Microsoft Research
微软亚洲研究院

Answering Questions with Complex Semantic Constraints on Open Knowledge Bases

Pengcheng Yin[#] Nan Duan[$] Ben Kao[#] Junwei Bao[$] Ming Zhou[$]

[#] The University of Hong Kong [$] Microsoft Research Asia

Speaker: Pengcheng Yin

pcyin@cs.hku.hk

The University of Hong Kong

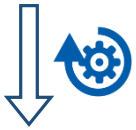
Agenda

- **Background & Motivation**
  - Knowledge-based Question Answering (KB-QA) systems
  - How to answer *complex* questions?
  - Problems with the state-of-the-arts
  - Our contribution
- Approach
- Results
- Summary

Knowledge-based Question Answering (KB-QA)

- Task: Answer factoid questions in natural languages

Who wrote Harry Potter?



Transformation

```
Select ?author.
Where{
    Harry_Potter written_by  ?x,
    ?x           name        ?author
}
```

Sparql query

Execution

Knowledge Bases (KBs)

J. K. Rowling

1. Transform questions into structured queries
2. Execute the query against the KB to retrieve answers

How to answer *complex* questions?

- Many existing KB-QA systems focus on answering questions with simple semantic constraints:

Q_1: What $[is\ the\ currency\ of]_{rel}$ Spain?

Answers: {  ,  }

Euro *Peseta*

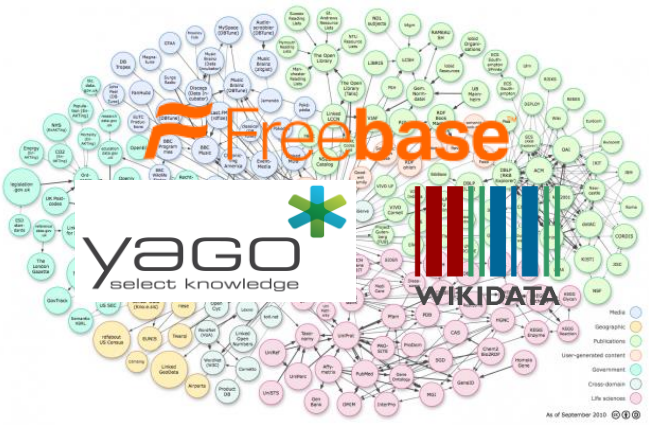
- How to answer questions with *complex* semantic constraints?

Q_2: What $[was\ the\ currency\ of]_{rel}$ Spain *before 2002*?




Expressed via prepositional/ adverbial modifiers

Two Families of KBs



curated KBs

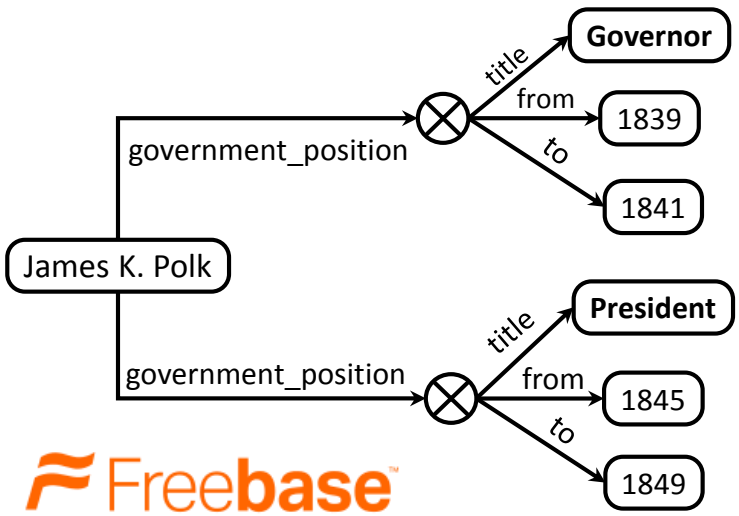
Subject	*Relation Phrase*	*Arguments*
James K. Polk	was	a governor; before he was president
the currency of Spain	was	the Peseta; before 2002
Peseta	was replaced	by Euro; as official tender of Spain; in 2002
Barack Obama	graduated	from Harvard Law School; in 1979 and 1991
Obama	graduated	magna cum laude; from Harvard Law School; in 1991
Barack Obama	attended	Harvard Law School

open KBs

Curated KBs

- Curated KBs
  - Manually created, structured KBs based on predefined schema
  - Accurate and precise
- Problem: non-trivial to transform complex questions to structured queries!

what was James K. Polk before he was president?



