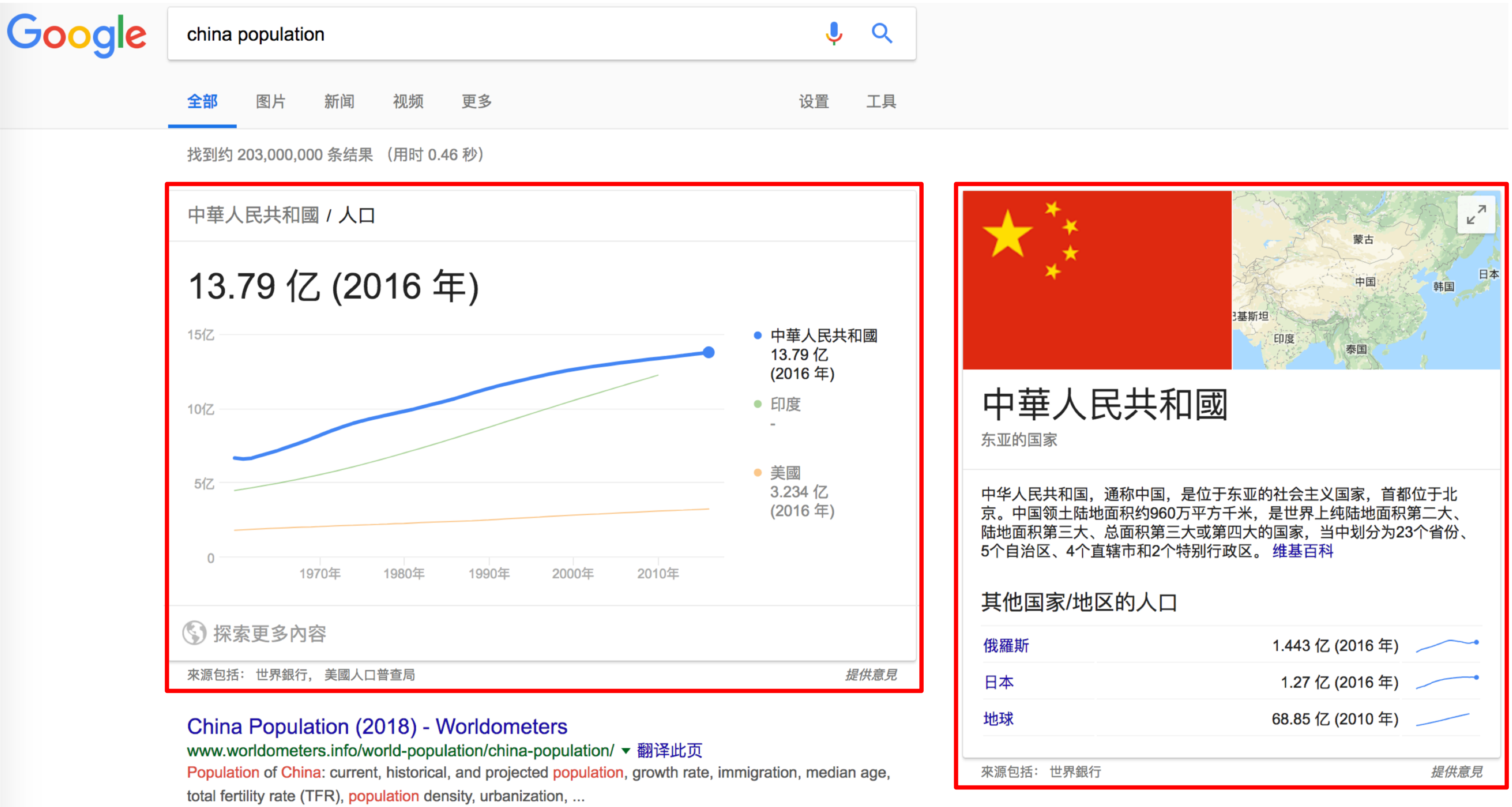
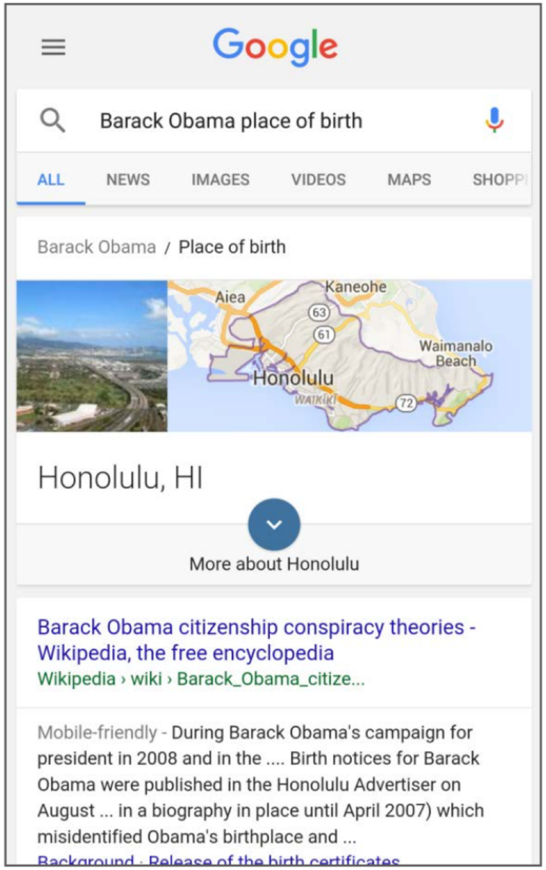
Knowledge Graph: Introduction, Problem and Application

Xusheng Luo (棋落)
Sept 06, 2018

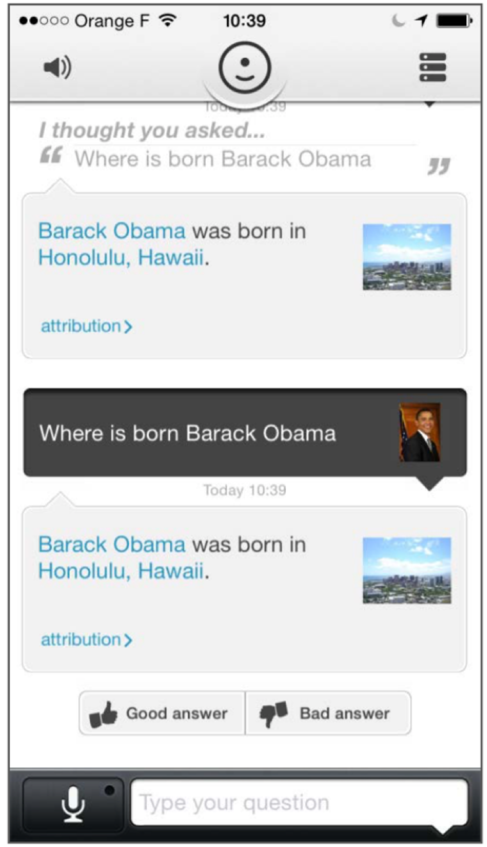

Google
china population
全部
图片
新闻
视频
更多
设置
工具
找到约 203,000,000 条结果 （用时 0.46 秒）
中華人民共和國 / 人口
13.79 亿 (2016 年)
15亿
10亿
5亿
0
1970年
1980年
1990年
2000年
2010年
中華人民共和國
13.79 亿
(2016 年)
印度
-
美國
3.234 亿
(2016 年)
探索更多內容
來源包括： 世界銀行， 美國人口普查局
提供意見
China Population (2018) - Worldometers
www.worldometers.info/world-population/china-population/ ▾ 翻译此页
Population of China: current, historical, and projected population, growth rate, immigration, median age, total fertility rate (TFR), population density, urbanization, ...
蒙古
中国
日本
韩国
印度
泰国
中華人民共和國
东亚的国家
中华人民共和国，通称中国，是位于东亚的社会主义国家，首都位于北京。中国领土陆地面积约960万平方千米，是世界上纯陆地面积第二大、陆地面积第三大、总面积第三大或第四大的国家，当中划分为23个省份、5个自治区、4个直辖市和2个特别行政区。 维基百科
其他国家/地区的人口
俄羅斯 1.443 亿 (2016 年)
日本 1.27 亿 (2016 年)
地球 68.85 亿 (2010 年)
來源包括： 世界銀行
提供意見


