When a Knowledge Base is not Enough

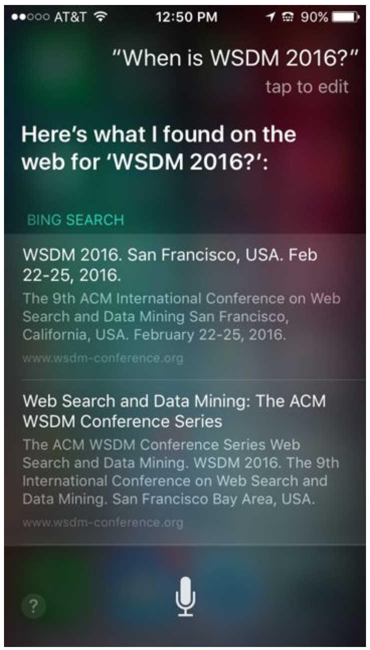
Question Answering over Knowledge Bases with External Text Data

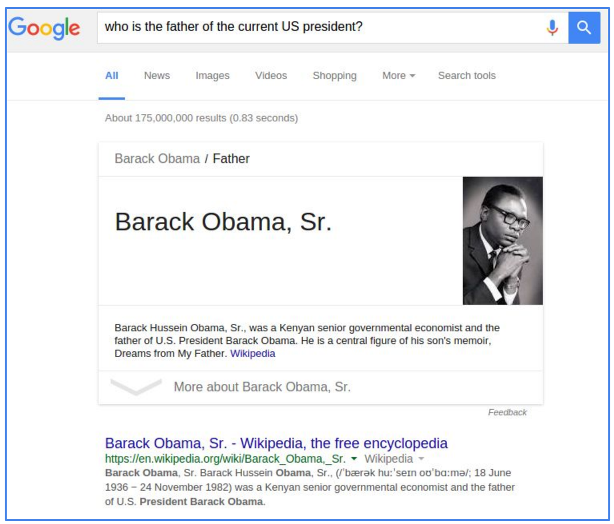
Denis Savenkov
Emory University
dsavenk@emory.edu

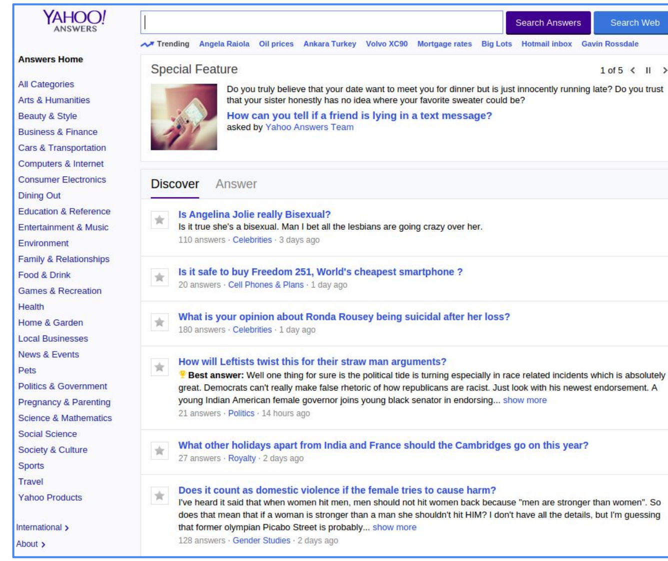
Eugene Agichtein
Emory University
eugene@mathcs.emory.edu

SIGIR 2016

Percentage of question search queries is growing[1]

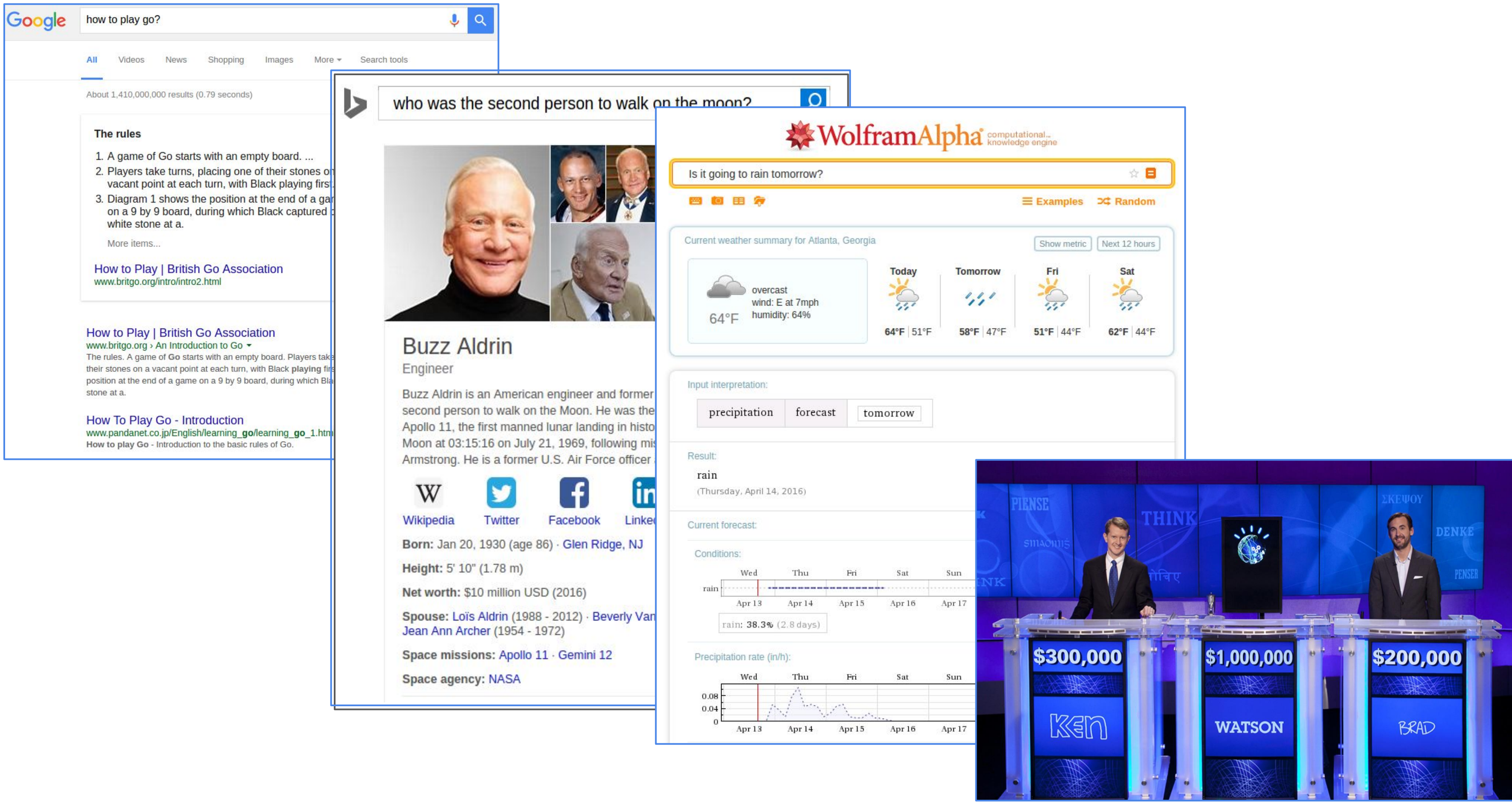






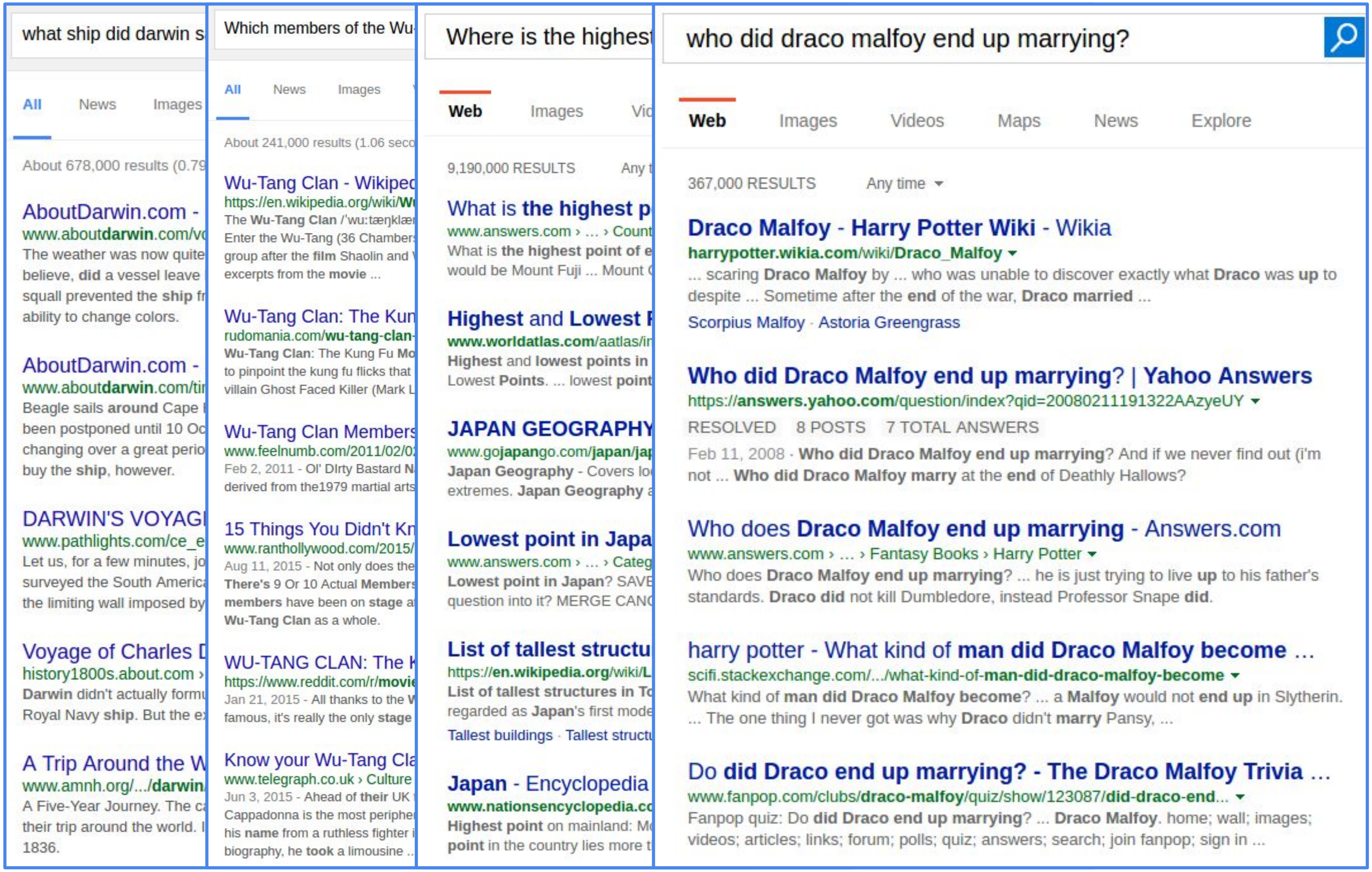
[1] "Questions vs. Queries in Informational Search Tasks", Ryen W. White et al, WWW 2015

Automatic Question Answering works relatively well for simple factoid questions



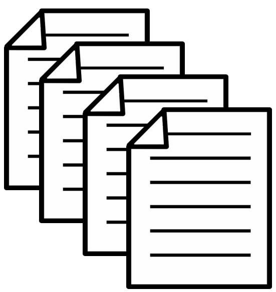
(AP Photo/Jeopardy Productions, Inc.)

For many questions we still have to dig into "10 blue links"



* the questions are taken from different QA datasets (WebQuestions, QALD-5, Yahoo! Answers Webscope)

Different data sources are used for question answering



San Francisco, California	
Consolidated city-county	
City and County of San Francisco	
Nickname(s): *The City by the Bay*; *Fog City*; *San Fran*;[1] *Frisco* (*locally disparaged*);[2][3][4][5] *The City that Knows How* (*past*);[6] *Baghdad by the Bay* (*past*);[7] *The Paris of the West* (*past*)[8]	
Motto: *Oro en Paz, Fierro en Guerra* (English: "Gold in Peace, Iron in War")	
Coordinates: 37°47′N 122°25′W	
Country	United States
State	California
CSA	San Jose–San Francisco–Oakland
Metro	San Francisco–Oakland–Hayward
Mission	June 29, 1776[9]
Incorporated	April 15, 1850[10]
Founded by	José Joaquin Moraga Francisco Palóu
Named for	St. Francis of Assisi
Government	
• **Type**	Mayor-council
• **Body**	Board of Supervisors
• **Mayor**	Edwin M. Lee (D)[11]
• **Supervisors**[13]	**List** [show]
• **Assembly members**[14][15]	David Chiu (D) Phil Ting (D)
• **State senator**	Mark Leno (D)[12]
• **United States Representatives**[16][17]	Nancy Pelosi (D) Jackie Speier (D)

2014 rank	City	State[5]	2014 estimate	2010 Census	Change	2014 land area	2010 population density	Location
1	*New York*[6]	New York	8,491,079	8,175,133	+3.86%	302.6 sq mi 783.8 km^2	27,012 per sq mi 10,430 km^{-2}	40.6643°N 73.9385°W
2	*Los Angeles*	California	3,928,864	3,792,621	+3.59%	468.7 sq mi 1,213.9 km^2	8,092 per sq mi 3,124 km^{-2}	34.0194°N 118.4108°W
3	*Chicago*	Illinois	2,722,389	2,695,598	+0.99%	227.6 sq mi 589.6 km^2	11,842 per sq mi 4,572 km^{-2}	41.8376°N 87.6818°W
4	*Houston*[7]	Texas	2,239,558	2,100,263	+6.63%	599.6 sq mi 1,552.9 km^2	3,501 per sq mi 1,352 km^{-2}	29.7805°N 95.3863°W
5	*Philadelphia*[8]	Pennsylvania	1,560,297	1,526,006	+2.25%	134.1 sq mi 347.3 km^2	11,379 per sq mi 4,394 km^{-2}	40.0094°N 75.1333°W
6	***Phoenix***	Arizona	1,537,058	1,445,632	+6.32%	516.7 sq mi 1,338.3 km^2	2,798 per sq mi 1,080 km^{-2}	33.5722°N 112.0880°W

Text documents

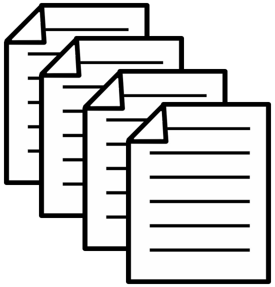
Web tables & infoboxes

Knowledge bases

Unstructured data **Semi-structured data** **Structured data**

Data Sources have different advantages and problems

Text documents



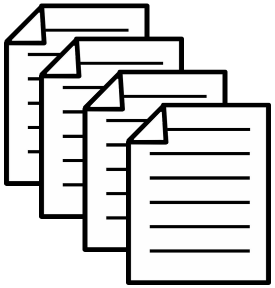
- easy to match against question text
- cover a variety of different information types
- each text phrase encodes a limited amount of information about mentioned entities

Knowledge bases

- aggregate the information around entities
- allow complex queries over this data using special languages (e.g. SPARQL)
- hard to translate natural language questions into special query languages
- incomplete (missing entities, facts and properties)

Advantages of one Data Source can compensate disadvantages of the other

Text documents



Knowledge bases

Text documents		Knowledge bases
+ easy to match against question text	→	- hard to translate natural language questions into special query languages
+ cover a variety of different information types	→	- incomplete (missing entities, facts and properties)
- each text phrase encodes a limited amount of information about mentioned entities	←	- aggregate the information around entities

Knowledge Base Question Answering (KBQA)

- Goal: translate natural language question into structured KB query (e.g. SPARQL) to retrieve correct entity or attribute value

When did Tom Hanks win his first Oscar?

```
PREFIX fb: <http://rdf.freebase.com/ns/>
SELECT ?year WHERE {
  fb:/m/0bxtg  fb:/award/award_winner/awards_won ?award .
  ?award fb:/award/award_honor/award fb:/m/0f4x7 .
  ?nomination fb:/award/award_honor/year ?year .
} ORDER BY ?year LIMIT 1
```

Knowledge Base Question Answering Challenges

1. Query analysis
    - How to identify question topic entity to anchor KB search?
2. Candidate generation
    - What predicates might correspond to words and phrases in the question?
    - What entities to include as candidate answers?
3. Evidence extraction
    - How to score correspondence between a certain candidate answer (e.g. involved predicates) and the question?
4. Answer selection
    - How to rank candidate answers to select the final response?