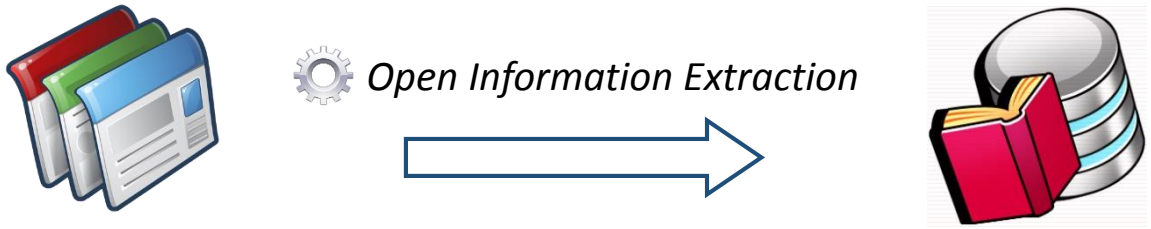

```
SELECT ?job_title.
FROM Freebase
WHERE{
   James K. Polk  government_position  ?job.
   ?job           title               ?job_title.

   ?job           to                  ?to_date.

   FILTER(?to_date < (
      SELECT ?start_date.
      WHERE{
         James K. Polk government_position ?job1.
         ?job1         title               President.
         ?job1         from                ?start_date.
      }
   ))
}
```

Open KBs

- open KBs [Galárraga *et al.*, 2014; Fader *et al.*, 2014]: Open domain KBs automatically extracted using open Information Extraction (IE) techniques



(massive) Web Docs

Open Knowledge Base

Peseta was replaced by Euro as official tender of Spain in 2002.

Open Information Extraction

n-tuple assertion

Subject	***Relation Phrase***	***Arguments***	***Confidence***
Peseta	was replaced	by Euro; as official tender of Spain; in 2002	0.89

- Each assertion has a subject field, a relation phrase field, and multiple argument fields

Different arguments are separated by semicolons

[Galárraga *et al.*, 2014; Fader *et al.*, 2014]

Open KBs

Open KB is a large collection of n-tuple assertions

Subject	*Relation Phrase*	*Arguments*	*Freq.*	*Conf.*
James K. Polk	was	a governor; before he was president	2	0.87
the currency of Spain	was	the Peseta; before 2002	3	0.95
Peseta	was replaced	by Euro; as official tender of Spain; in 2002	3	0.81
Barack Obama	graduated	from Harvard Law School; in 1979 and 1991	4	0.77
Obama	graduated	magna cum laude; from Harvard Law School; in 1991	5	0.93
Barack Obama	attended	Harvard Law School	3	0.90

- Knowledge is modeled as n-tuple ($n \geq 3$) assertions
- Open KBs are unnormalized

Open KB-QA and Open KBs

- n-tuple assertions contain rich semantic information
- Naturally handles complex questions

What was James K. Polk before he was president?

⇩ ***Query Transformation***

n-tuple query *<James K. Polk; was; ?x, before he was president>*

⇩ ***Querying***

Open KB assertions

Subject	*Relation Phrase*	*Arguments*	*Freq.*	*Confidence*
James K. Polk	was	a governor; **before he was a president**	2	0.93
Peseta	was replaced	by Euro; as official tender of Spain; in 2002	3	0.89

⇩ ***Answer Extraction***

Governor

[Fader *et al.*, 2014]

Existing Open KB-QA systems

- Existing open KB-QA systems are designed to work on open KBs of *triplet* assertions (with *single* argument)
  - Triplet assertions are NOT semantically rich!

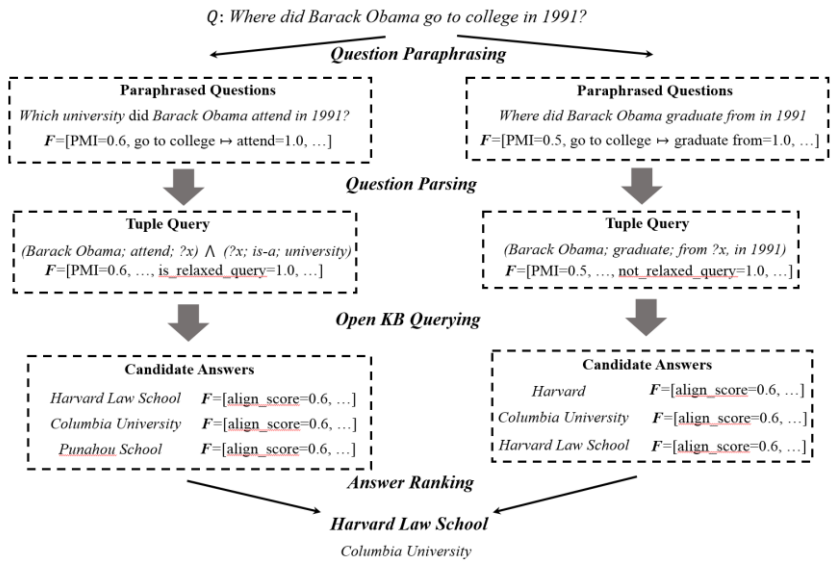
Subject	***Relation Phrase***	***Argument***	***Freq.***	***Confidence***
James K. Polk	was	**a governor**	2	0.93
Peseta	was replaced	by Euro	3	0.89