Bai du 百度
姚明老婆的身高
百度一下
网页 资讯 知道 图片 视频 贴吧 文库 地图 音乐 更多»
百度为您找到相关结果约2,270,000个
搜索工具
叶莉身高：
190cm
叶莉，1981年11月20日出生，上海人。1996年进入上海体育运动技术学院，1997年在上海体育运动技术学院，1998年入选国家青年队，1999年第一次入选国家队。2002年5月，叶莉... 详情>>
来自百度百科


搜狗搜索
姚明妻子的女儿的身高
搜狗搜索
网页 新闻 微信 知乎 图片 视频 明医 英文 问问 学术 更多
搜狗已为您找到约18,789条相关结果
全部时间
姚沁蕾身高
110cm
姚沁蕾是著名篮球明星姚明和叶莉的女儿。2010年5月22日凌晨出生于休斯敦，体重3.6公斤。2011年4月21日姚明在上海梅陇基地召开了回国后的首次媒... 详情>>
搜狗立知 | 纠错



Google



EVI

(Amazon)

Siri

(Apple)

Structured search within the graph

Outline

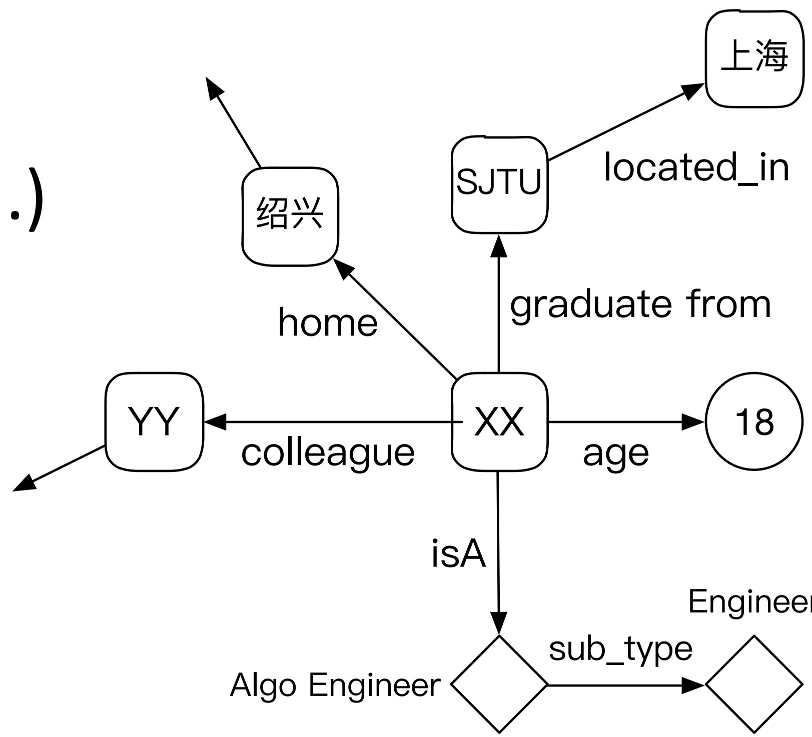
- Introduction to Knowledge Graphs
- Applications
  – Entity Linking
  – Factoid Question Answering
  – E-Commerce Application
  – Knowledge Based Recommendation
  – ……

Outline

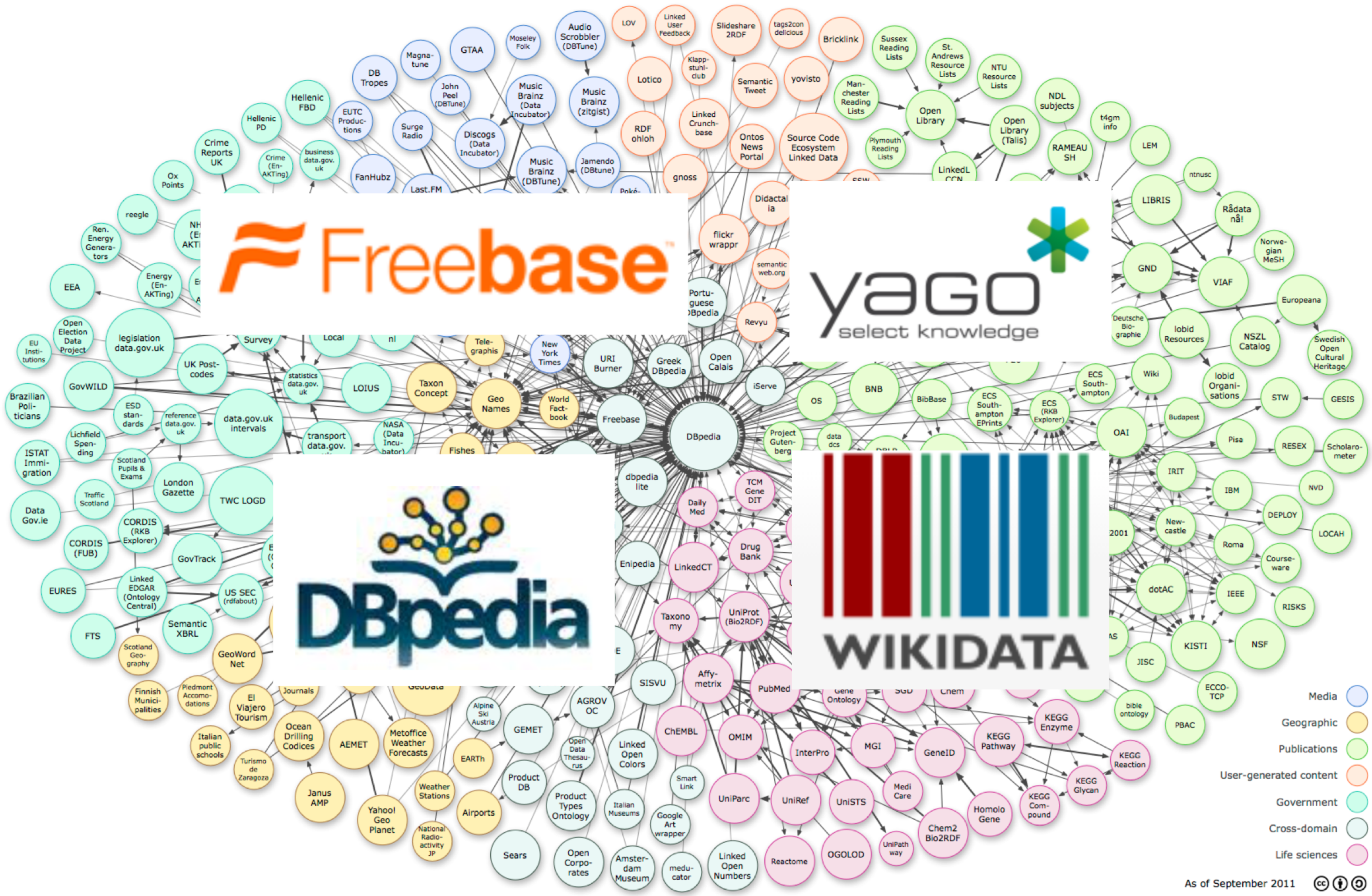
- Introduction to Knowledge Graphs
- Applications
  - Entity Linking
  - Factoid Question Answering
  - E-Commerce Application
  - Knowledge Based Recommendation

What is a Knowledge Graph?

- Node
  - entity
  - concept (type/cate...)
  - value (str/num)
- Edge
  - relation
  - attribute



Common Knowledge Graphs



Common Knowledge Graphs

Name	Instances	Facts	Types	Relations
DBpedia (English)	4,806,150	176,043,129	735	2,813
YAGO	4,595,906	25,946,870	488,469	77
Freebase	49,947,845	3,041,722,635	26,507	37,781
Wikidata	15,602,060	65,993,797	23,157	1,673
NELL	2,006,896	432,845	285	425
OpenCyc	118,499	2,413,894	45,153	18,526
Google's Knowledge Graph	570,000,000	18,000,000,000	1,500	35,000
Google's Knowledge Vault	45,000,000	271,000,000	1,100	4,469
Yahoo! Knowledge Graph	3,443,743	1,391,054,990	250	800