Existing Text-KB hybrid approaches

✓ **Open QA** [A.Fader et al. 2014]
- → Use Open Information Extraction to build semi-structured KB from text
- → Joint QA over extracted and curated KB

✓ **Extended Knowledge Graphs** [S. Elbassuoni et al 2009, M.Yahya et al 2016]
- → Extend triples in knowledge base with keywords
- → SPARQL query relaxation techniques to use keyword matches

✓ **"Open Domain Question Answering via Semantic Enrichment"** [H.Sun et al 2015]
- → Annotate text with entity mentions
- → Use entity types and textual KB descriptions to imrove text-based QA

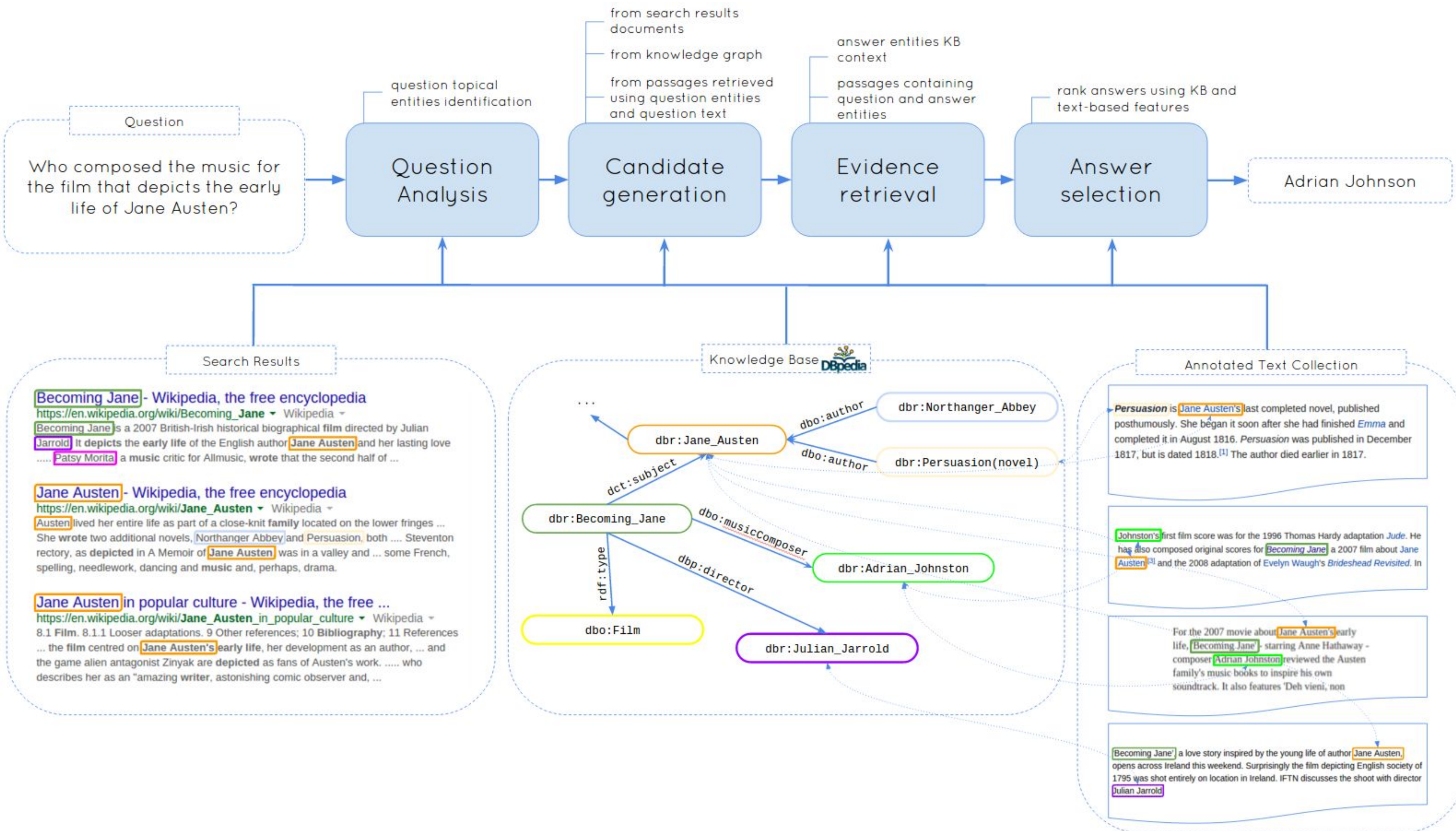
✓ **"Question Answering on Freebase via Relation Extraction and Textual Evidence"** [K. Xu et al. 2016]
- → Using text documents to refine answers, generated by KBQA system

✓ **Memory Networks** [A. Bordes et al 2015]
- → encode curated and OpenIE triples into NN memory

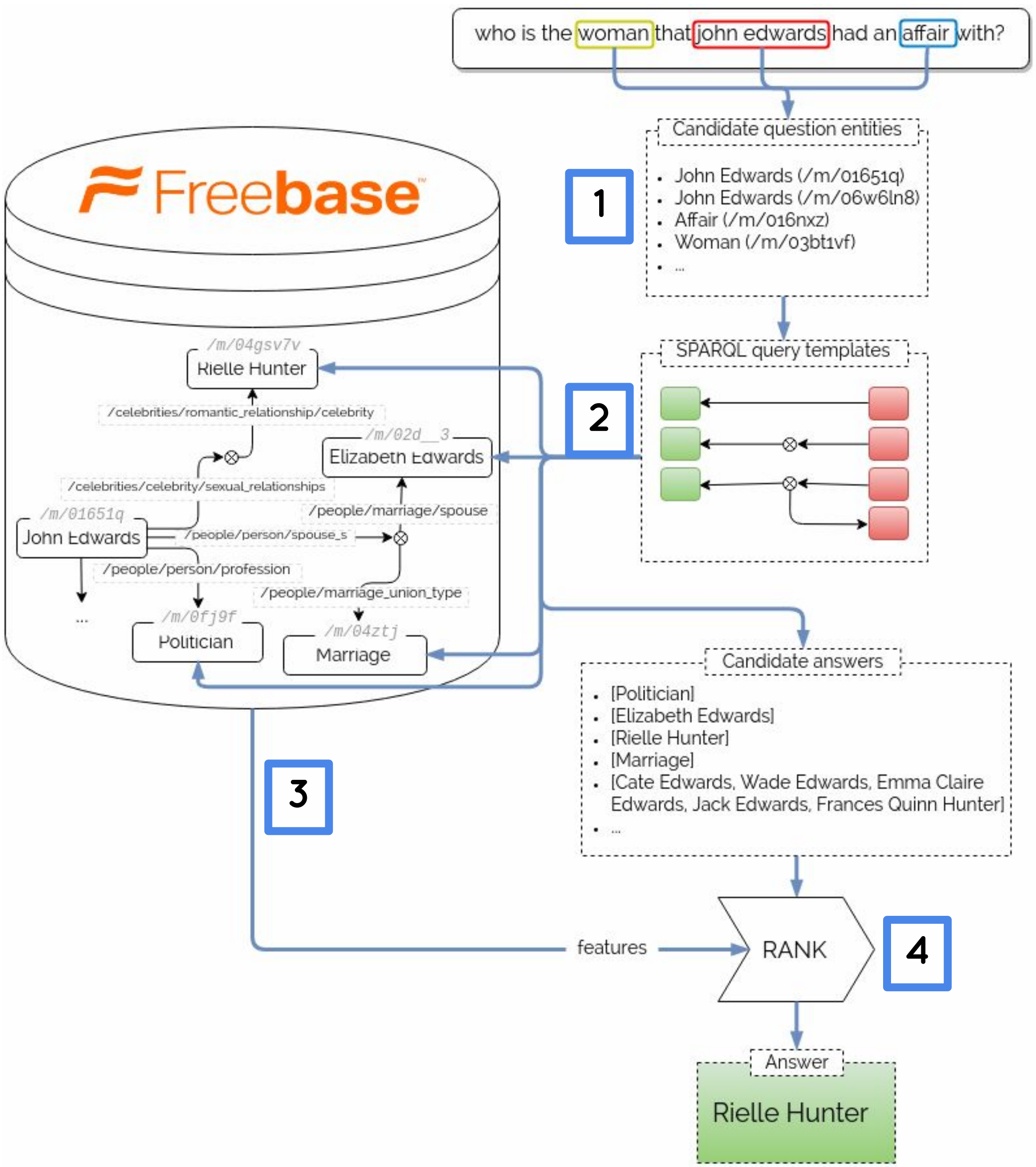
Text2KB: main idea

- ✓ Improve different stages in Knowledge Base Question Answering using various textual data
  - query analysis
    - ✓ question topic entity identification using web search results
  - candidate generation
    - ✓ Mine associations patterns between question terms and predicates from CQA data
  - evidence extraction
    - ✓ build language model for candidate question-answer entity pairs based on annotated corpus of text documents
  - answer selection
    - ✓ Score answer candidates using a combination of KB and text-based features

Text2KB: Incorporating Text in Answering Process



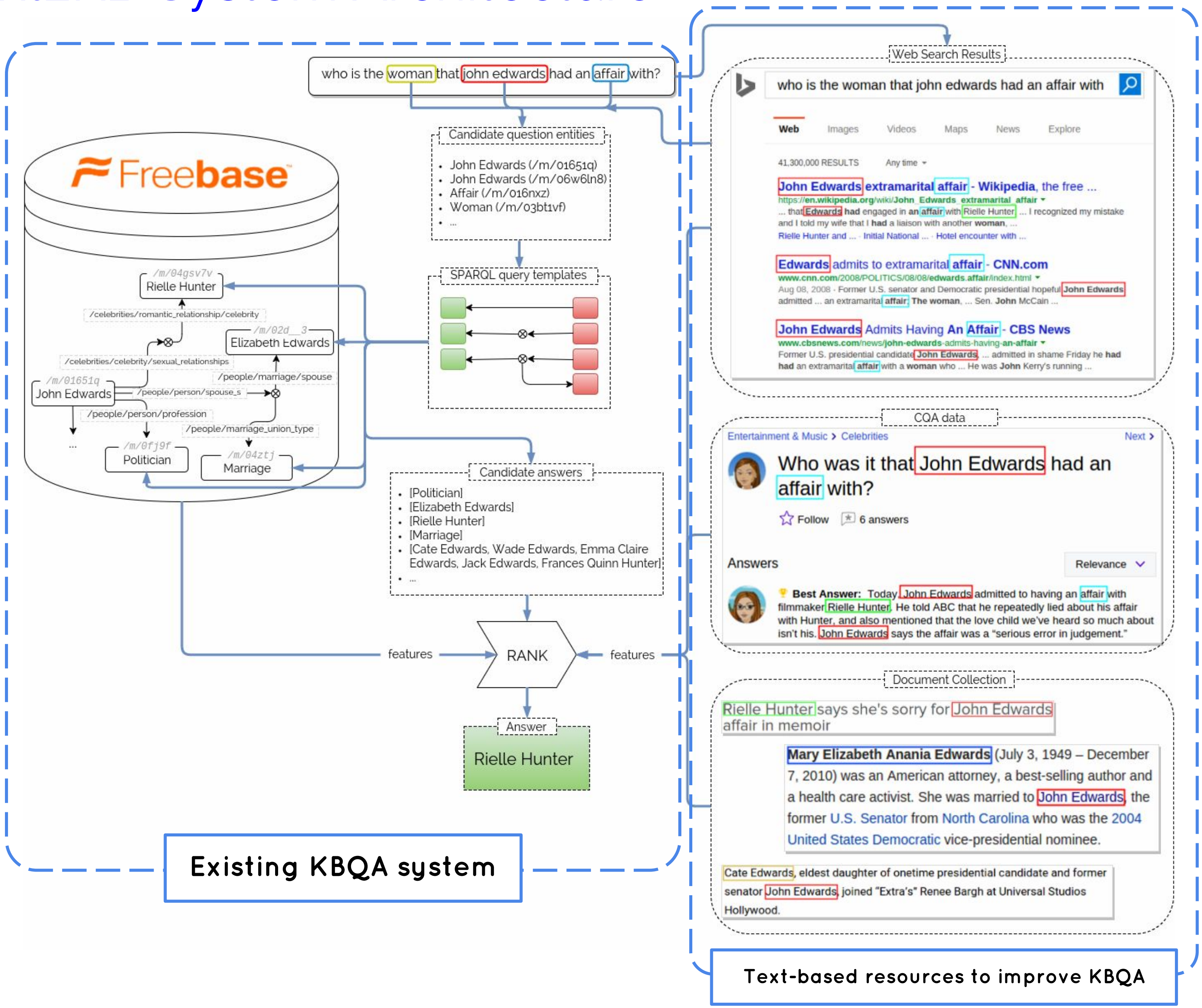
Baseline system architecture*



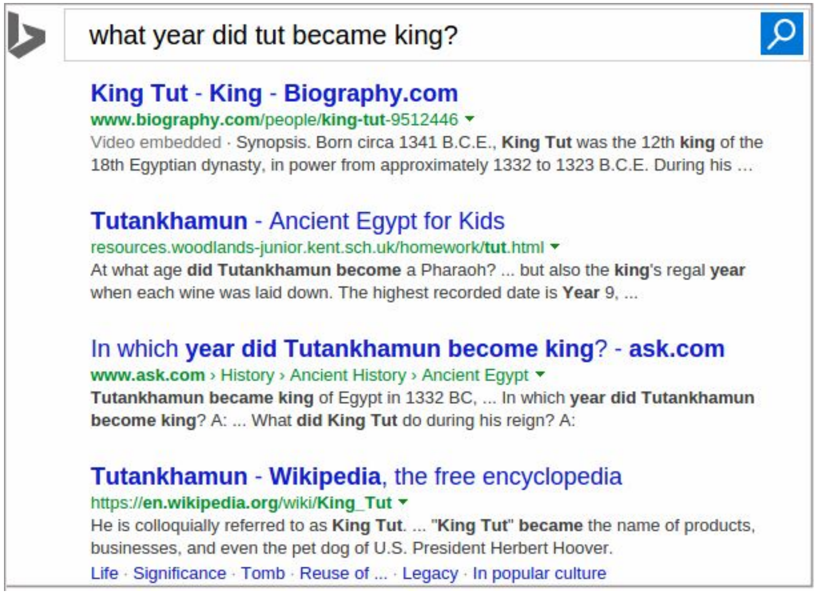
1. **Detecting question topic entity**: multiple candidates are detected using dictionary of names and aliases
2. **Answer candidate generation**: instantiate candidate SPARQL queries from the neighborhood of question entities using a set of template queries
3. **Evidence generation**: each candidate is represented with a set of features, describing the detected topic entity, predicates on KB path connecting topic and answer entities, etc.
4. **Answer selection**: candidate answers are ranked using a trained ranking model and top scoring one is returned as the answer

