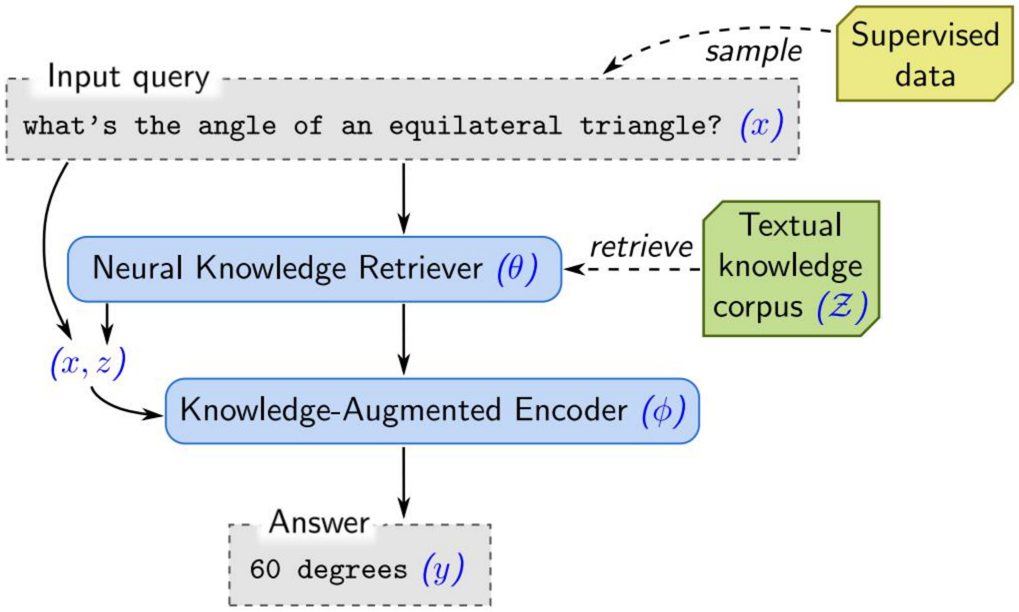
$$\texttt{join}_{\texttt{BERT}}(x_1, x_2) = \texttt{[CLS]}\, x_1\, \texttt{[SEP]}\, x_2\, \texttt{[SEP]}$$

Realm: Retrieval-augmented language model pre-training
K Guu, K Lee, Z Tung, P Pasupat, MW Chang - arXiv :2002.08909

Pre-train Retrieval Model with REALM



Unsupervised pre-training



Supervised fine-tuning

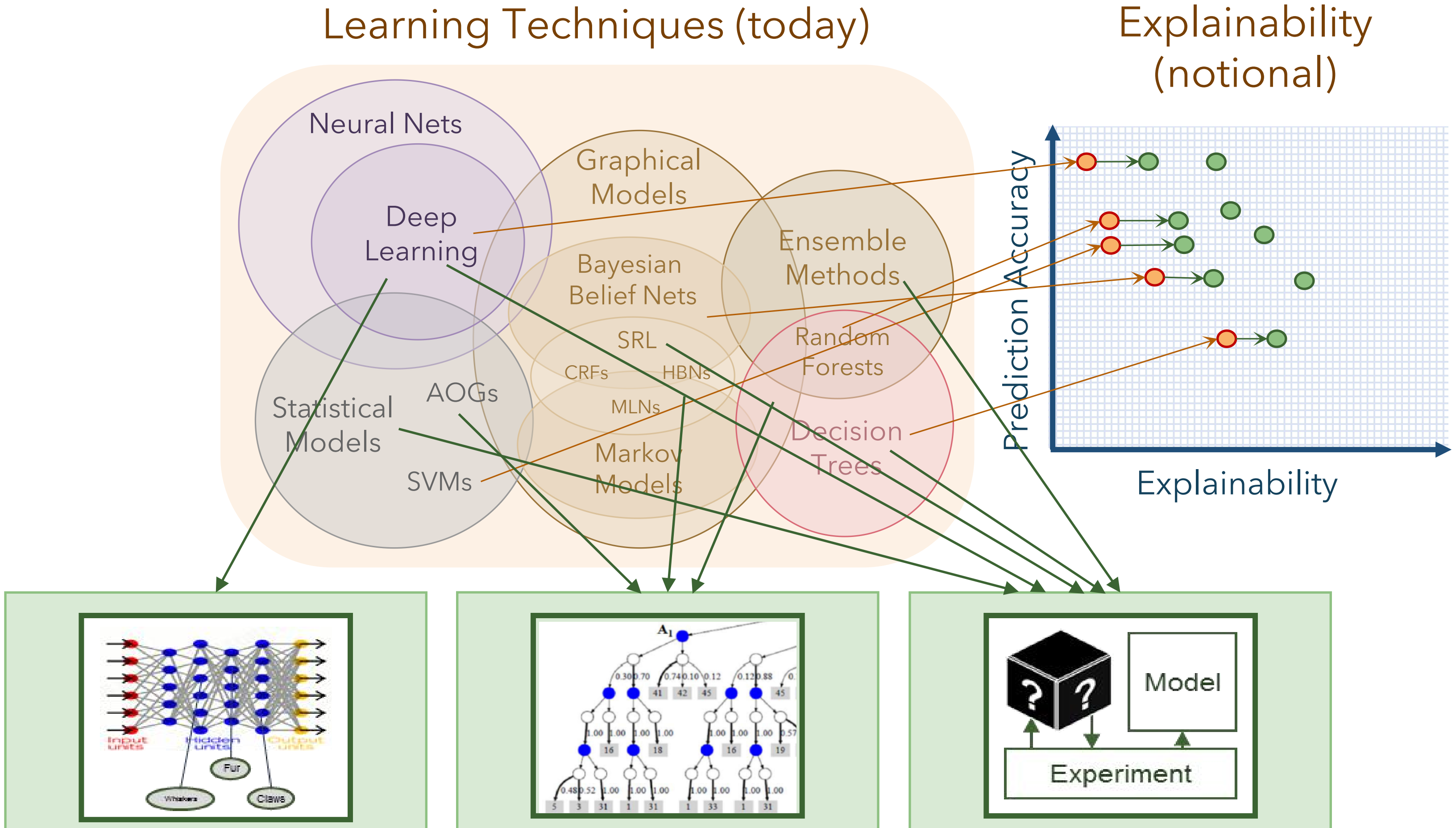
Realm: Retrieval-augmented language model pre-training
K Guu, K Lee, Z Tung, P Pasupat, MW Chang - arXiv :2002.08909

Summary

- Topics covered by this talk
  - Logic-based Models in NLP
  - Neural-Symbolic Models in NLP
  - Evidence-based Models in NLP

- Directions worth pursuing
  - Interpretable models and methods
  - Deep understanding with reasoning ability

Challenge: Performance vs. Explainability



Deep Explanation
Modified deep learning techniques to learn explainable features

Interpretable Models
Techniques to learn more structured, interpretable, causal models

Model Induction
Techniques to infer an explainable model from any model as a black box

Explainable Artificial Intelligence (XAI), David Gunning, DARPA/I2O

We are hiring!

- Both interns and employees.

- Topics:
  - Semantic Parsing
  - Code Intelligence
  - Machine Reasoning

- Send email to dutang@microsoft.com