2016

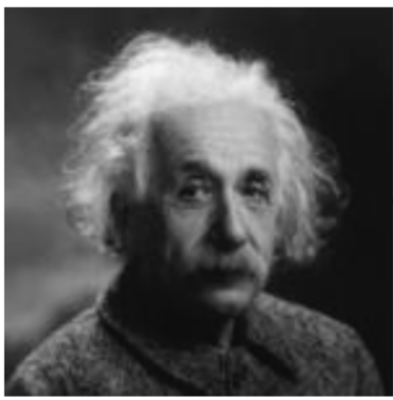
bing Satori

Google Search › Knowledge Graph Search API

NELL: Never-Ending Language Learning

Subject-Predicate-Object (SPO) triples in Freebase

</m/0jcx, **/m/04m8,** **/m/019xz9>**



/people/person/place_of_birth

Place of birth

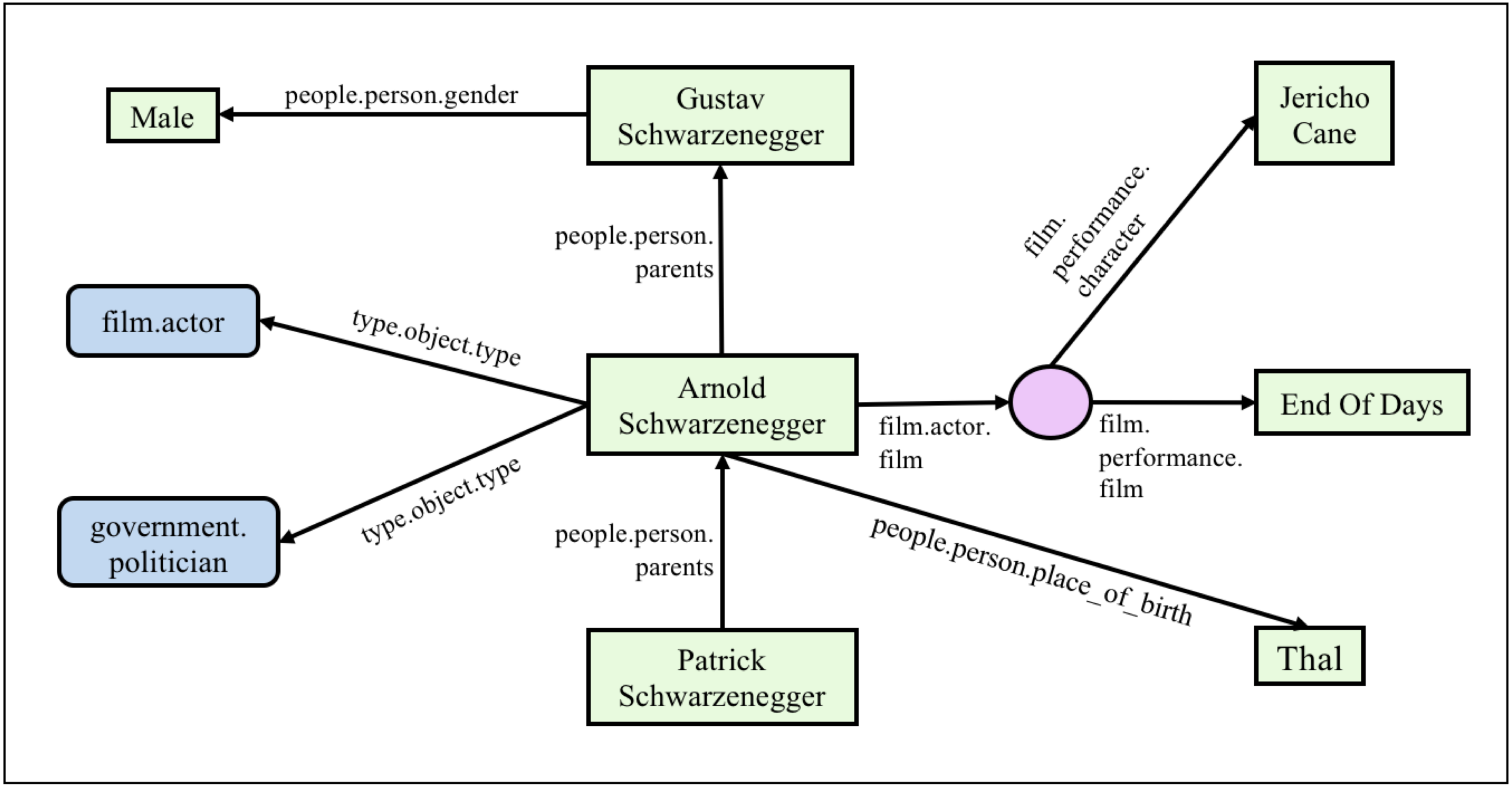
/en/albert_einstein

Albert Einstein

/en/ulm

Ulm

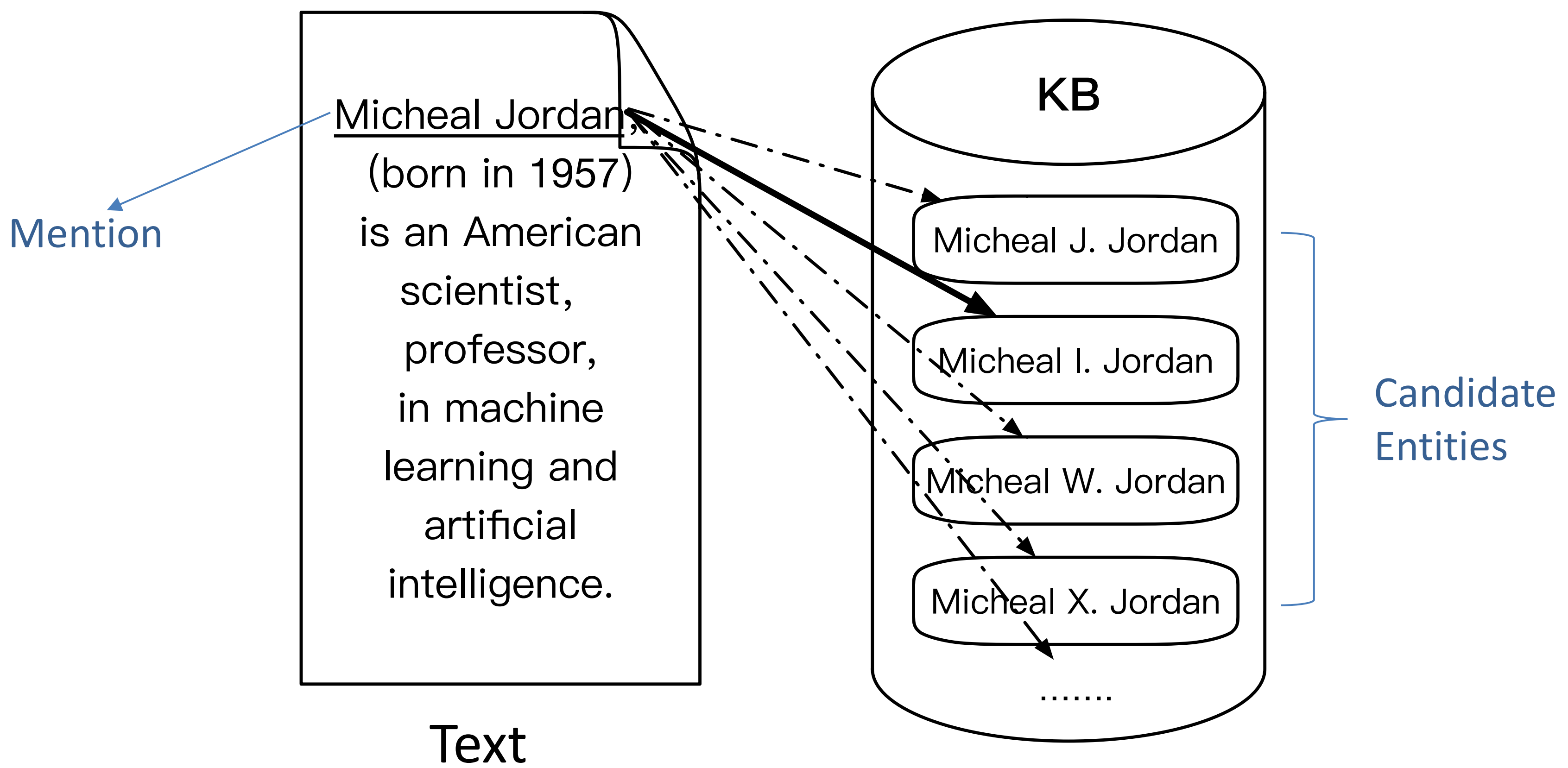
Freebase as an example



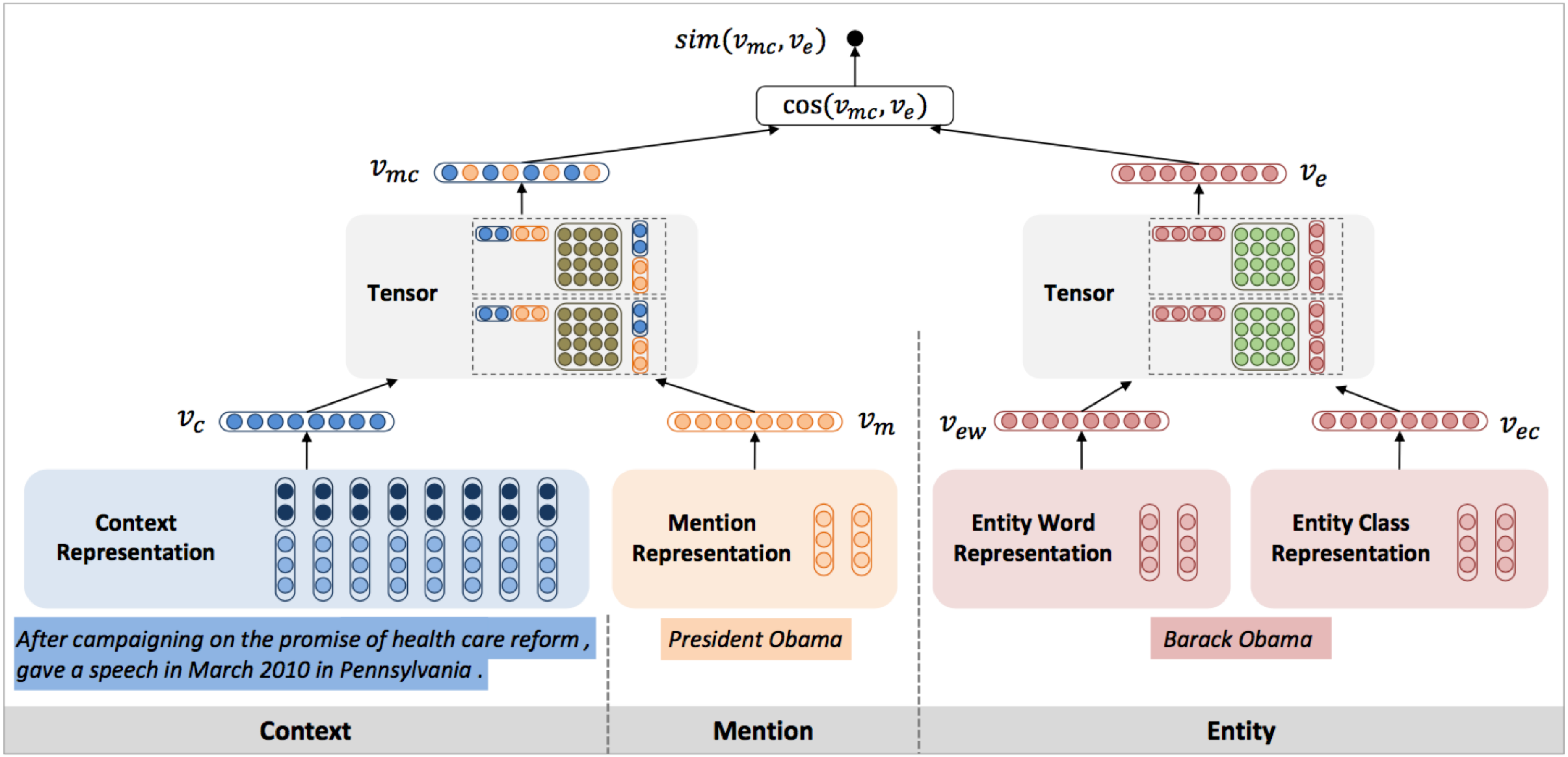
Outline

- Introduction to Knowledge Graphs
- Applications
  - Entity Linking
  - Factoid Question Answering
  - E-Commerce Application
  - Knowledge Based Recommendation