* “More Accurate Question Answering on Freebase” by Hannah Bast et al, 2015

Text2KB System Architecture

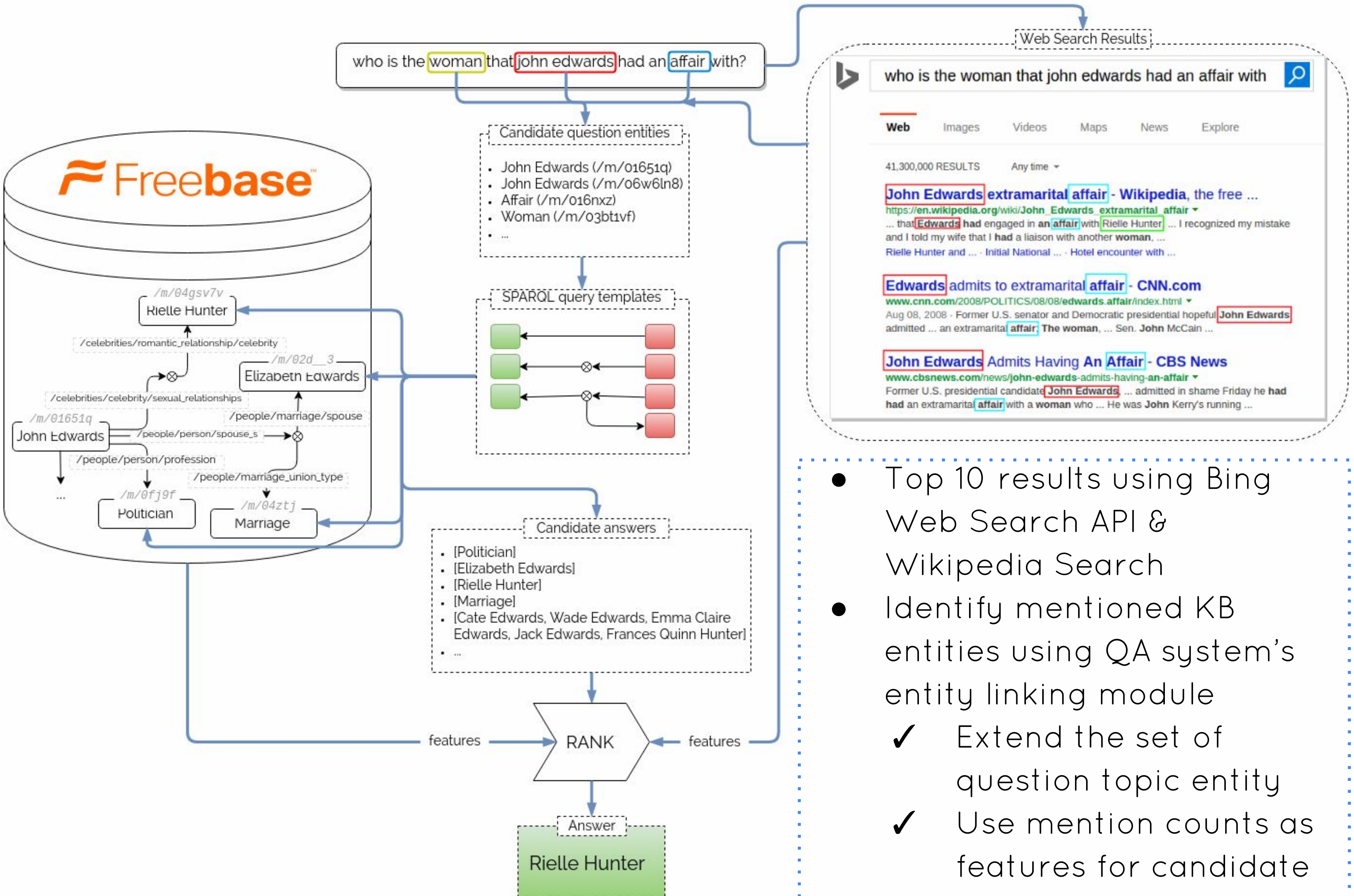


Question Analysis: Entity Linking



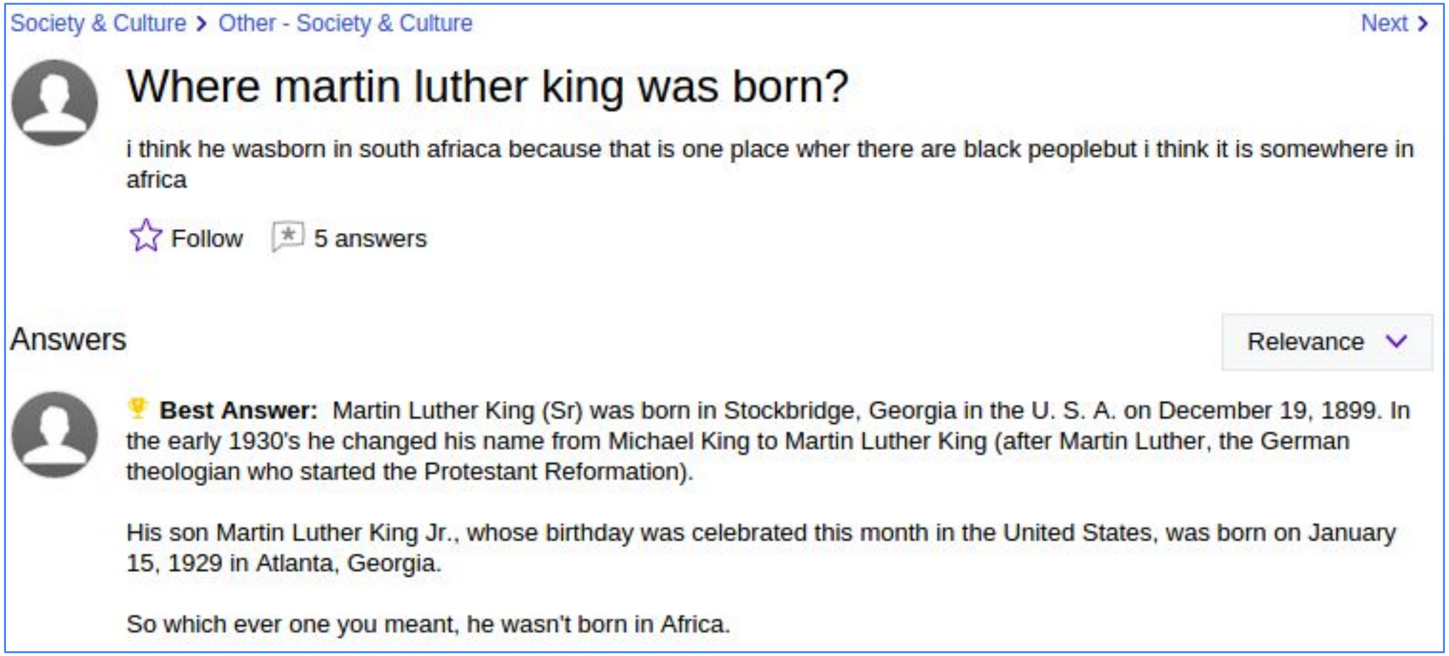
- ✓ Web Search Results can help entity linking and provide textual evidence to answer candidates
- ✓ Contains multiple mentions of the question topic entity, often in variations, which might help entity linking
- ✓ Search results often contain the answer to the question itself, which is exploited by text-based question answering systems

Text2KB System Architecture: web search results



- Top 10 results using Bing Web Search API & Wikipedia Search
- Identify mentioned KB entities using QA system's entity linking module
  - ✓ Extend the set of question topic entity
  - ✓ Use mention counts as features for candidate ranking

Community Question Answering data can help map question phrases to predicates



Society & Culture > Other - Society & Culture Next >

Where martin luther king was born?

i think he wasborn in south afriaca because that is one place wher there are black peoplebut i think it is somewhere in africa

Follow 5 answers

Answers Relevance

Best Answer: Martin Luther King (Sr) was born in Stockbridge, Georgia in the U. S. A. on December 19, 1899. In the early 1930's he changed his name from Michael King to Martin Luther King (after Martin Luther, the German theologian who started the Protestant Reformation).

His son Martin Luther King Jr., whose birthday was celebrated this month in the United States, was born on January 15, 1929 in Atlanta, Georgia.

So which ever one you meant, he wasn't born in Africa.

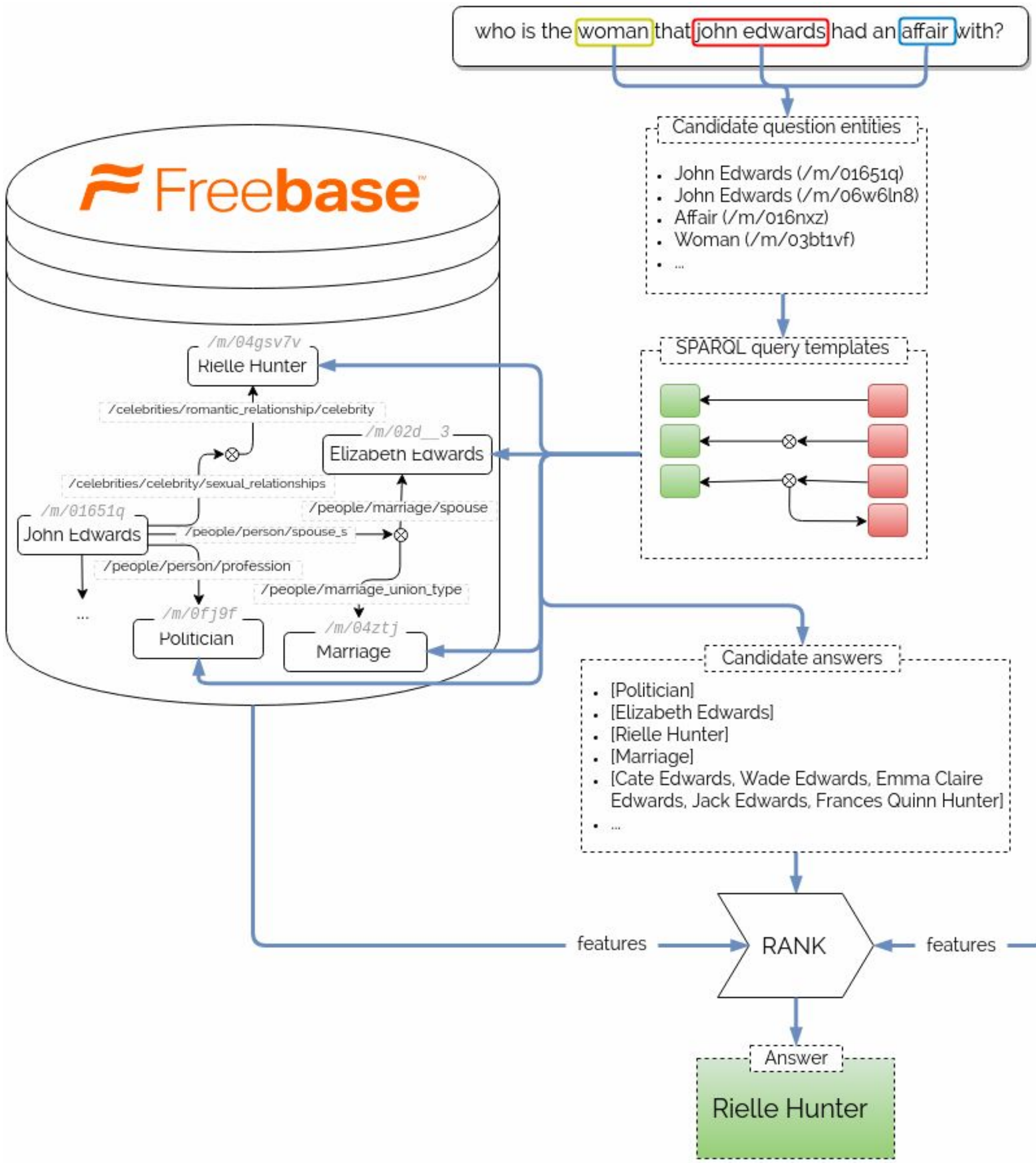
- ✓ Huge number of question-answer pairs, but noisy (most of the questions aren't factoid, answers are verbose and contain redundant information)
- ✓ Can be helpful to learn associations between the language of a question and KB predicates using distant supervision assumption

Examples of term-predicate associations computed using CQA data

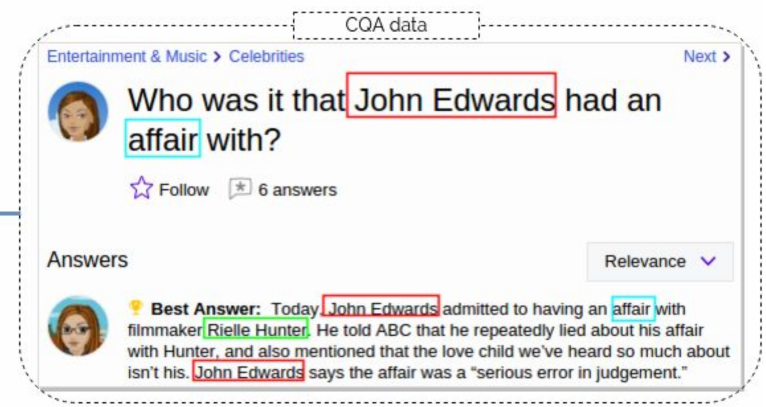
Term	Predicate	PMI score
born	people.person.date_of_birth people.person.date_of_death location.location.people_born_here	3.67 2.73 1.60
kill	people.deceased_person.cause_of_death book.book.characters	1.70 1.55
currency	location.country.currency_formerly_used location.country.currency_used	5.55 3.54
school	education.school.school_district people.education.institution sports.school_sports_team.school	4.14 1.70 1.69
win	sports.sports_team.championships sports.sports_league.championship	4.11 3.79

✓ Despite the noisy distant supervision labeling, top scoring predicates are indeed related to the corresponding word

Text2KB System Architecture: CQA data



- Distant supervision to label question-answer pairs from Yahoo! Answers WebScope collection with KB predicates
- Learn associations between question terms and predicates using PMI scores
  - Use these PMI scores as features to score candidate answer predicates



Text around mentions of pairs of entities in documents help explain relationships between the entities

Rielle Hunter says she's sorry for John Edwards affair in memoir

Mary Elizabeth Anania Edwards (July 3, 1949 – December 7, 2010) was an American attorney, a best-selling author and a health care activist. She was married to John Edwards, the former U.S. Senator from North Carolina who was the 2004 United States Democratic vice-presidential nominee.

Cate Edwards, eldest daughter of onetime presidential candidate and former senator John Edwards, joined "Extra's" Renee Bargh at Universal Studios Hollywood.