  - Cannot handle complex questions
    - *What was James K. Polk before he was president?*
    - *What was the currency of Spain before 2012?*
  - It's non-trivial to directly extended existing systems to work on n-tuple open KBs

Our Contribution

Subject	Relation Phrase	Arguments
James K. Polk	was	a governor; before he was president
the currency of Spain	was	the Peseta; before 2002
Peseta	was replaced	by Euro; as official tender of Spain; in 2002
Barack Obama	graduated	from Harvard Law School; in 1979 and 1991
Obama	graduated	magna cum laude; from Harvard Law School; in 1991
Barack Obama	attended	Harvard Law School

nOKB: n-tuple open KB



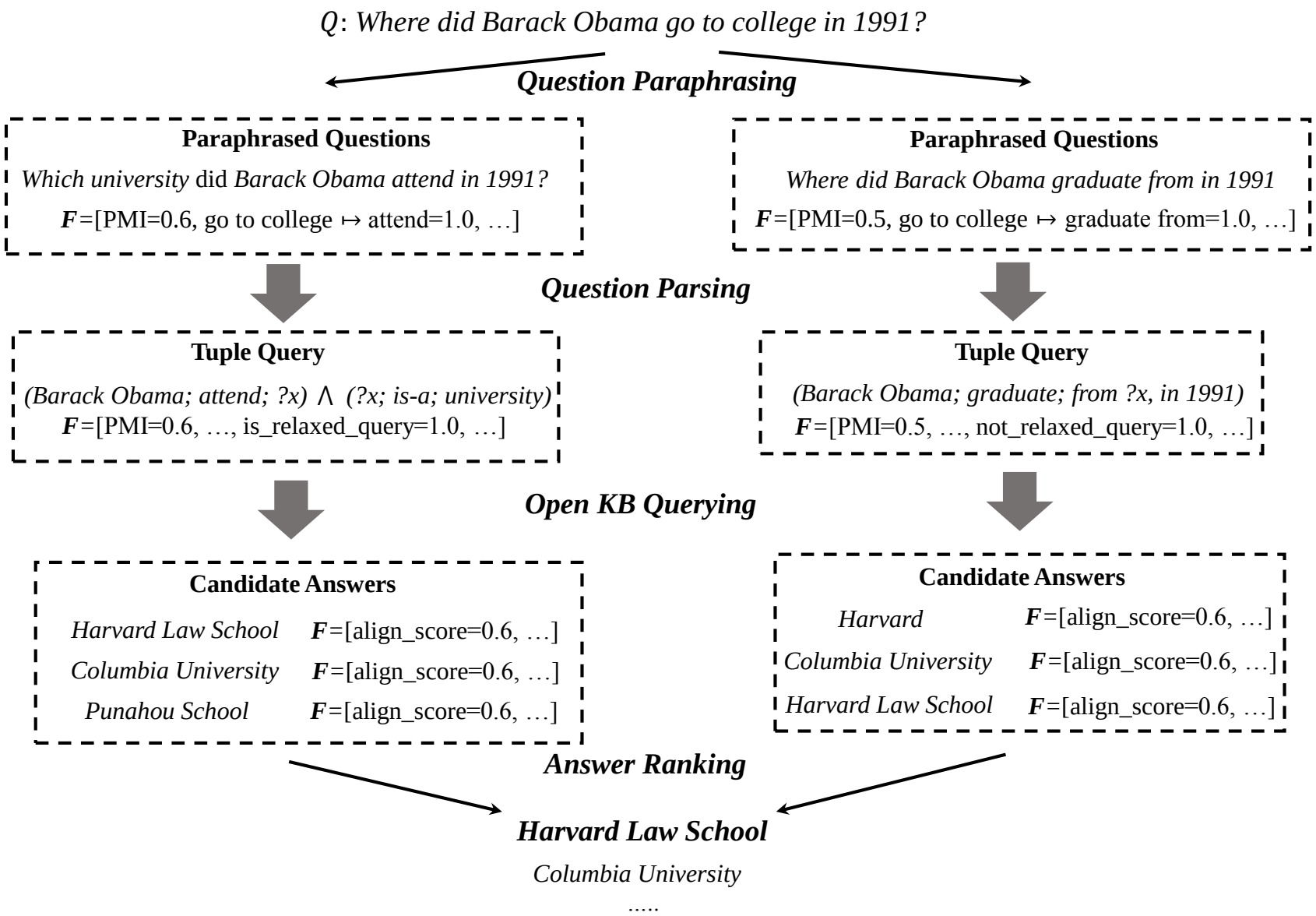
TAQA: n-**T**uple **A**ssertion-based **Q**uestion **A**nswering