Entity Linking



Entity Linking

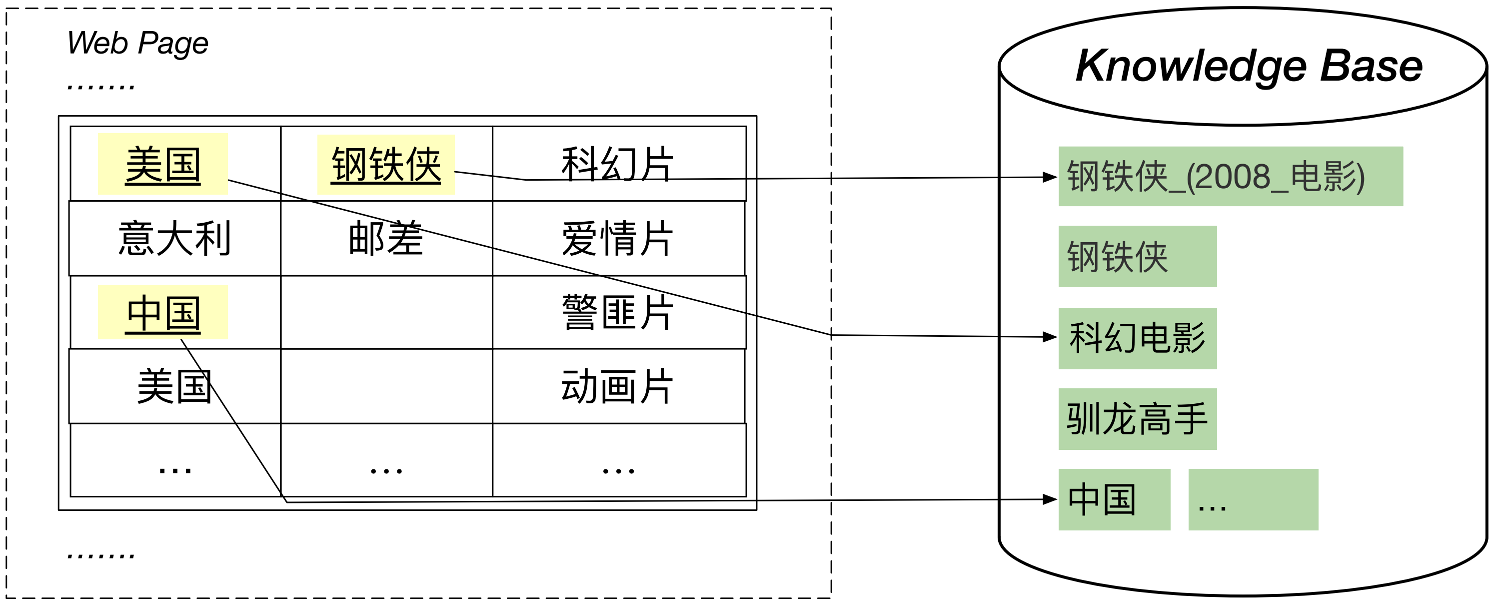


[Sun, 2015] ***IJCAI15***

Table Linking

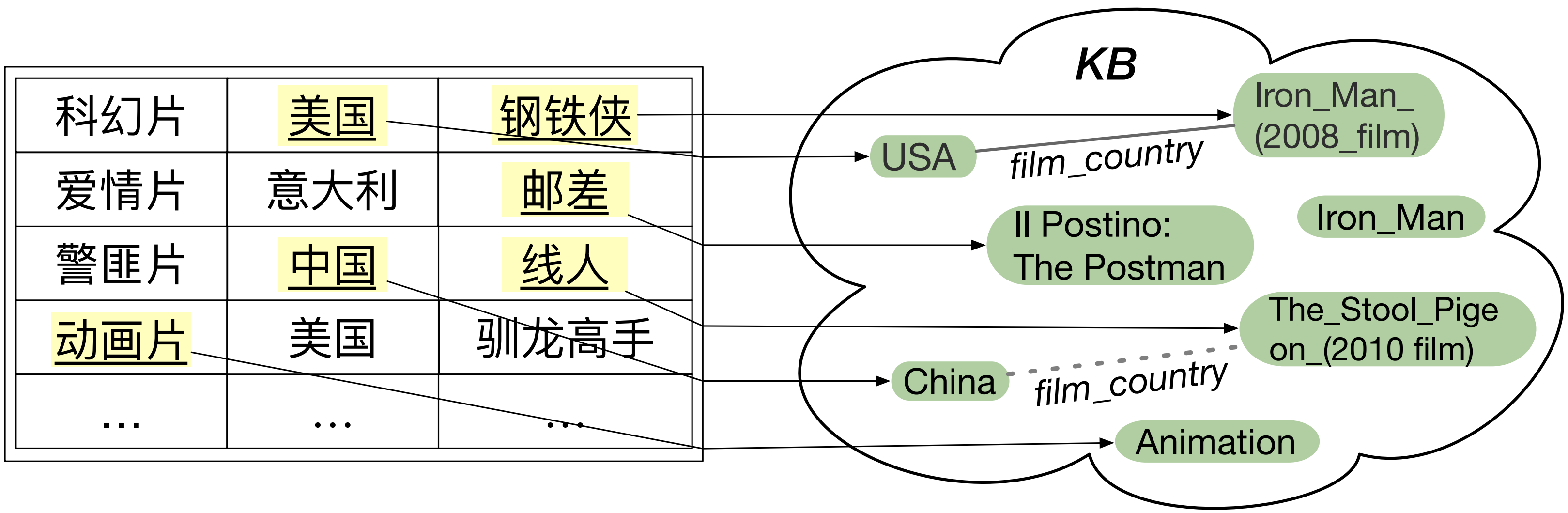
– Entity Linking for Web Tables

– Existing works are Mono-Lingual



Cross-lingual Entity Linking for Web Tables

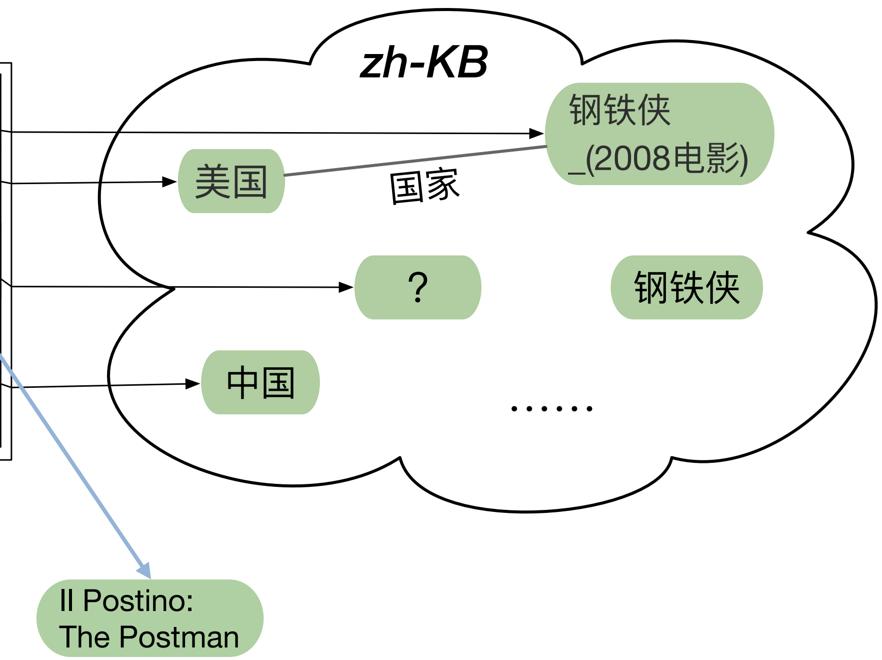
AAAI 2018



Motivation

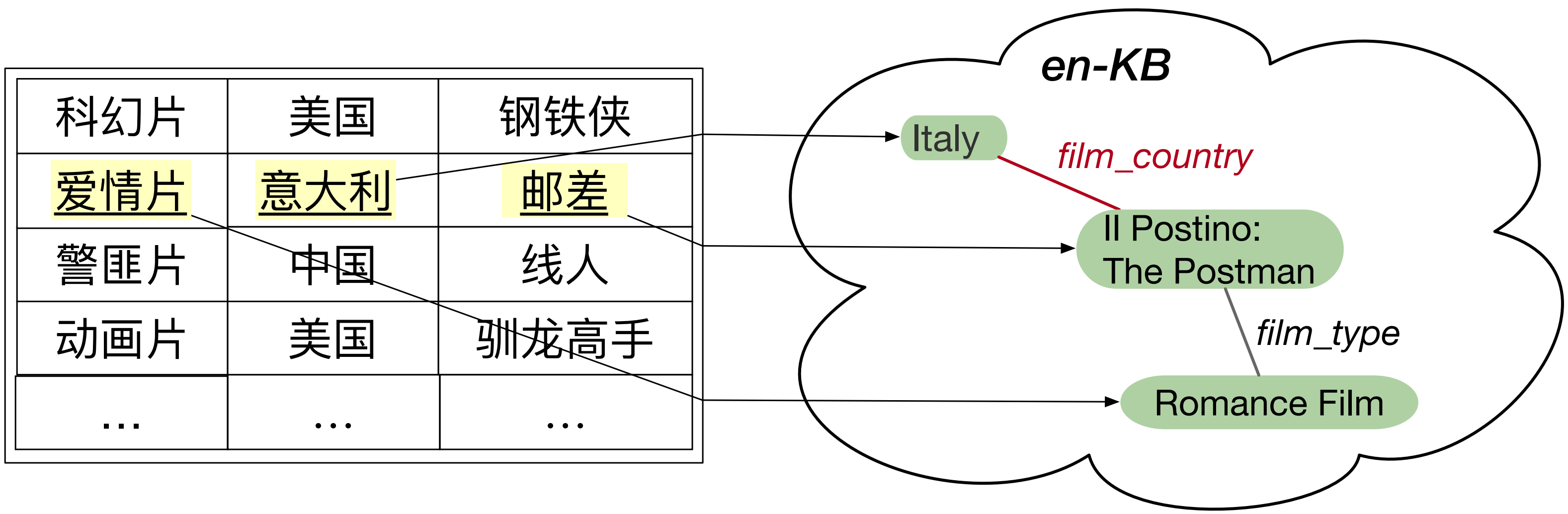