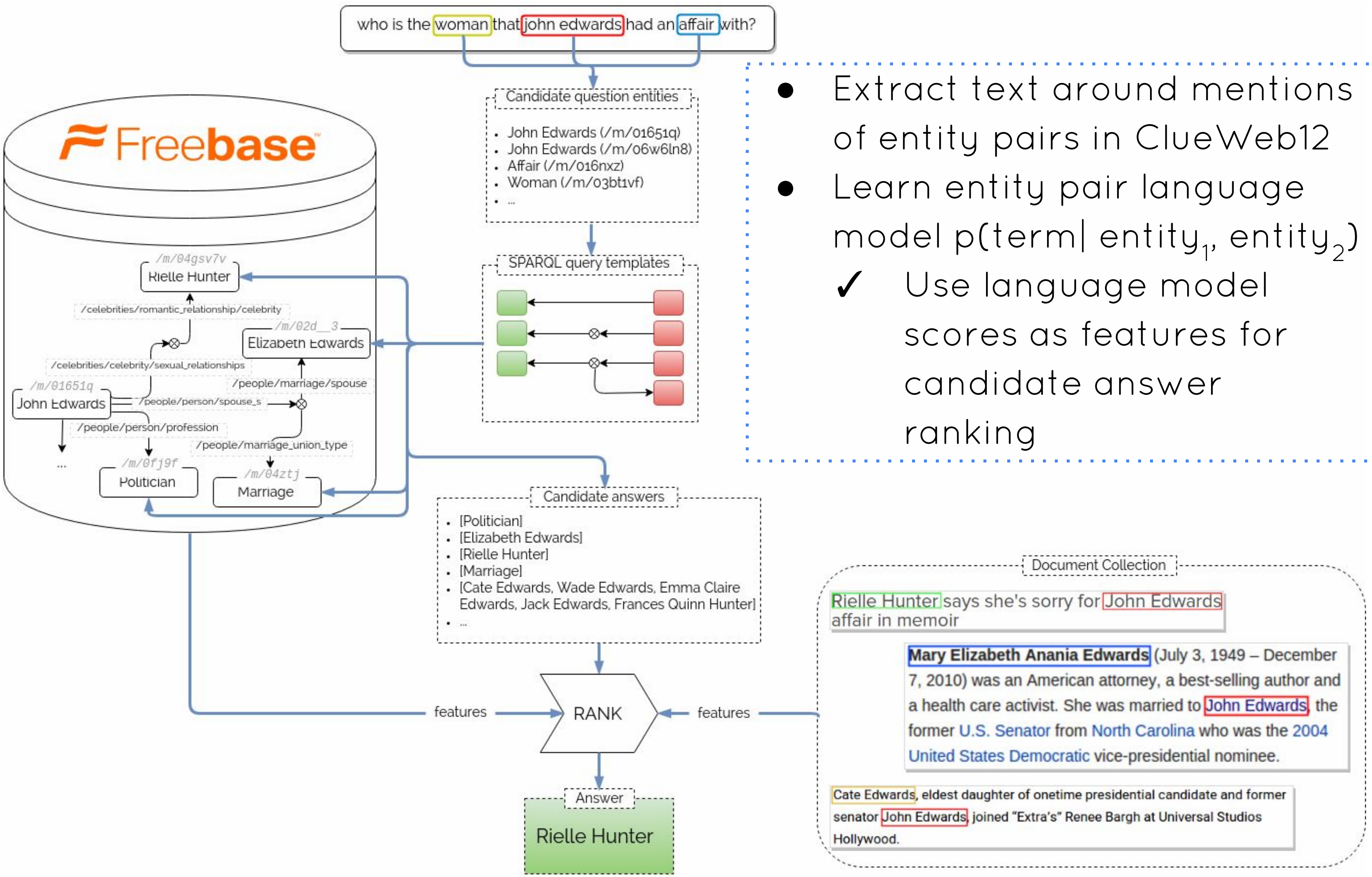
- ✓ Sentences and passages that mention multiple entities often express some facts about them
- ✓ Terms used in these passages can explain the relationships between the entities

Examples of entity pair language models

Entity 1	Entity 2	Term counts
John Edwards	Rielle Hunter	campaign, affair, mistress, child, former ...
John Edwards	Cate Edwards	daughter, former, senator, courthouse, greensboro, eldest ...
John Edwards	Elizabeth Edwards	wife, hunter, campaign, affair, cancer, rielle, husband ...
John Edwards	Frances Quinn	daughter, john, rielle, father, child, former, paternity...

✓ Terms most frequently used around mention of a pair of entities indeed shed some light on the relationship between the entities

Text2KB System Architecture: document collection



- Extract text around mentions of entity pairs in ClueWeb12
- Learn entity pair language model p(term| entity$_1$, entity$_2$)
  - ✓ Use language model scores as features for candidate answer ranking

Evaluation

- ✓ WebQuestions dataset
  - 3,778 training and 2,032 test questions
- ✓ Metrics:
  - Average F1: $avg\ F1 = \frac{1}{|Q|} \sum_{q \in Q} f1(a_q^*, a_q)$

$$f1(a_q^*, a_q) = 2 \frac{precision(a_q^*, a_q) recall(a_q^*, a_q)}{precision(a_q^*, a_q) + recall(a_q^*, a_q)}$$

- ✓ Methods compared:
  - Aqqu (Bast et al, 2015) - our KB-only baseline
  - STAGG (Yih et al, 2015) - SOTA at the moment of publication
  - our Text2KB (Web search)
  - our Text2KB (Wikipedia search)

Results

	Recall	Precision	F1
OpenQA [A.Fader et al 2014]	-	-	0.35
STAGG [H.Sun et al 2015]	0.607	0.528	0.525
Aqqu (baseline) [H.Bast et al 2015]	0.604	[illegible]	0.494
Text2KB (wikipedia search)	0.632	0.498	[illegible]
Text2KB (web search)	0.635	0.506	0.522

+5.7%

- ✓ Text2KB significantly improves upon the baseline Aqqu system (0.494 -> 0.522 avg F1 score)
- ✓ Text2KB reaches the performance of STAGG, best result at the moment of publication
    - but this work is orthogonal to improvements in STAGG and therefore can be combined

Component ablation

System	avg F1
Aqqu	0.494
+ Entity linking from search results	0.508
+ Search results, CQA and Clueweb features for ranking	0.514
Text2KB	0.522

System	avg F1
Aqqu	0.494
Text2KB (Web search)	0.522
- Web search data	0.513
- CQA data	0.519
- ClueWeb data	0.523
+ Web search data only	0.522
+ CQA data only	0.508
+ ClueWeb data only	0.514

- ✓ Both entity linking using web search results and features for answer ranking contribute to improvements
- ✓ Search results have the largest contribution to the overall performance, but CQA and ClueWeb are also useful

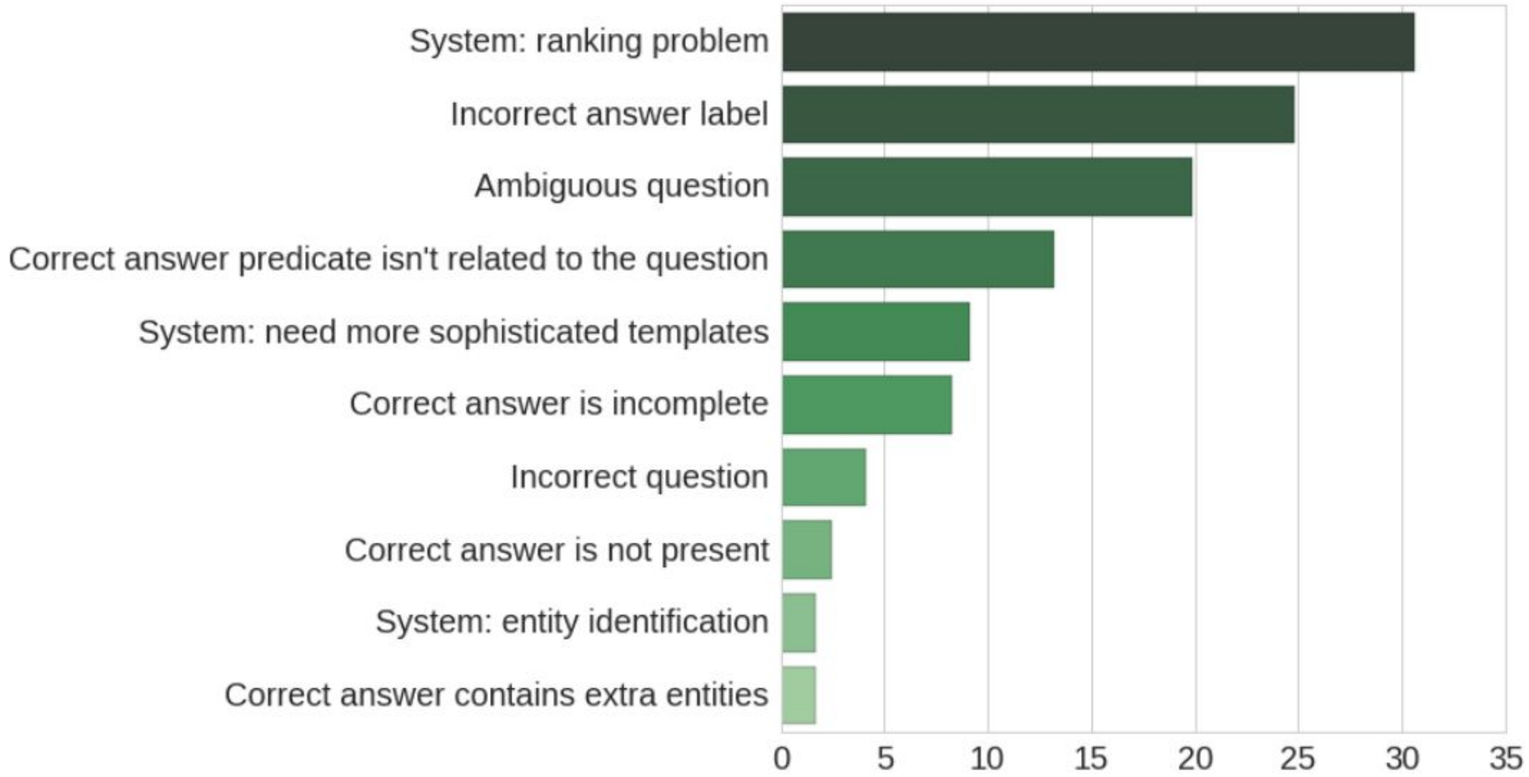
Combining Text2KB & STAGG

System	avg F1
STAGG (Yih et al, 2015)	0.525
Text2KB + STAGG (takes STAGG answers if it has less entities)	0.532
Text2KB + STAGG (Oracle: chooses answer with higher F1 score)	0.606

✓ Combining results of Text2KB and STAGG suggests that our ideas could benefit it as well

- Heuristic combination: take Text2KB or STAGG answer, which contains less entities
- Oracle combination always choose the answer with higher F1

Error analysis



- ✓ Majority of errors (F1 < 1) are ranking errors
- ✓ But there are also many problems in questions and labels
  - ✓ Check out the new WebQuestionsSP dataset: https://goo.gl/eQF0tM

Current & Future work

- Overall, our system is <u>most helpful</u>:
  - ➢ Question topic entity is hard to identify (uncommon alias, misspelling)
  - ➢ Form of the question or ground truth predicate is less frequent in the training set
- Our system has the <u>following problems</u>:
  - ➢ Less effective for tail and abstract entities, whose mentions are harder to find in text. For example entity “Associated Press Male Athlete of the Year” isn’t linked correctly (unless mentioned exactly by name)
  - ➢ Our use of text doesn’t help much to solve KB incompleteness (e.g. missing facts or predicates)
- <u>Future work</u>:
  - ➢ Instead of improving KBQA, move to more open scenario
    - new hybrid model that will use all the information available in different data sources
    - new dataset of entity-centric factoid questions

Conclusions

- Textual data sources provide additional information, that can compensate disadvantages of structured knowledge bases
- Our Text2KB system uses a combination of structured and unstructured data to improve Knowledge Base Question Answering
  - Improve avg F1 on WebQuestions dataset: **0.494 -> 0.522**

Acknowledgements

Denis Savenkov is planning to defend in December 2016 and will be on the market for postdoc and industry research positions

dsavenk@emory.edu

Thank you!

This work was partially supported by the Yahoo Labs Faculty Research Engagement Program (FREP).

What is Knowledge based Question Answering

Question:
Who is the president of the United States?

Answer:
Donald Trump

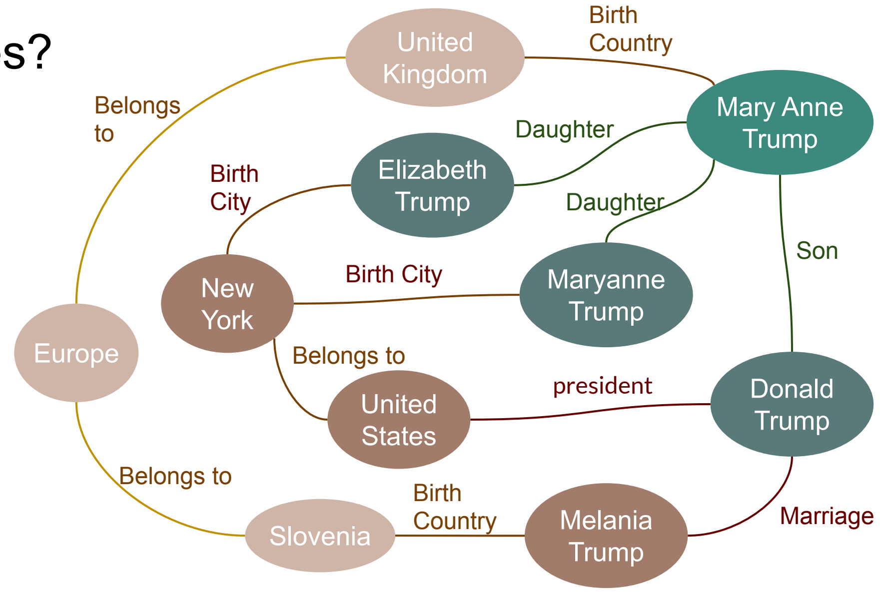
What is Knowledge based Question Answering

Question:

Who is the president of the United States?

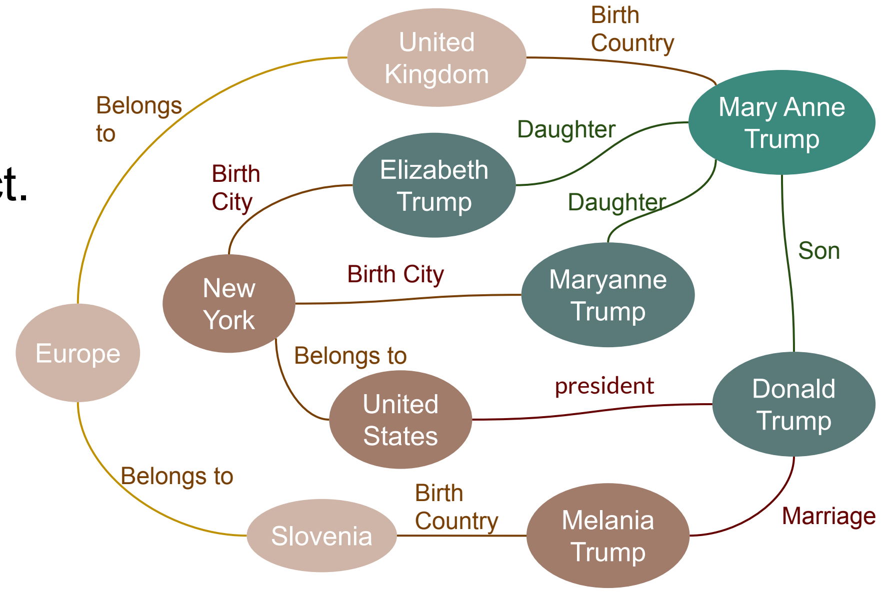
Answer:

Donald Trump



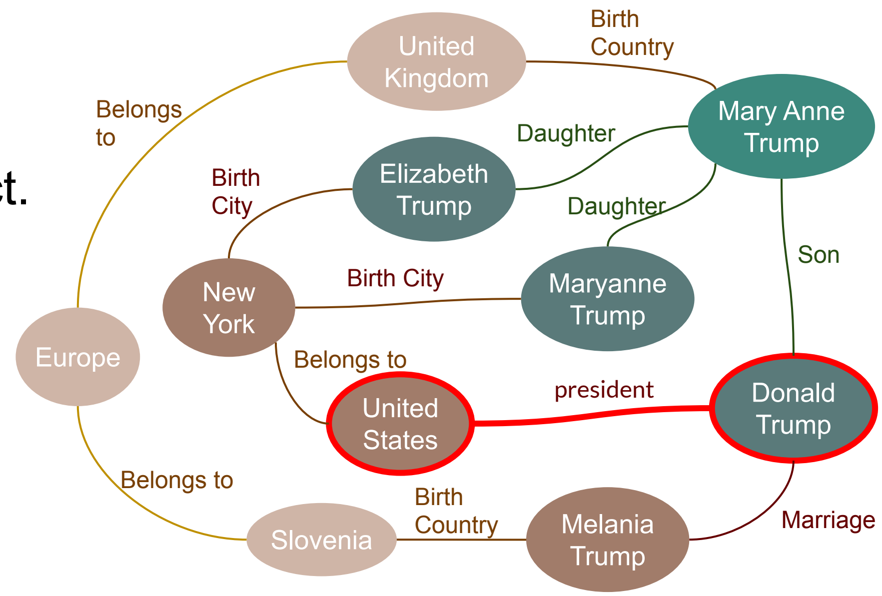
Knowledge Graph

- Each node e is an entity.
- Each edge r represents a relation between two connected entities.
- A triplet $(e_{head}; r; e_{tail})$ is called a fact.