leverage rich semantic information in n-tuple assertions to answer *complex* questions

Agenda

- Background & Motivation
- **Approach**
  - **TAQA**: n-**T**uple **A**ssertion-based **Q**uestion **A**nswering
  - Question Paraphrasing
  - Question Parsing
  - Open KB Querying
  - Answering Ranking
- Results
- Summary

TAQA's workflow – An example



Question Paraphrasing

- paraphrase (rewrite) input question into multiple semantically similar questions
  - Motivation: bridge the lexical/syntactical gap between input questions and relevant KB assertions

Question *Where did [Barack Obama] go to college in 1991 ?*

Noun Phrase[1] *Prepositional Phrase*

Relevant assertion in open KB

Subject	***Rel. Phrase***	***Arguments***
Barack Obama	*graduated*	*from Harvard Law School; in 1979 and 1991*

5 million Paraphrasing Templates

Where did **[None Phrase]** *go to college* ↦
Where did **[None Phrase]** *graduate from*

(Templates provided by Fader et al., 2014)

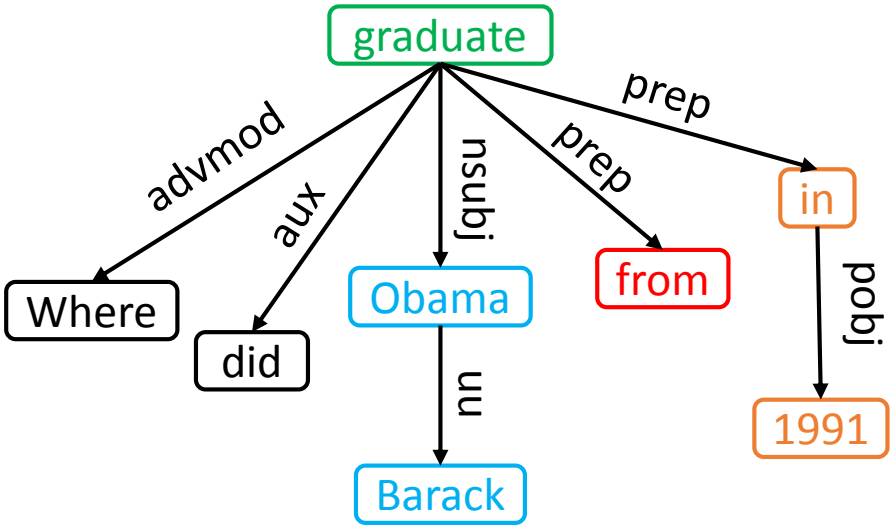
Paraphrased Question *Where did Barack Obama graduate from in 1991?*

[1]*Detected via phrase chunking*

Question Parsing

- Parse natural language questions into ***tuple queries***
- Use *dependency parsing* to parse questions into ***tuple queries***
  - Dep. parsing: powerful tool to analyze semantic relations between constituents
- Parse a question into tuple query by traveling through its *dependency tree*
- Generate fields in the tuple query from the descendants of the root node

Where did Barack Obama graduate from in 1991?