- non-English KB not comprehensive enough
  - Chinese Wiki: 1/6 size of English version
  - Increasing number of non-English Web tables

科幻片	美国	钢铁侠
爱情片	意大利	邮差
警匪片	中国	线人
动画片	美国	驯龙高手
…	…	…



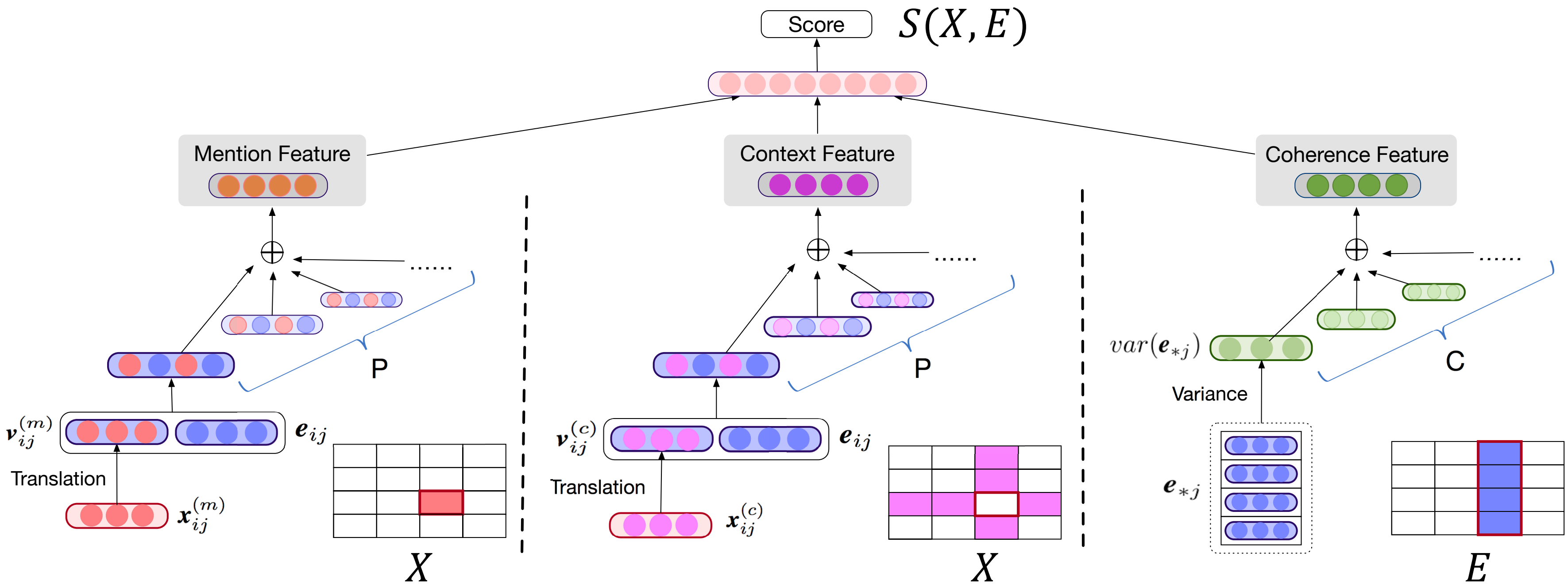
Another Motivation

- Enrich English KB with multi-lingual facts
  - English KB contains long-tail entities
  - non-English tables are rich sources for rare entities