Knowledge Graph

- Each node e is an entity.
- Each edge r represents a relation between two connected entities.
- A triplet $(e_{head}; r; e_{tail})$ is called a fact.



Fact:

(United States, President, Donald Trump)

What is Knowledge based Question Answering

Question:

Who is the sister of the president of the United States?

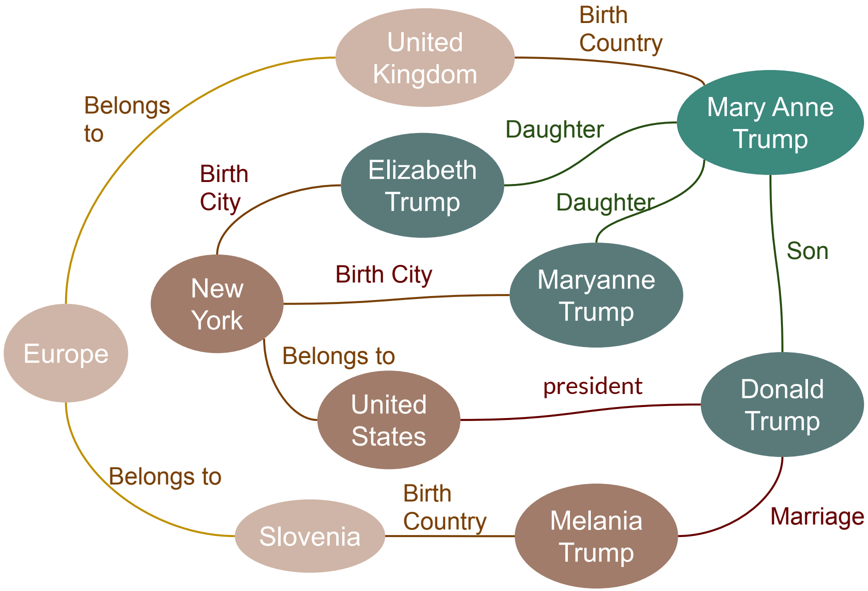
Who is the president of the United States?

Who is the mother of Donald Trump?

Who are the daughters of Mary MacLeod?

Answer:

Maryanne Trump / Elizabeth Trump



What is Knowledge based Question Answering

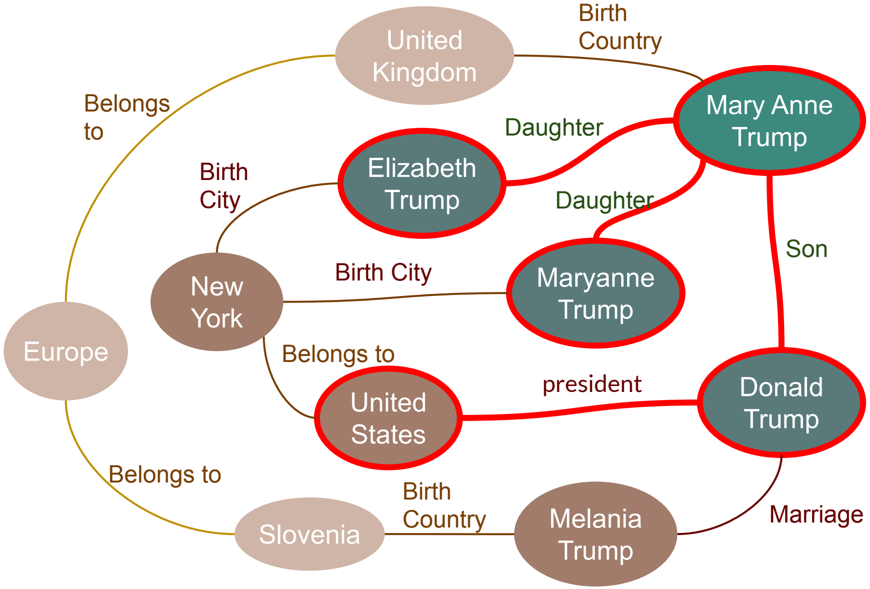
Question:
Who is the sister of the president of the United States?

(United States, President, Donald Trump)

(Donald Trump, Mother, Mary Anne Trump)

(Mary Anne Trump, Daughter, Maryanne/Elizabeth)

Answer:
Maryanne Trump / Elizabeth Trump



What is Knowledge based Question Answering

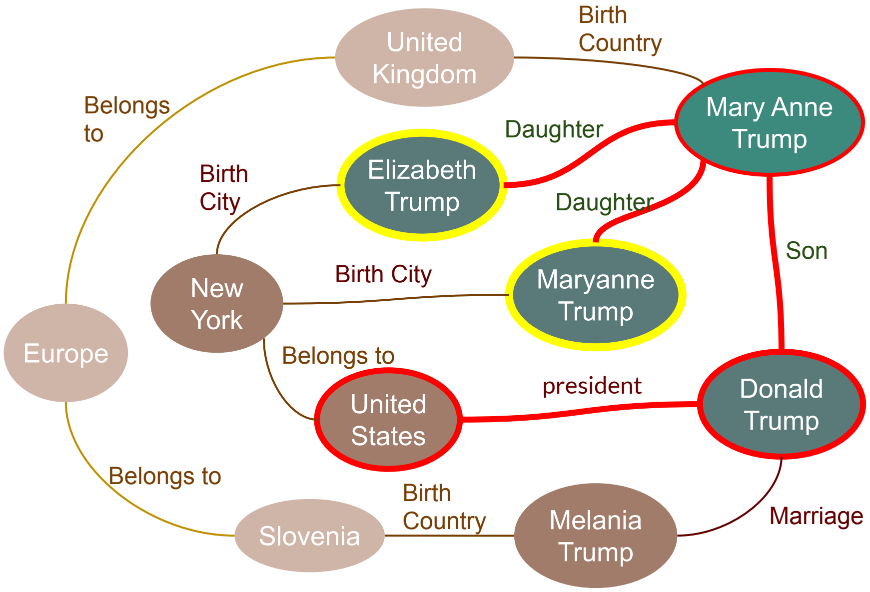
Question:
Who is the sister of the president of the United States?

(United States, President, Donald Trump)

(Donald Trump, Mother, Mary Anne Trump)

(Mary Anne Trump, Daughter, Maryanne/Elizabeth)

Answer:
Maryanne Trump / Elizabeth Trump



What is Knowledge based Question Answering

Question:

Who is the sister of the president of the United States?

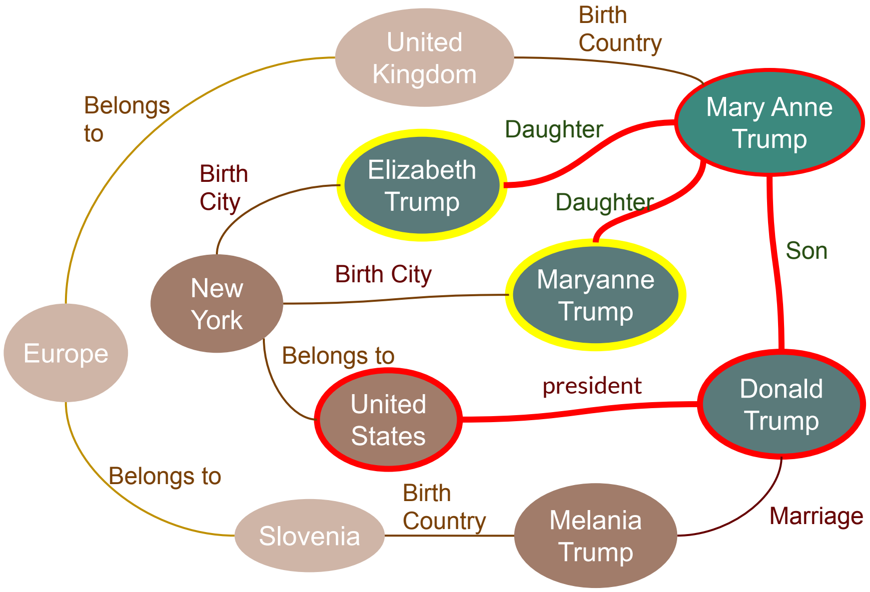
(United States, President, Donald Trump)

(Donald Trump, Mother, Mary Anne Trump)

(Mary Anne Trump, Daughter, Maryanne/Elizabeth)

Answer:

Maryanne Trump / Elizabeth Trump



Latent info = reasoning path (Highlight in red)

Reasoning Path as Latent Variable

$$p(y|x) = \sum_z p(y|z)p(z|x)$$

x: question

Who is the sister of the president of the United States?

z: reasoning path

United States→President→Donald Trump→Mother→Mary Anne Trump→Daughter→

y: answer

Maryanne Trump / Elizabeth Trump

Notations

For a given question x, a reasoning path z is a sequence in the form:

$$z = e_0 \rightarrow r_1 \rightarrow e_1 \rightarrow, \ldots, \rightarrow e_{T-1} \rightarrow r_T$$

that points to the answer:

$$z \rightarrow (e_T = y)$$

Notations

$$p(y|x) = \sum_z p(y|z)p(z|x)$$

$$p(y|z)=p(e_T|e_0,r_1,e_1,r_2,\ldots,e_{T-1},r_T)$$

$$p(z|x)=p(e_0,r_1,e_1,r_2,\ldots,e_{T-1},r_T|x)=p(e_0|x)p(r_1|x,e_0)p(e_1|x,e_0,r_1)\ldots p(r_T|x,e_0,r_1,\ldots,e_{T-1})$$

Notations

$$p(y|x) = \sum_z p(y|z)p(z|x)$$

$$p(y|z)=p(e_T|e_0,r_1,e_1,r_2,\ldots,e_{T-1},r_T)$$

$$p(z|x)=p(e_0,r_1,e_1,r_2,\ldots,e_{T-1},r_T|x)=p(e_0|x)p(r_1|x,e_0)p(e_1|x,e_0,r_1)\ldots p(r_T|x,e_0,r_1,\ldots,e_{T-1})$$

Notations

$$p(y|x) = \sum_z p(y|z)p(z|x)$$

$p(y|z)=p(e_T|e_0,r_1,e_1,r_2,\ldots,e_{T-1},r_T)=p(e_T|e_{T-1},r_T)$

$p(z|x)=p(e_0,r_1,e_1,r_2,\ldots,e_{T-1},r_T|x)=p(e_0|x)p(r_1|x,e_0)p(e_1|x,e_0,r_1)\ldots p(r_T|x,e_0,r_1,\ldots,e_{T-1})$

We just need to model two terms $p(e|*)$ and $p(r|*)$.

Entity Probability *p(e|*)*

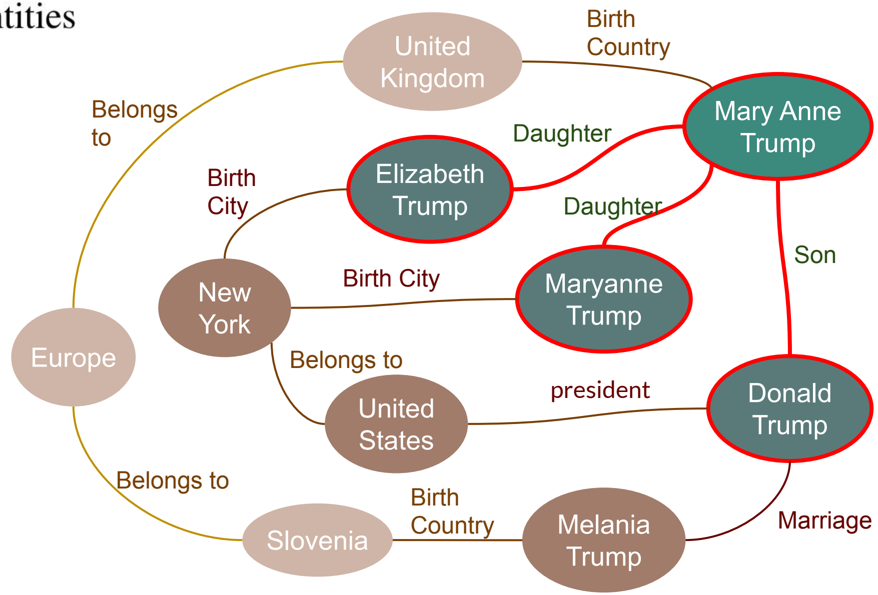
$$p(e_t|e_{t-1}, r_t) = \begin{cases} 1/M & \text{if } e_t \text{ is one of the } M \text{ matched entities} \\ 0 & \text{if } e_t \text{ is not a matched entity} \end{cases}$$

p(Elizabeth_Trump|...,Daughter,Mary Anne)=½

p(Maryanne_Trump|...,Daughter,Mary Anne)=½

p(Donald_Trump|...,Daughter,Mary Anne)=0

p(Donald_Trump|...,Son,Mary Anne)=1



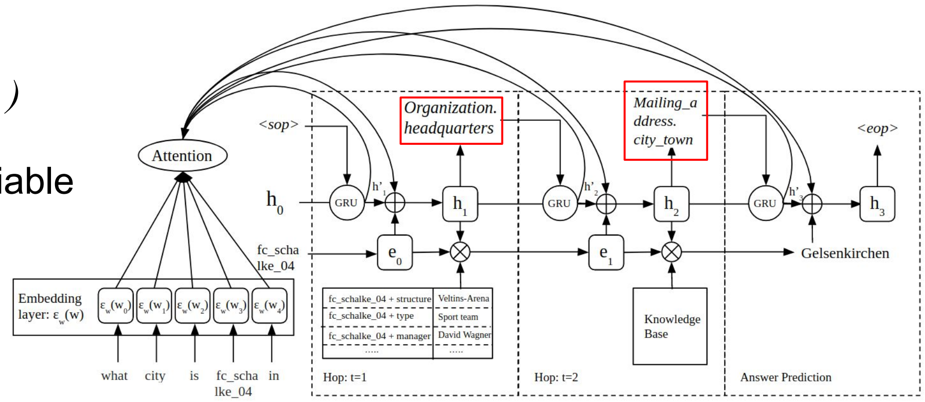
Relation Probability $p(r|*)$

At each timestep t, given r_{t-1} and e_{t-1}, we estimate $p(r_t|...)$ using a recurrent structure:

$$p(r_t|e_0,r_1,...,e_{t-1})=softmax(\ [f(e_0,...,e_{t-1});\ f(r_1,...,r_{t-1});\ f(x)]\ )$$

Where $f(*)$ is a mapping function from random variable to its vector representation.

Therefore $f(e_{t-1})$, $f(r_{t-1})$, and $f(x)$ are vector representations of the previous entity, previous relation, and the input query.