Dependency Parse Tree

tuple query

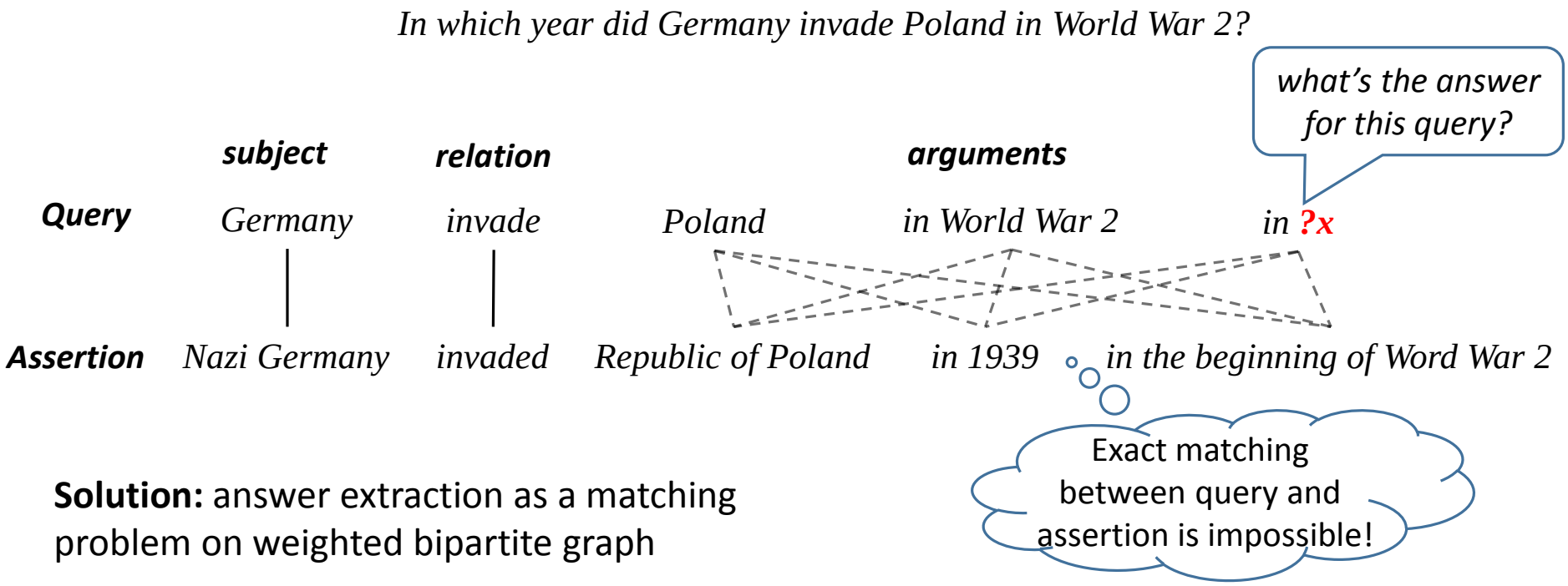
Subject	***Rel. Phrase***	***Arguments***
Barack Obama	*graduate*	*from ?x, in 1991*

Query Focus

nsubj: nominal subject
prep: preposition
pobj: prepositional object
advmod: adverbial modifiers
aux: auxiliary verb

Open KB Querying

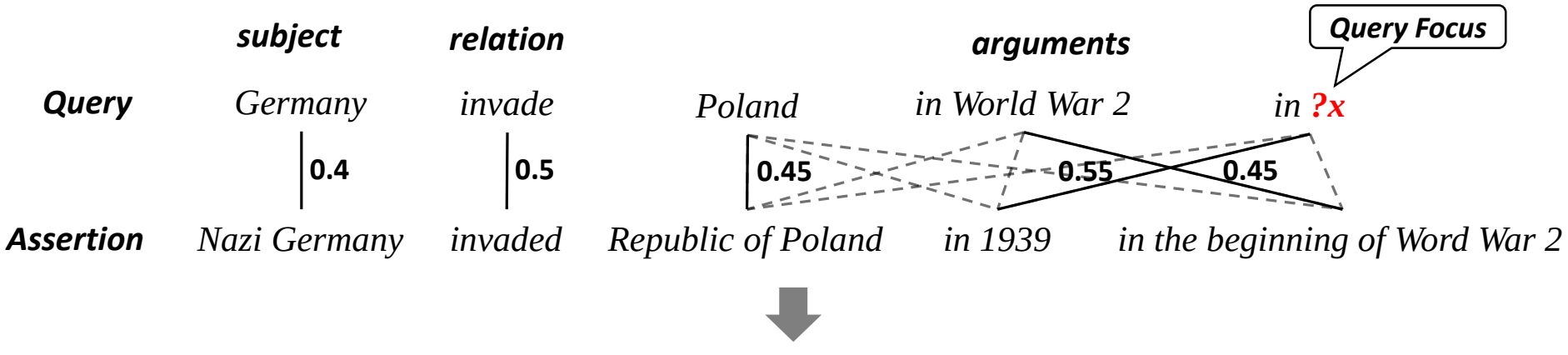
- Given a tuple query, how to query the open KB to retrieve answers?
- Key challenge: both the query and assertions in the open KB may have arbitrary number/order of arguments



Solution: answer extraction as a matching problem on weighted bipartite graph

Open KB Querying

In which year did Germany invade Poland in World War 2?