TabEL Overview

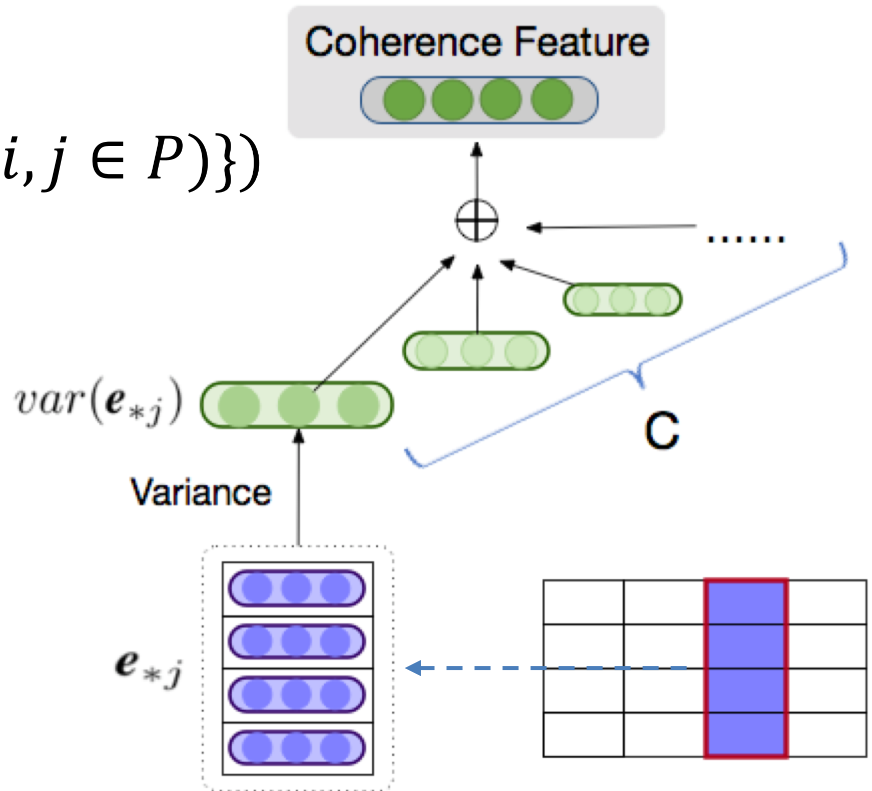
- Joint model



Coherence Feature

- Intuition: entities in the same column (row) tend to own the same type
- Element-wise variance for all entity vectors in the same column
  - $\boldsymbol{h}^{(coh)} = \frac{1}{|C|}\sum_j \boldsymbol{var}(\{\boldsymbol{e}_{ij}|(i,j \in P)\})$

科幻片	美国	钢铁侠
爱情片	意大利	邮差
警匪片	中国	线人
动画片	美国	驯龙高手
…	…	…
Movie Type	Movie Country	Movie Name



Training & Predication

- One positive & n negative tables
  – Randomly corrupted
  – Random number & positions
- Loss function
  – Hinge Loss
  – Pairwise Ranking Loss
- Iterative prediction
  – Local Search Descent

End-2-End Results

Approach	Micro Acc.	Macro Acc.
$TabEL_B$	0.512	0.507
$TabEL_W$	0.514	0.519
$TextEL$	0.472	0.458
Ours (Baidu Only)	0.576	0.573
Ours (Full, - pre-train)	0.606	0.591
Ours (Full, + pre-train)	**0.629**	**0.614**

$TabEL_B$, $TabEL_W$:

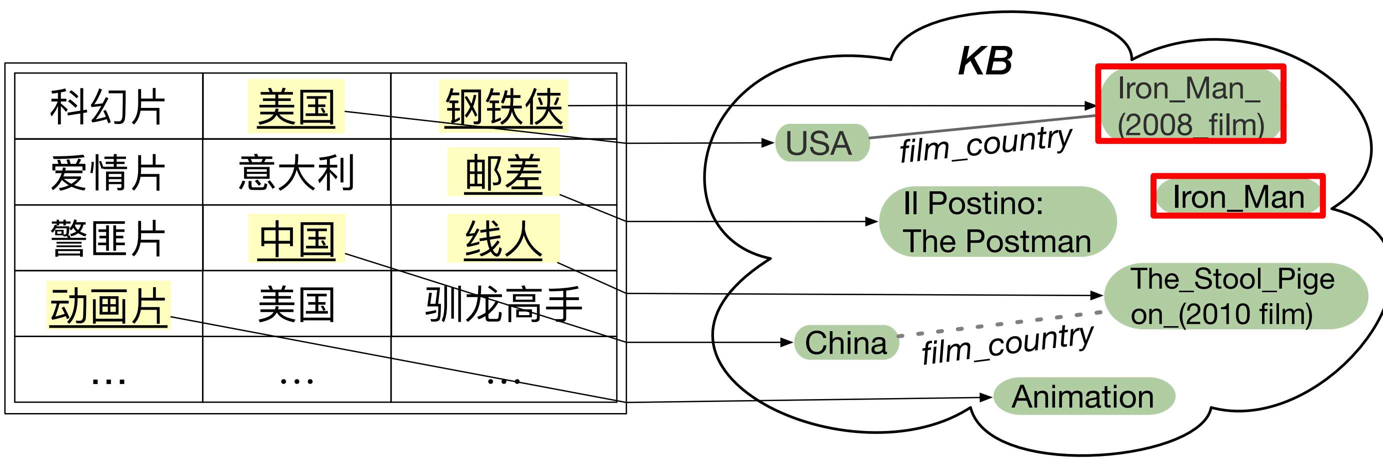
translation + mono-lingual table linking

TextEL:

transform table to text + cross-lingual text linking

Feature Variations

Feature Combination	Micro Acc.	Decrease in Acc. (%)
Mention Only	0.604	12.7
Context Only	0.576	16.7
Coherence Only	0.279	59.6
Mention + Context	0.652	5.78
Full	0.692	0.00



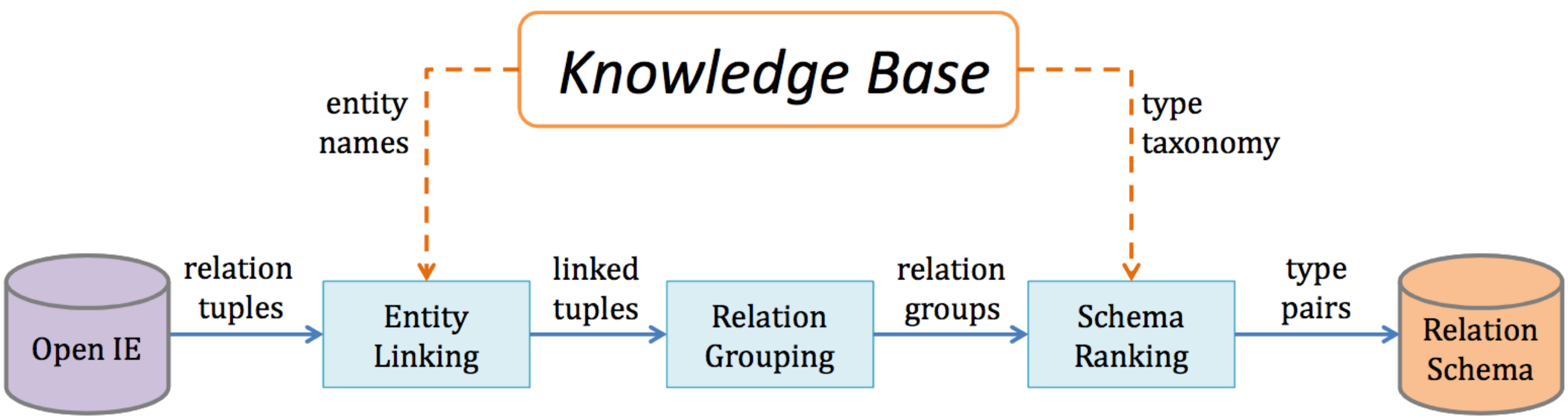
coherence works!

Type Inference for Relation Arguments

EMNLP 2015

A played in B
C plays in D
.....