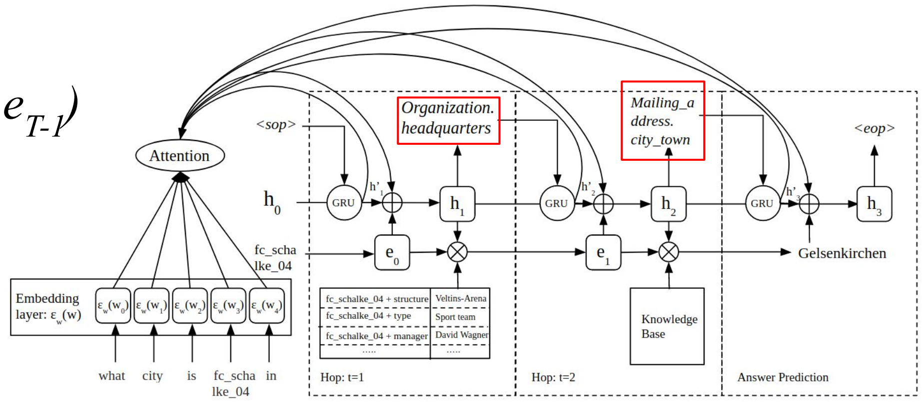


Latent Reasoning Path Prediction $p(z|x)$

$$p(z|x)$$
$$=p(e_0,r_1,e_1,r_2,\ldots,e_{T-1},r_T|x)$$
$$=p(e_0)p(r_1|e_0)p(e_1|e_0,r_1)\ldots p(r_T|e_0,r_1,e_1,r_2,\ldots,e_{T-1})$$

1. e_0 is identified by entity linking tool.

2. At each timestep t, we estimate $p(r_t|*)$ and $p(e_t|*)$ as discussed.



Estimate Values of z in Preprocessing

$$p(y|x) = \boxed{\sum_z} p(y|z)p(z|x)$$

To train the model without using labeled z, we use graph algorithm to select reasoning paths from the graph.

Preliminary Experimental Results

Properties:

- **Model multiple reasoning paths:** consider multiple reasoning paths for each question answer pair make the model more stable than using a single path in most existing work.
- **Reasoning path as latent variable:** our model can be trained without using labeled reasoning paths.
- **Easy to implement:** fit with any base models (we use RNN structure).

	Extra Supervision	Model $p(e)$	Different Setup	WQSP	CWQ
STAGG_SP	Y		Semantic Parsing	**71.7**	-
HR-BiLSTM	Y			62.3	31.2
KBQA-GST	Y	Y		67.9	36.5
NSM	Y		Neural Program Generation	**69.0**	-
KV-MemNN				38.6	-
STAGG_Answer			Semantic Parsing	66.8	-
GRAFT-Net		Y		62.8	26.0
Our Method		Y		67.9	**41.9**

Proposed Work: Advanced Path Selection

$$p(y|x) = \boxed{\sum_{z}} p(y|z)p(z|x)$$

The summation makes training process intractable.

We need to consider all valid paths between e_0 and e_{answer}.

A real-world knowledge graph contains **billions** of entity-relation facts. Between two nodes, there are a very large number of valid paths!

More importantly, not all the valid paths are good enough to serve as a reasoning path.

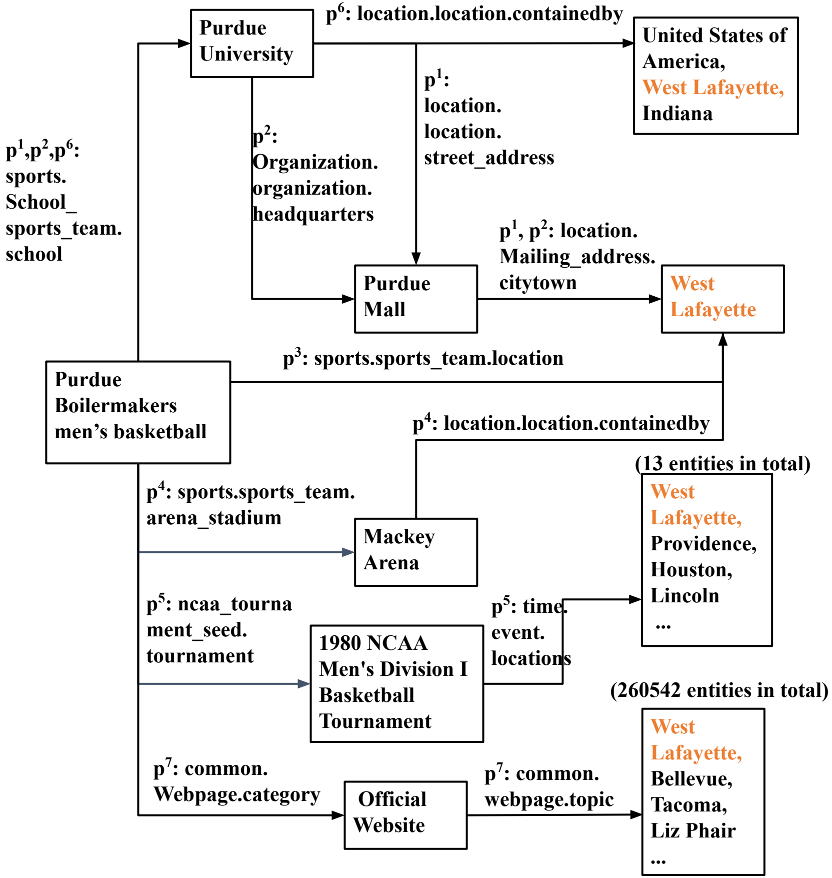
Path Selection - Rule #1

Question:

What city is home to the University that is known for Purdue Boilermakers men's basketball?

Answer:

West Lafayette



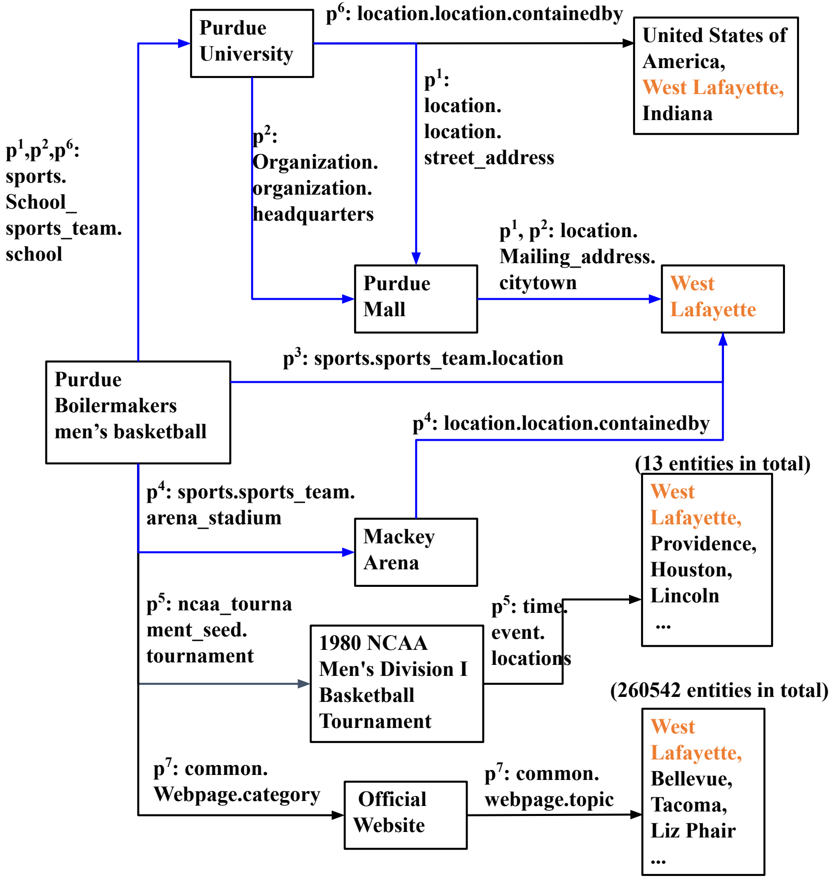
Path Selection - Rule #1

Question:

What city is home to the University that is known for Purdue Boilermakers men's basketball?

Answer:

West Lafayette



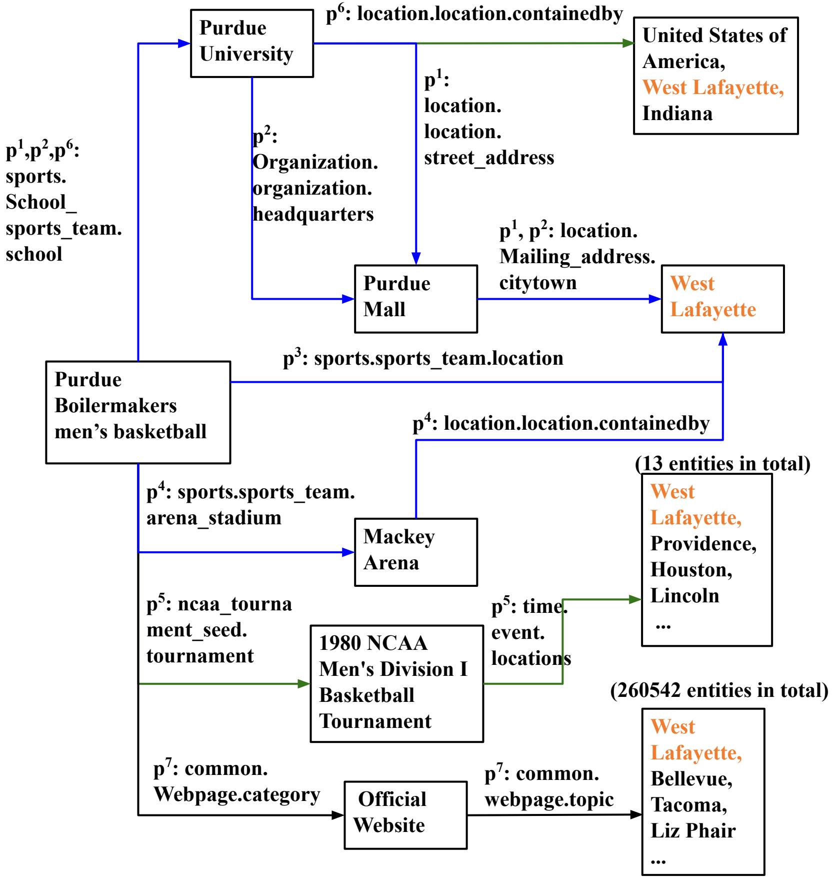
Path Selection - Rule #1

Question:

What city is home to the University that is known for Purdue Boilermakers men's basketball?

Answer:

West Lafayette



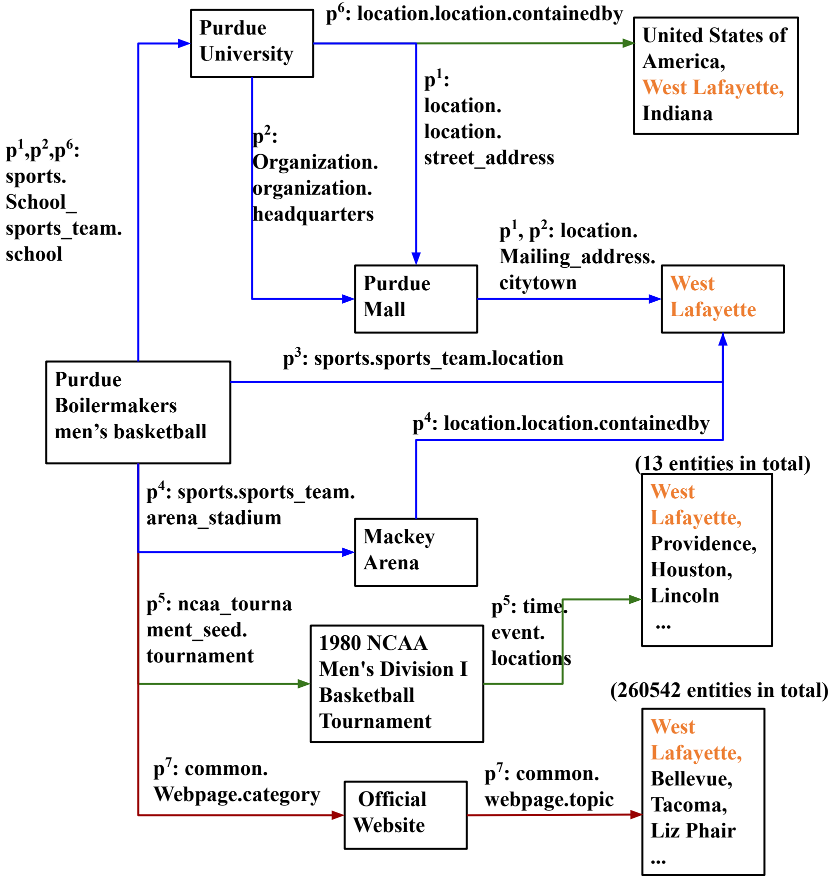
Path Selection - Rule #1

Question:

What city is home to the University that is known for Purdue Boilermakers men's basketball?

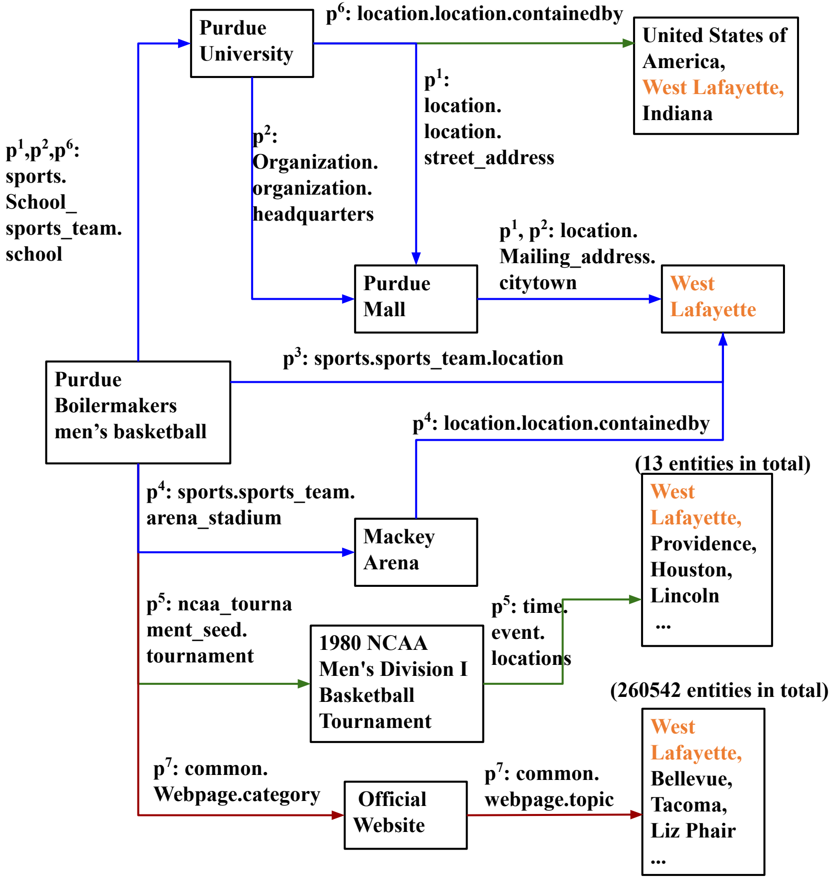
Answer:

West Lafayette



Path Selection - Rule #1

Rule 1: We want to filter out paths pointing to too many entities.



Path Selection - Rule #2

Question: Who was the owner of kfc?

Answer: Colonel Sanders

Path 1: kfc→organization.organization.founders→Colonel Sanders

Path 2: kfc→advertising_characters.product.advertising_characters→Colonel Sanders

Path Selection - Rule #2

Question: Who was the owner of kfc?