Answer is extracted from the field aligned with *?x* *in optimal solutions*

?x = 1939 *matching score = 2.35*

- Fields in the query and assertion form two sets of nodes in the bipartite graph
- Define pairwise similarity between fields in query and assertion (1-to-1 matching)
- Get optimal matching solutions by maximizing global matching score

Answer Ranking

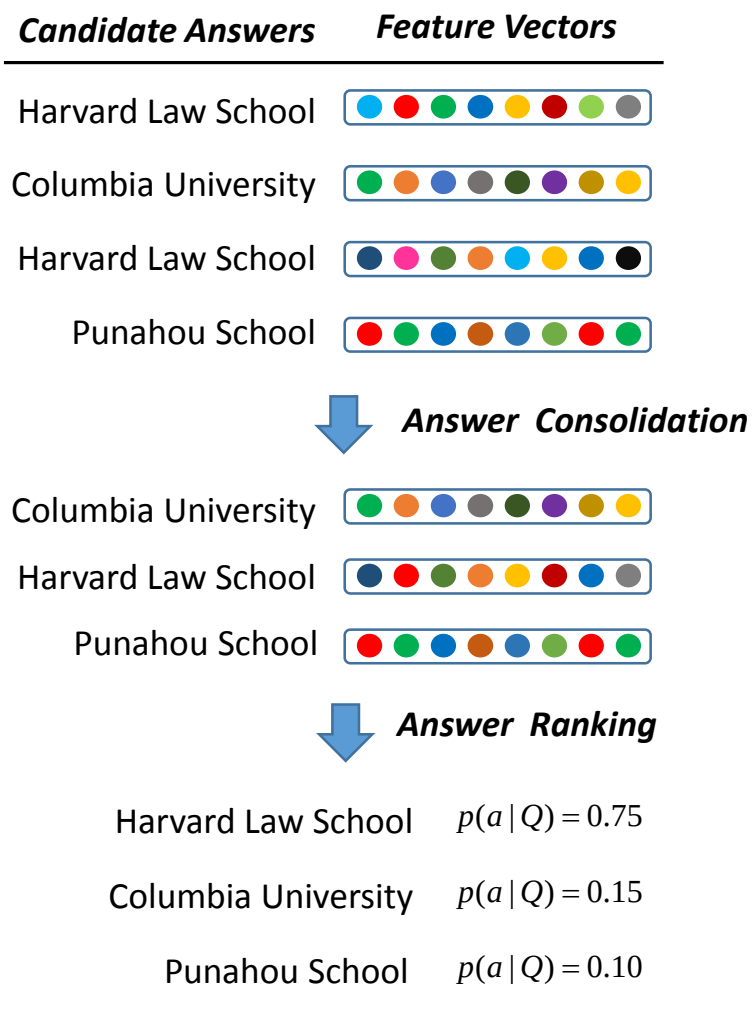
- Answer consolidation:
  - The same answers can be derived from paraphrasing-parsing-querying pipeline
  - Merge the feature vectors of the answers with the same surface text form
- Log-linear ranking model
  - Use over **20K** features to measure the probability that a candidate answer a is an answer of Q

$$p(a|Q) = \frac{exp\{\sum_{k=1}^{M} \lambda_k \cdot f_k(a)\}}{\sum_{a' \in \mathcal{A}} exp\{\sum_{k=1}^{M} \lambda_k \cdot f_k(a')\}}$$

  - Training the model by maximizing the log-likelihood on a set of question-answer pairs

$$\mathcal{L}(\mathcal{D}; \boldsymbol{\lambda}) = \sum_{t=1}^{N} \log p(a_t|Q_t; \boldsymbol{\lambda})$$

  - Stochastic Gradient Descent learning (AdaGrad)



Agenda

- Background & Motivation
- Approach
- **Results**
  - n-tuple Open KB
  - Questions Sets
  - Experiment results
- Summary

n-tuple open KB

Web Collection

5.7 million Web Pages

+

Web crawls using evaluation questions

Sentence Extraction

163 million sentences

Open Information Extraction

120 million n-tuple assertions

n-tuple Open KB

- **All** the existing open KBs are in **triplet** form (assertions contain only ONE argument)
- We build nOKB+: ensemble of all extracted n-tuple assertions and existing triplet open KBs (Probase, NELL, *etc*)
- Size of nOKB+: ***0.8 billion*** assertions

Evaluation Question Sets

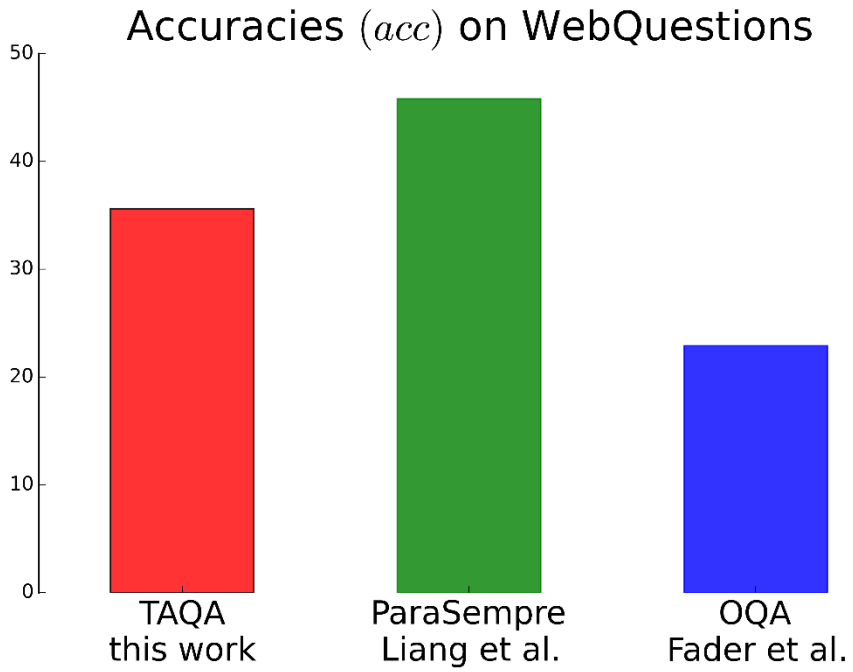
- **`WebQuestions`** [Berant *et al.*, 2013]:
  - 5810 questions, mostly (95%+) with simple constraints
  - Manually answered on Freebase by crowdsourcing workers

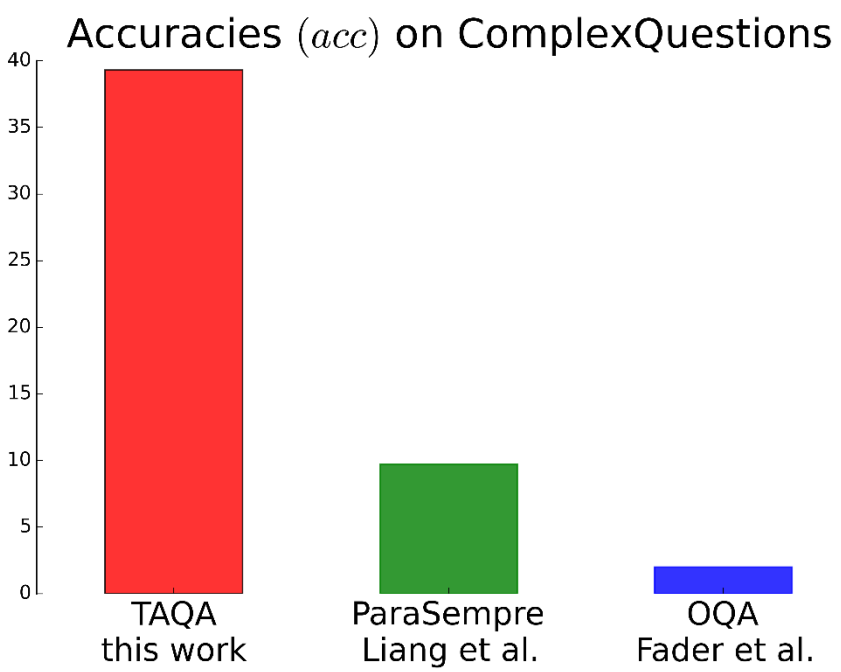
Question	Answer
Who played Jacob Black in twilight?	*Taylor Lautner*
Who did roger federer married?	*Dominica*

- **`ComplexQuestions`** [our work]: 300 questions with complex semantic constraints
  - 80 from **`WebQuestions`**, 220 newly collected
  - Manually labeled with gold-standard answers

Question	Answer
*What team did Shaq play for **before the Lakers**?*	*Orlando Magic*
*What country gained its independence **from Britain in 1960**?*	*Cyprus*
*What did France lose **to the British in the treaty of Paris in 1763**?*	*Dominica*