Answer: Colonel Sanders

Path 1: kfc→organization.organization.founders→Colonel Sanders

~~Path 2: kfc →advertising_characters.product.advertising_characters →Colonel Sanders~~

Rule 2: We want to filter out paths that are not relevant to the question.

Reasoning Path as Latent Variable

Step 1: Use graph algorithm to collect all valid paths between topic entity e_0 and answer e_{answer}.

Step 2: Select paths based on rule #1 and rule #2.

Step 3: Update model parameters by maximizing likelihood $p(y|x)$ based on selected paths.

Repeat step 2 and step 3 until the model converges.

Timeline

Timeline	Task
by June 2020	Designing evaluation experiments for QA task - Human identification - Major claim extraction - Discourse relation classification
by Winter 2020	Improving path selection - Use current trained model to select good paths - Use advanced bootstrapping methods to select good paths - Explore other directions to solve the problem - Evaluate performance of the proposed method
by Winter 2020	Refining model architecture - Neural Transformer - Memory Network - Propose novel model structures - Evaluate performance of the proposed model
by Summer 2021	Handling noisy tags in multi-label classification - Propose novel ideas to handle noisy tags - Propose novel model structures - Evaluate performance of the proposed model
by Fall 2021	Thesis writing and defense.

Thank you!

Questions?

Other Work

Use latent topic to predict a winner in a debate:

Winning on the Merits: The Joint Effects of Content and Style on Debate Outcomes (TACL), 2017.

Use latent conversation structure information to generate meeting minutes:

Joint Modeling of Content and Discourse Relations in Dialogues (ACL), 2017.

Capture label dependencies in multi-label prediction task:

Learning to Calibrate and Rerank Multi-label Predictions (ECML PKDD), 2019.

Ranking-Based AutoEncoder for Extreme Multi-label Classification (NAACL-HLT), 2019.

Supervised Learning

#Car=1

Input x *y=f(x)* *Output y*

Supervised Learning

#Car=3

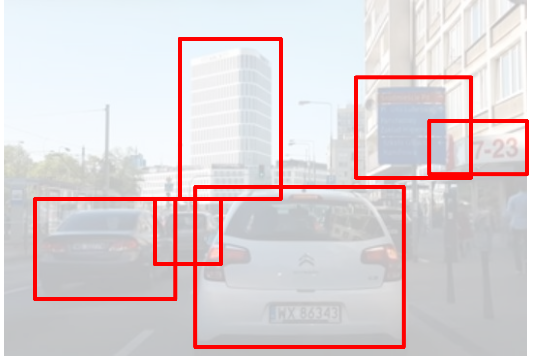
Input x $y=f(x)$ *Output y*

Supervised Learning

#Car=3

Input x $y=f(x)$ *Output y*

Supervised Learning with Latent Information

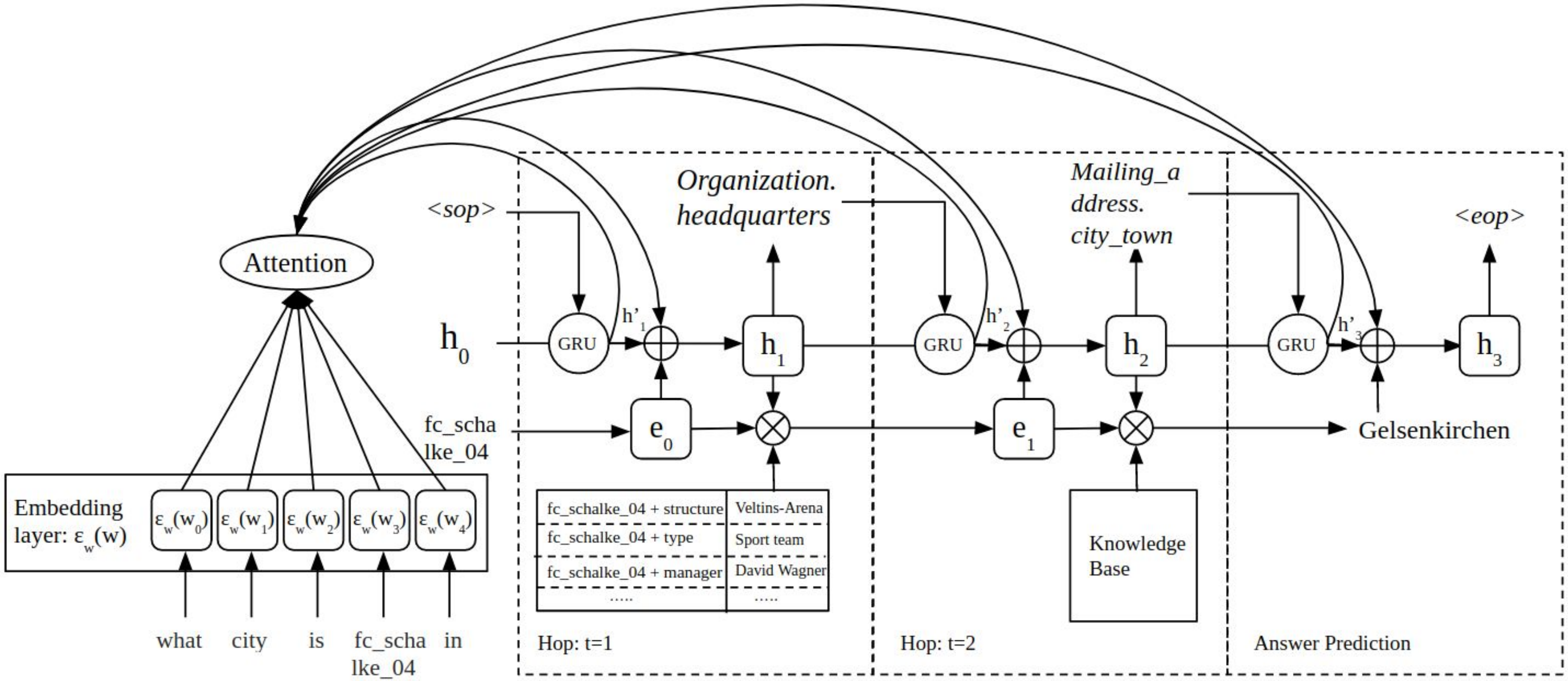


#Car=3

Input x *z=possible locations* *Output y*

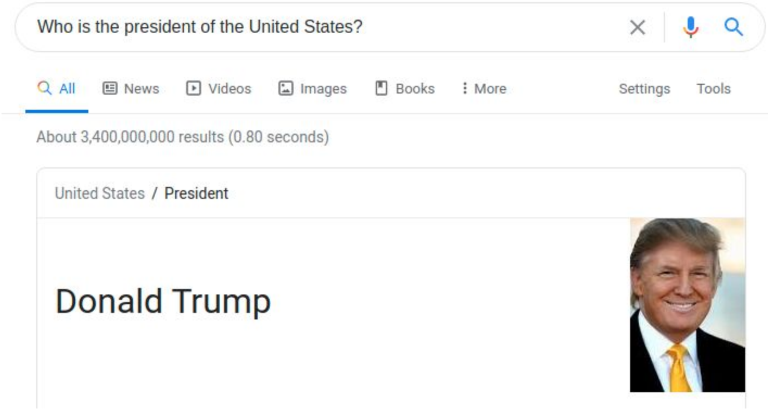
$z=f(x)$ $y=f(z)$

Model Structure



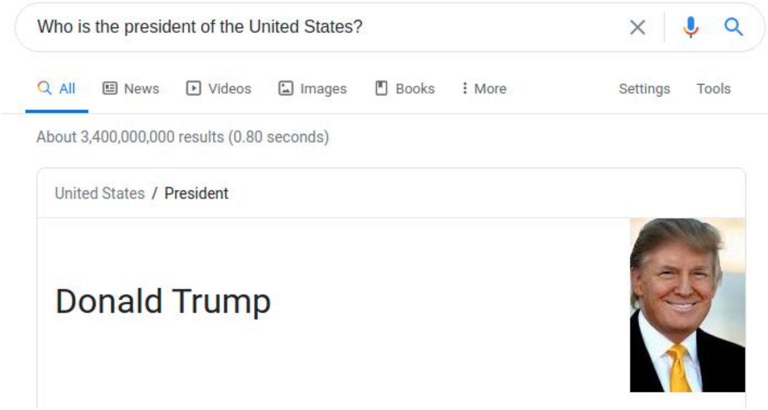
Supervised Learning

Who is the president of the United States?



Supervised Learning

Who is the president of the United States? ⟹ Feature x

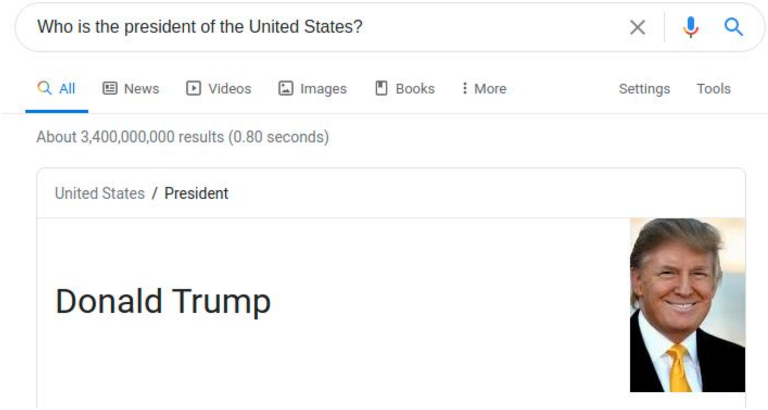


⟹ Target y

Supervised Learning

Who is the president of the United States? $\Rightarrow$ Feature x

$y=f(x)$

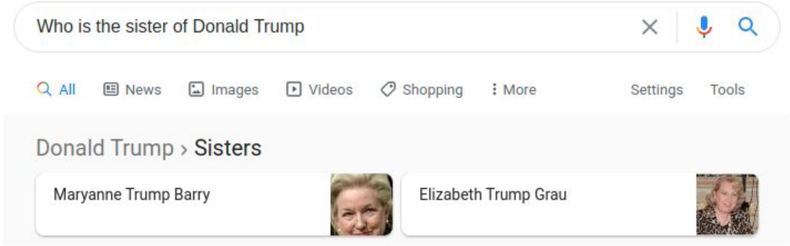


$\Rightarrow$ Target y

Supervised Learning

Who is the sister of Donald Trump?

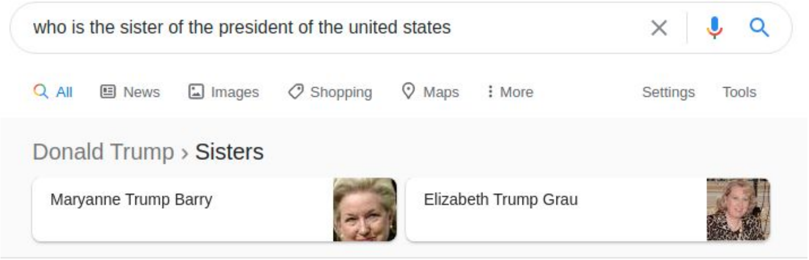
$$y=f(x)$$



Supervised Learning

Who is the sister of the president of the United States?

$y=f(x)$



?

Latent Information

Who is the sister of the president of the United States? x

$z=f(x)$

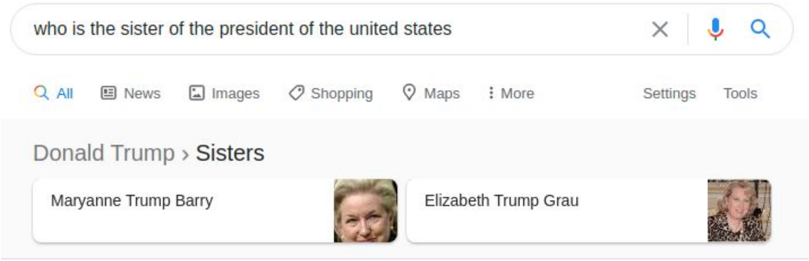
Who is the president of the United States? → Donald Trump

Who is the sister of Donald Trump?

Latent information z

$y=f(z)$

y



Latent Information

Who is the sister of the president of the United States? x

Who is the president of the United States?→ Donald Trump

$z=f(x)$

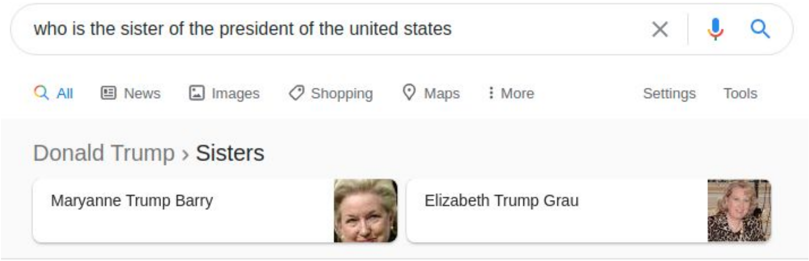
Who are the parents of Donald Trump?→ XXX

Who are the daughters of XXX?

Latent information z

$y=f(z)$

y

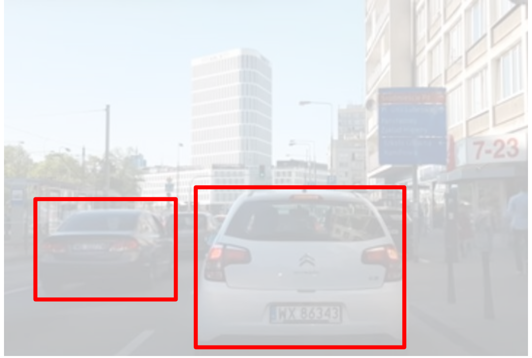


Supervised Learning

Car

Feature x $y=f(x)$ Target y

Supervised Learning with Latent Information

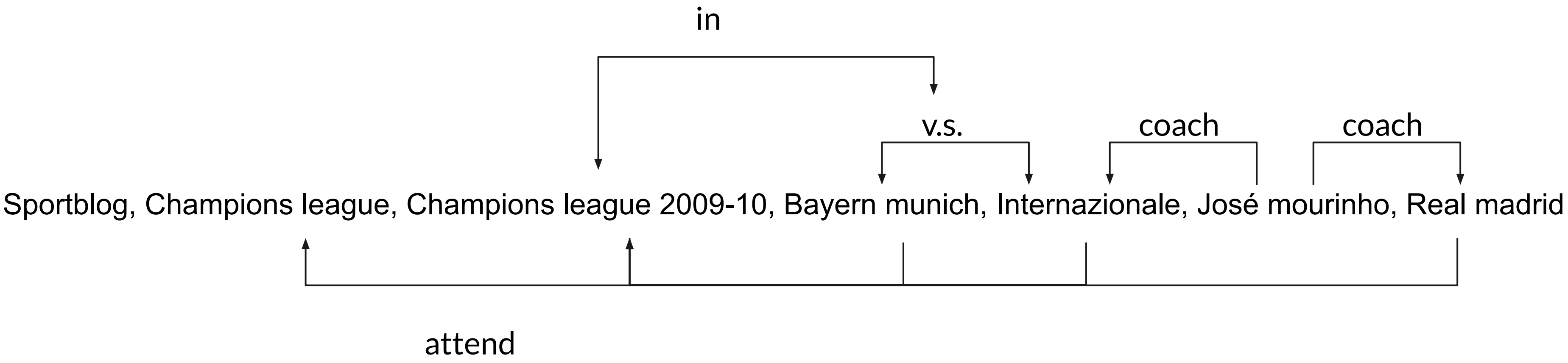


Car

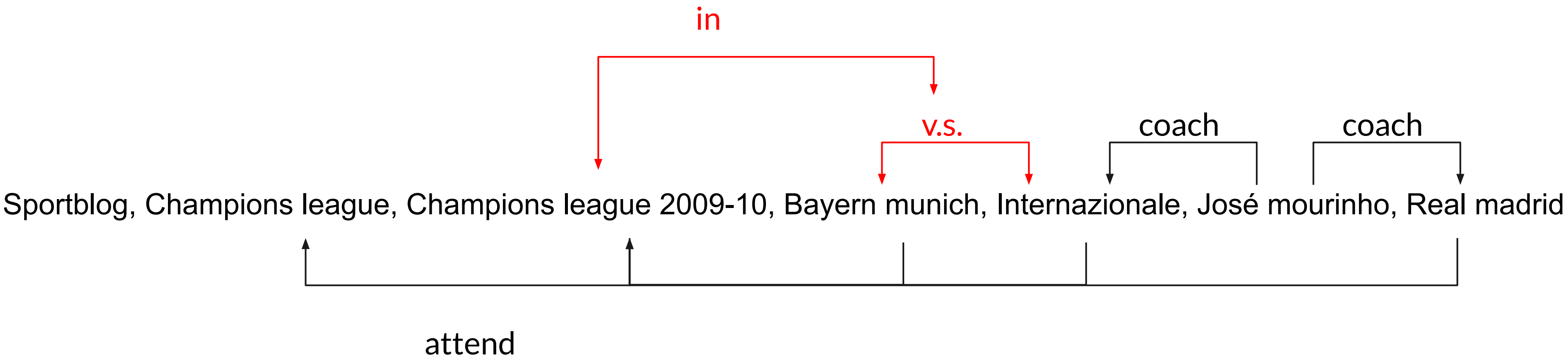
Feature x | Latent information z | Target y