End-to-End Evaluation

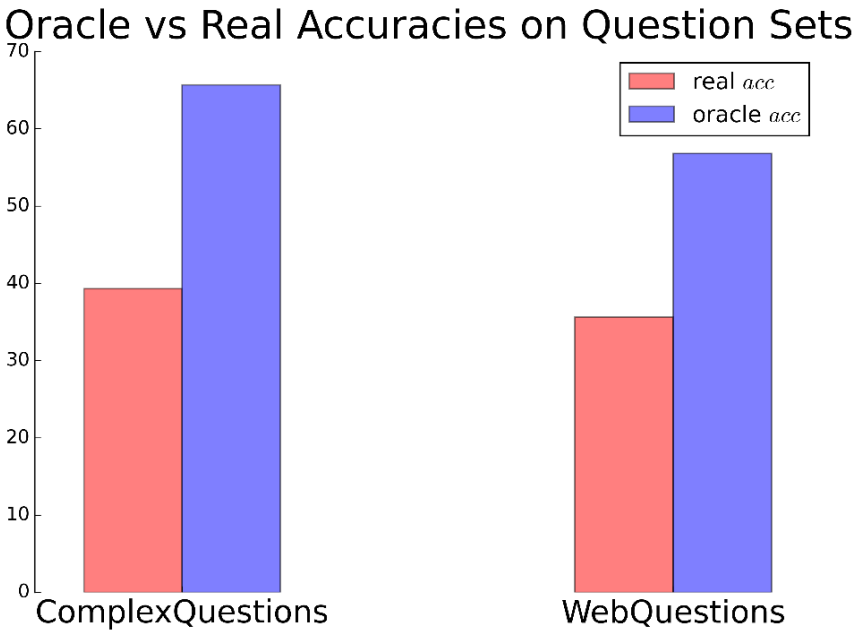




$$acc = \frac{\text{number of correctly answered questions}}{\text{total number of questions}}$$

- Baseline systems:
  - ParaSempre [Berant *et al.*, 2014]: SOTA curated KB-QA system
  - OQA [Fader *et al.*, 2014]: SOTA open KB-QA system
- TAQA performs well on both simple and complex questions

How can we further improve ?



- Oracle Accuracy approximately 1.5-2.0x higher than (actual) accuracy
- Ideally TAQA is capable of answering at most 60% questions in two question sets

Rank:	1	2-5	6-10	11-20	>20
ComplexQuestions	59.9%	19.8%	8.1%	4.1%	8.1%
WebQuestions	62.7%	15.8%	6.4%	4.9%	10.2%

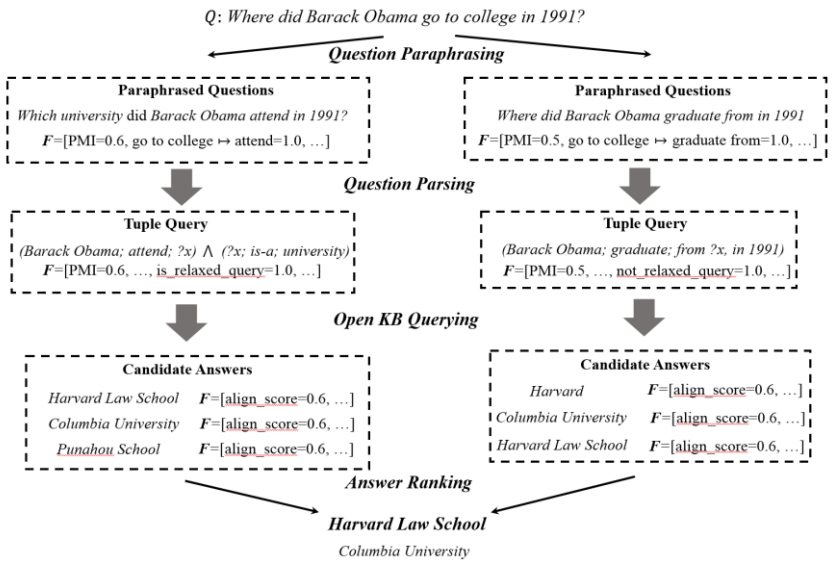
distribution of rank positions for oracle answers

- Insight: we can effectively improve TAQA's accuracy by **1/3** if we can improve the answer ranker to rank correct answers within top-5 up to top-1

Summary

Subject	Relation Phrase	Arguments
James K. Polk	was	a governor; before he was president
the currency of Spain	was	the Peseta; before 2002
Peseta	was replaced	by Euro; as official tender of Spain; in 2002
Barack Obama	graduated	from Harvard Law School; in 1979 and 1991
Obama	graduated	magna cum laude; from Harvard Law School; in 1991
Barack Obama	attended	Harvard Law School

nOKB: n-tuple open KB



TAQA: n-**T**uple **A**ssertion-based **Q**uestion **A**nswering

Thanks!

TAQA: n-**T**uple **A**ssertion-based **Q**uestion **A**nswering

Demo, Open KB and question set available at:
http://taqa.pcyin.me