$z=f(x)$ $y=f(z)$

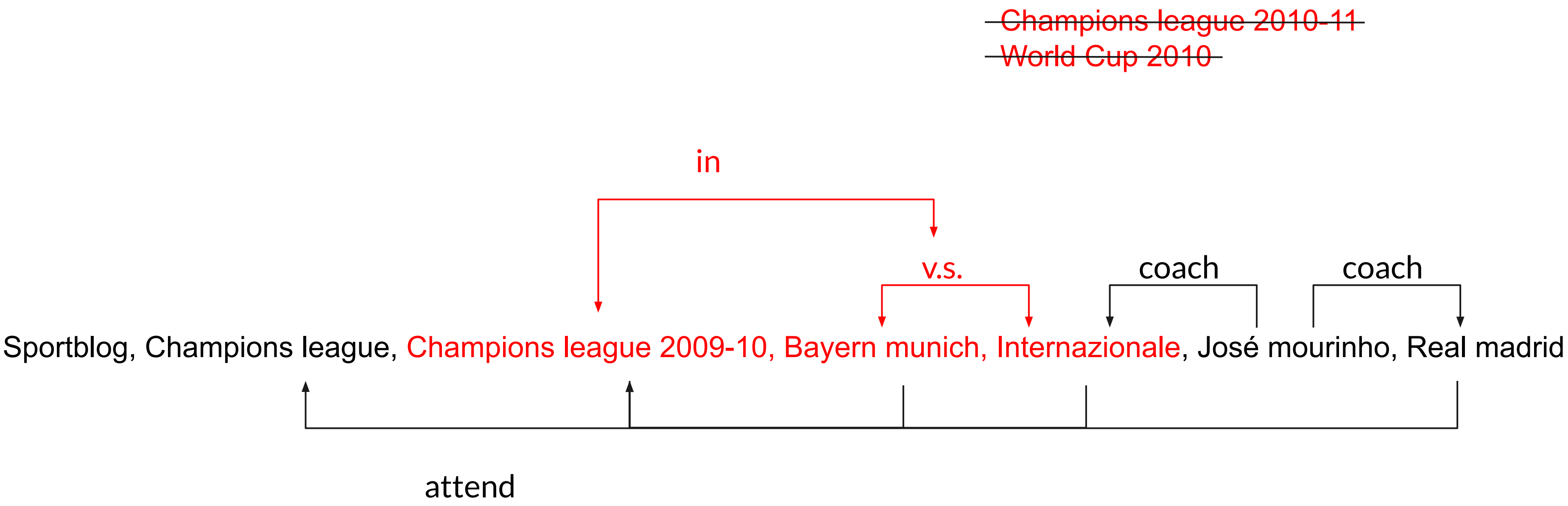
Label Dependencies



Label Dependencies



Label Dependencies



Different Ways to Sort Labels (classifiers)

Frequency:
Sportblog→Champions league→Real_madrid→José mourinho→Internazionale→Champions league 2009-10→Bayern munich

Hierarchy:
Sportblog→Champions league→Champions league 2009-10→Bayern munich→Internazionale→Real_madrid→José mourinho

Alphabeta:
Bayern munich→Champions league→Champions league 2009-10→Internazionale→José mourinho→Real_madrid→Sportblog

What is latent information?

Answer Prediction $p(y|z)$

$(e_0, r_1, e_1, r_2, \ldots, e_{T-1}, r_T) \rightarrow e_{T-1} = y$, Our final goal is to estimate answer y.

$$p(e_t|e_{t-1}, r_t) = \begin{cases} 1/M & \text{if } e_t \text{ is one of the } M \text{ matched entities} \\ 0 & \text{if } e_t \text{ is not a matched entity} \end{cases}$$

Probabilistic Classifier Chain (PCC)

***x*:**

José Mourinho's treble – now for the Real story

Champions League glory completes the set for Inter but José Mourinho looks certain to quit for Real Madrid

▲ Jose Mourinho, the coach of Internazionale, during the Champions League final. Photograph: Jason Cairnduff/Action Images

José Mourinho's only problem is that he will run out of targets. A first league title for Chelsea in 50 years, Inter's first European Cup crown since 1965 and now the chance to manage Cristiano Ronaldo and Kaká at Real Madrid.

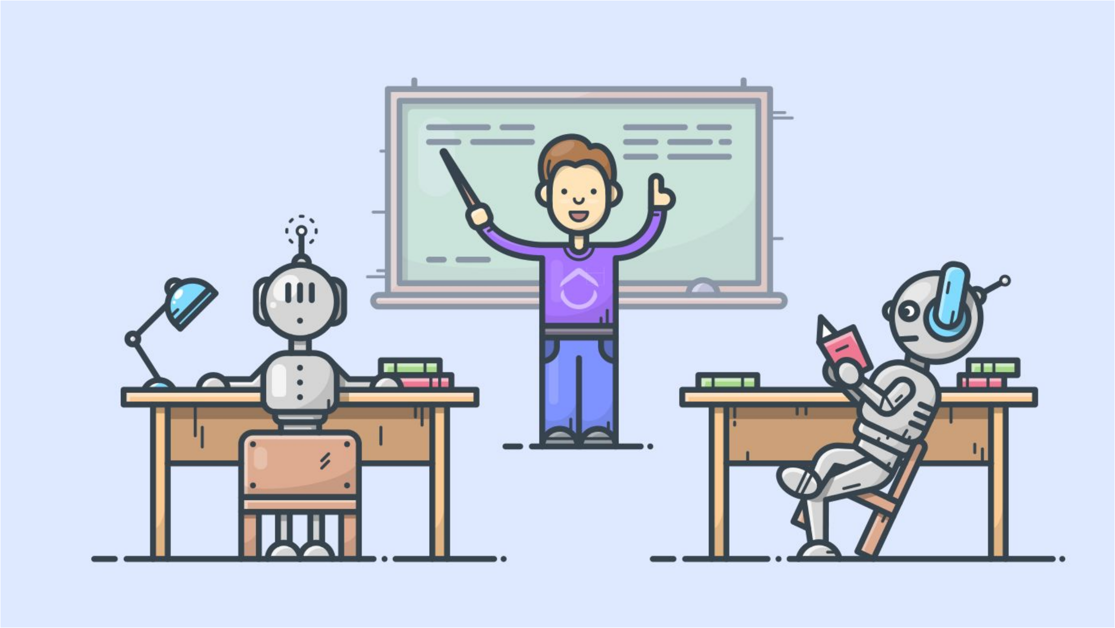
"I want to become the only coach to win the Champions League with three different clubs. I'm not leaving Inter, I'm leaving Italy," Mourinho said after Inter's 2-0 victory over Bayern Munich on a melodramatic night, thus confirming an open secret. A European champion with Porto six years ago,

$b_1(y_1|x)$
$b_2(y_2|x,y_1)$
$b_3(y_3|x,y_1,y_2)$
…
$b_n(y_n|x,y_1,\ldots,y_{n-1})$

***y*:**

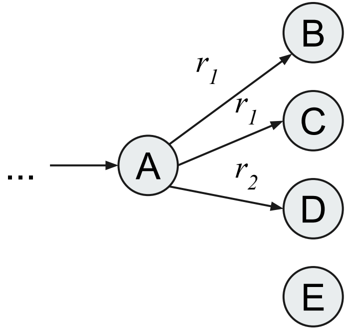
Champions league→Sportblog→
José mourinho→Internazionale→
Real_madrid→Bayern munich→
Champions league 2009-10

Teach Machines to Think like Humans



Entity Probability *p(e|*)*

$$p(e_t|e_{t-1}, r_t) = \begin{cases} 1/M & \text{if } e_t \text{ is one of the } M \text{ matched entities} \\ 0 & \text{if } e_t \text{ is not a matched entity} \end{cases}$$



$p(B|\ldots,A,r_1)=1/2$

$p(C|\ldots,A,r_1)=1/2$

$p(B|\ldots,A,r_2)=0$

$p(D|\ldots,A,r_2)=1$

$p(E|\ldots,A,r_*)=0$

Different Ways to Sort Labels (classifiers)

Alphabeta:
Bayern munich→Champions league→Champions league 2009-10→Internazionale→José mourinho→Real_madrid→Sportblog

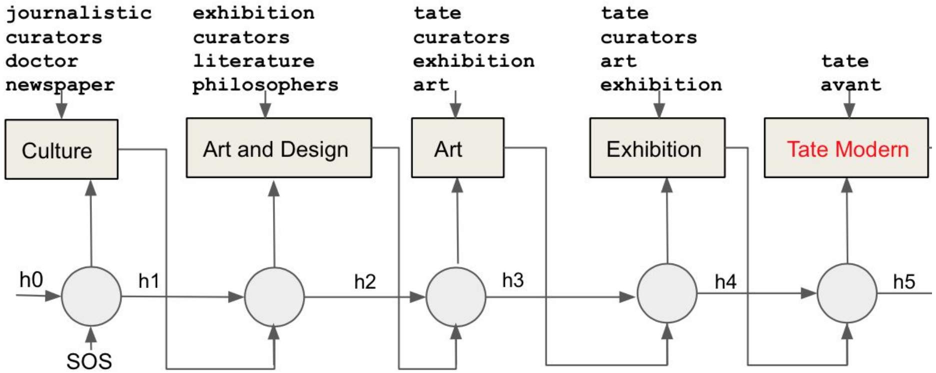
Frequency:
Sportblog→Champions league→Real_madrid→José mourinho→Internazionale→Champions league 2009-10→Bayern munich

Hierarchy:
Sportblog→Champions league→Champions league 2009-10→Bayern munich→Internazionale→Real_madrid→José mourinho

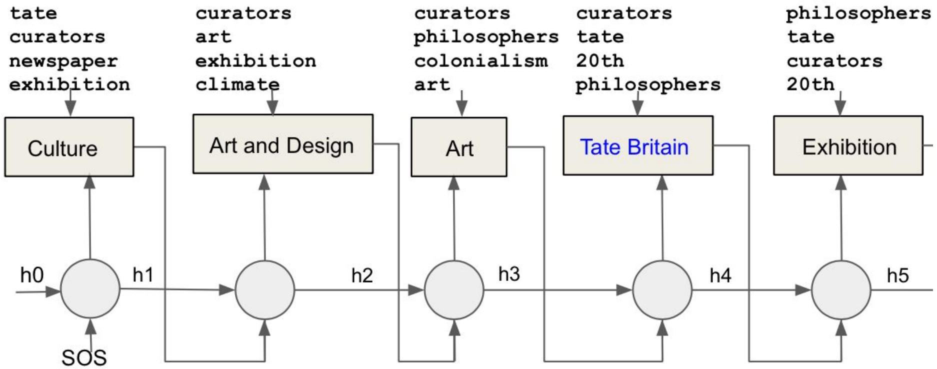
Manually:
Sportblog→Champions league→Champions league 2009-10→Bayern munich→Internazionale→José mourinho→Real_madrid

Case Study

RNN trained with fixed label order:



RNN trained with latent label order:



More examples of Latent Variable Models

- Gaussian Mixture Models (GMMs)
- Latent Dirichlet Allocation (LDA)
- Probabilistic Latent Semantic Analysis (pLSA)
- Hidden Markov Models (HMMs)
- Principal Component Analysis (PCA)
- ...

Problem of Using a Predefined Label Order

x:

José Mourinho's treble - now for the Real story

Champions League glory completes the set for Inter but José Mourinho looks certain to quit for Real Madrid

▲ Jose Mourinho, the coach of Internazionale, during the Champions League final. Photograph: Jason Cairnduff/Action Images

José Mourinho's only problem is that he will run out of targets. A first league title for Chelsea in 50 years, Inter's first European Cup crown since 1965 and now the chance to manage Cristiano Ronaldo and Kaká at Real Madrid.

"I want to become the only coach to win the Champions League with three different clubs. I'm not leaving Inter, I'm leaving Italy," Mourinho said after Inter's 2-0 victory over Bayern Munich on a melodramatic night, thus confirming an open secret. A European champion with Porto six years ago,

$y_1,\ldots,y_t$:

Frequency:

Sportblog→Champions league→Real_madrid→José mourinho→y_t=Cristiano Ronaldo

Hierarchy:

Sportblog→Champions league→Champions league 2009-10→Bayern munich→y_t=Internazionale (→Real_madrid→José mourinho)

ORDER MATTERS!

That is our latent information!

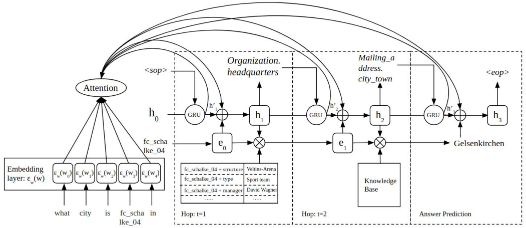
Rule 1: filter out paths leading to too many entities

$$p(y|x) = \sum_z p(y|z)p(z|x)$$

$$p(e_t|e_{t-1}, r_t) = \begin{cases} 1/M & \text{if } e_t \text{ is one of the } M \text{ matched entities} \\ 0 & \text{if } e_t \text{ is not a matched entity} \end{cases}$$

Rule 2: filter out irrelevant paths

Question: Who was the owner of kfc?

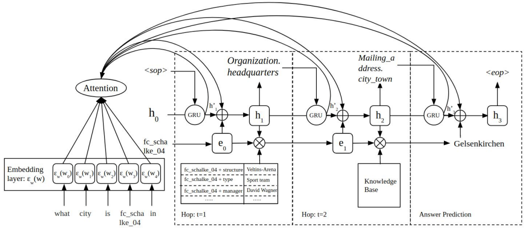


$$p(kfc \rightarrow organization.organization.founders \rightarrow Colonel\ Sanders|x) = 0.8$$

$$p(kfc \rightarrow advertising\_characters.product.advertising\_characters \rightarrow Colonel\ Sanders) = 0.2$$

Rule 2: filter out irrelevant paths

Question: Who was the owner of kfc?



$$p(kfc \rightarrow organization.organization.founders \rightarrow Colonel\ Sanders|x) = 0.8$$

$$\cancel{p(kfc \rightarrow advertising\_characters.product.advertising\_characters \rightarrow Colonel\ Sanders) = 0.2}$$