⟨**film actor**, *play in*, **film**⟩
⟨**athlete**, *play in*, **sports league**⟩



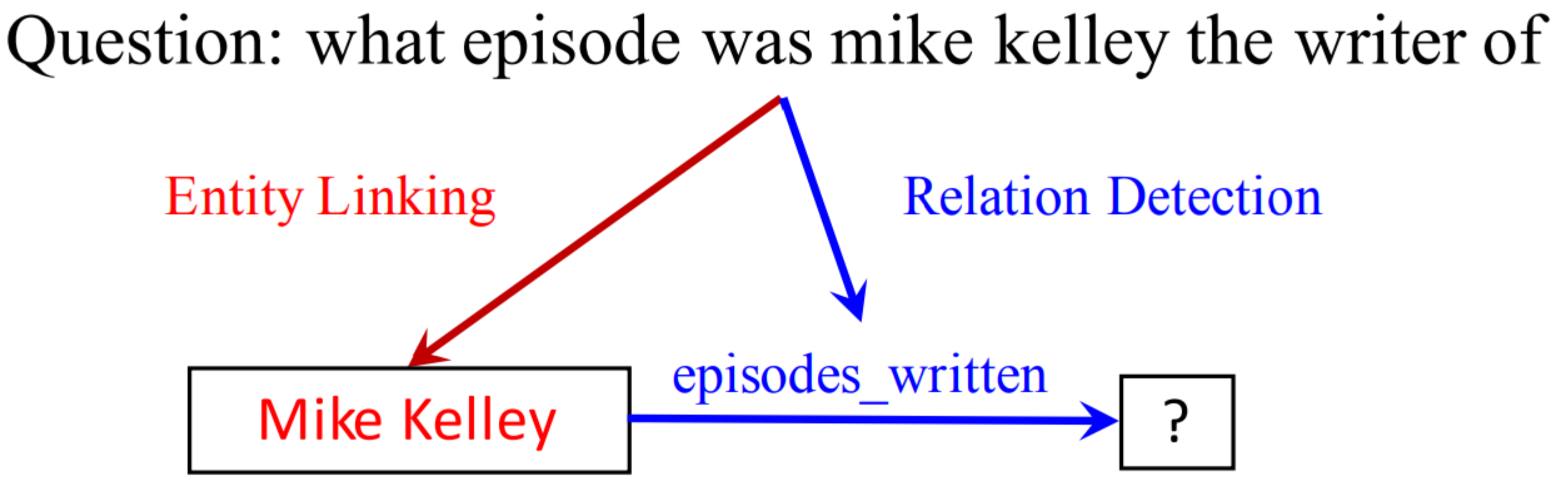
Outline

- Introduction to Knowledge Graphs
- Applications
  - Entity Linking
  - Factoid Question Answering
  - E-Commerce Application
  - Knowledge Based Recommendation

KBQA Introduction

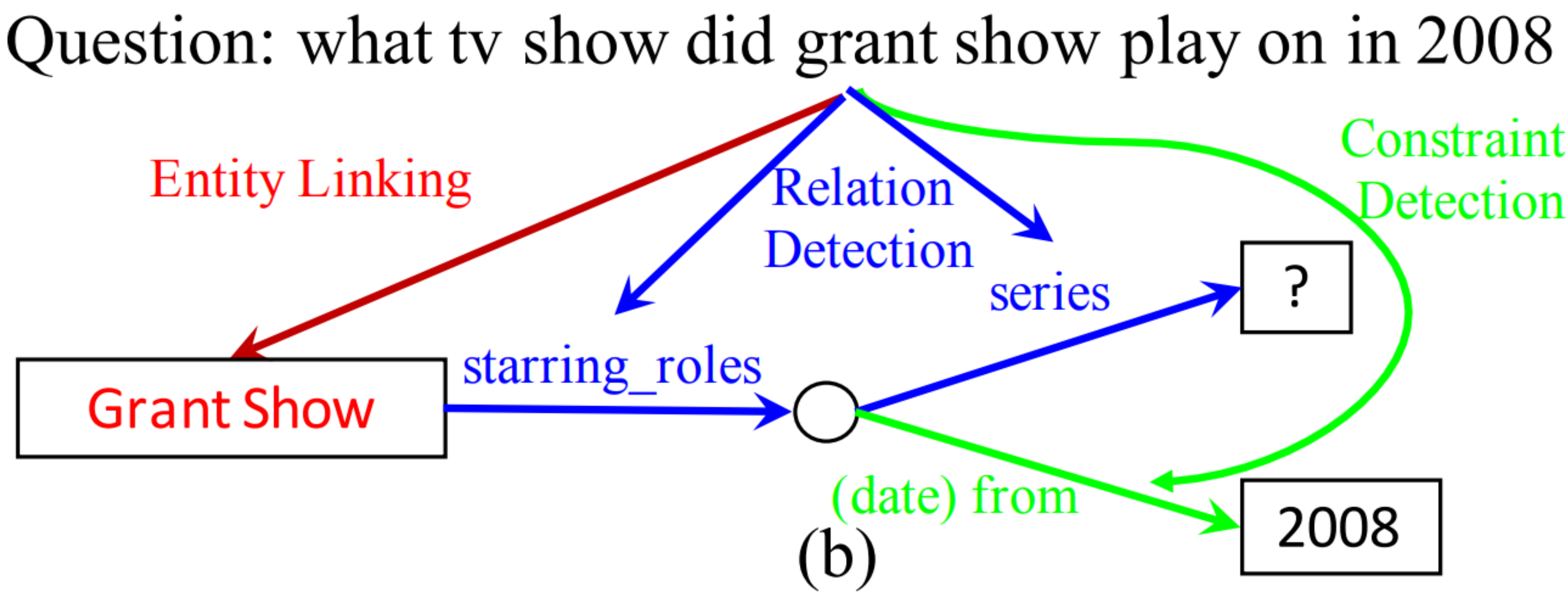
- Question Answering over Knowledge Bases
- Focus on *factoid questions*
  - The answer is a single entity in the KB
  - What/where/when, not how/why
- “where was obama born? ”
  - Answer: “Hawaii”
  - Structure:

Semantic Parsing



Query Structure

Semantic Parsing

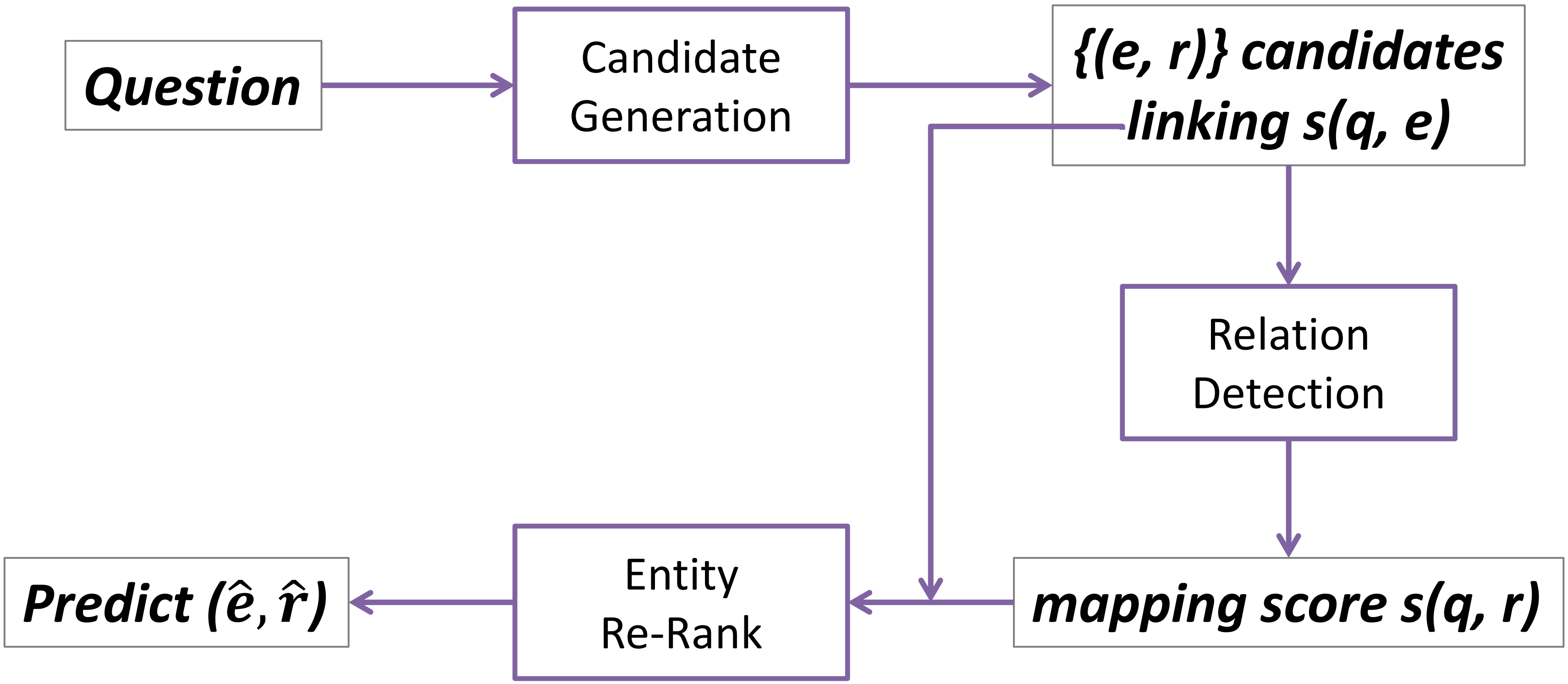


Query Structure

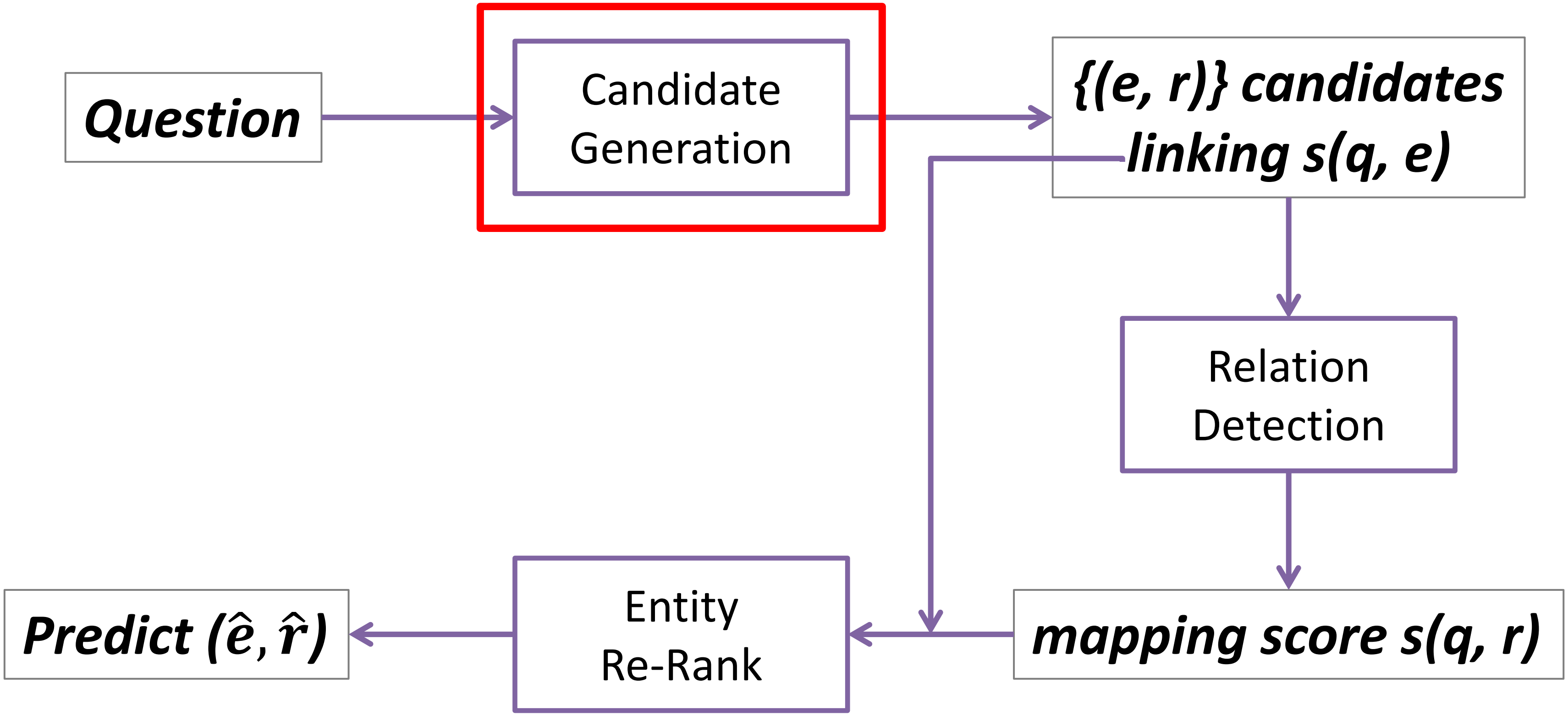
KBQA Datasets

- Free917 [Cai & Yates, 2013] ***ACL 2013***
  - 917 <question, lambda calculus forms>
- WebQuestions [Berant et al., 2013] ***EMNLP 2013***
  - 5810 QA pairs, Google Search API + Human Efforts
- SimpleQuestions [Bordes et al., 2015] ***arXiv***
  - 100K+ <question, focus entity, relation> triples
  - Simple semantics (one relation to answer)
- ComplexQuestions [Bao et al., 2016] ***COLING 2016***
  - 2100 QA pairs
  - Answer is derived from multiple constraints

KBQA Framework

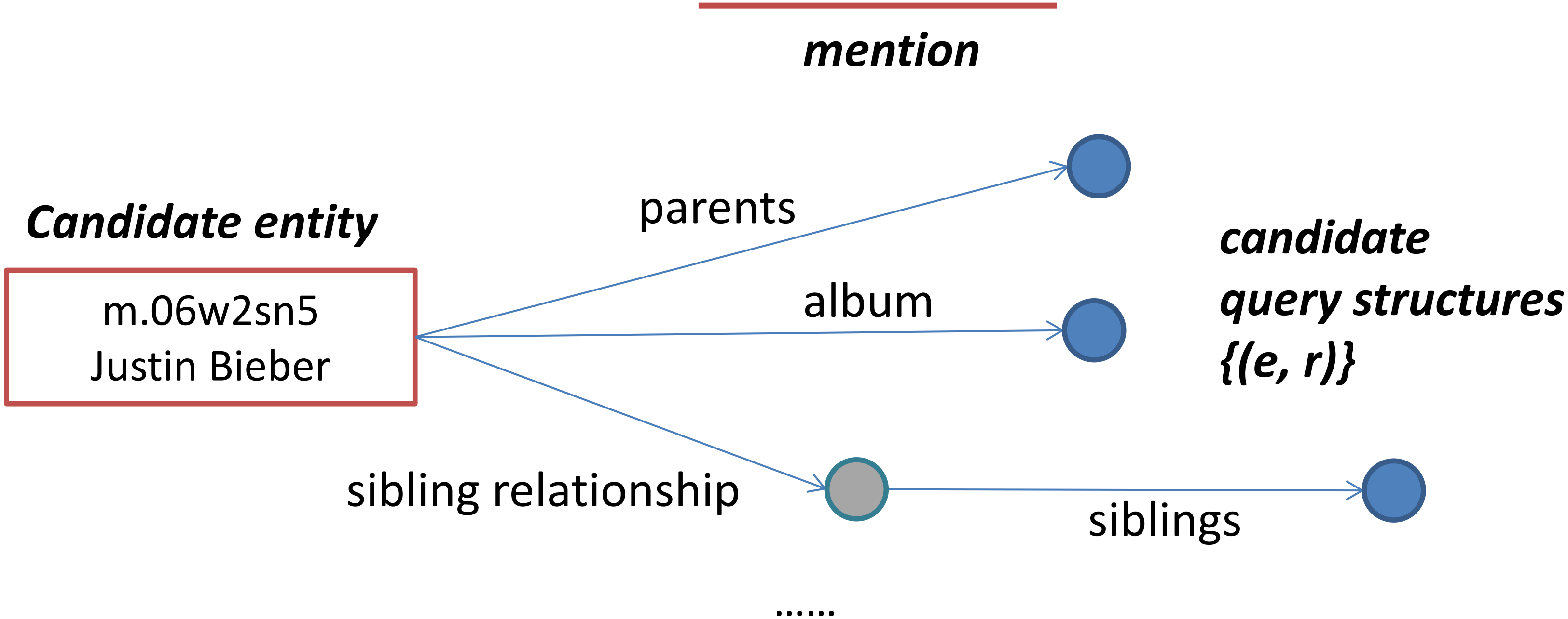


KBQA Framework



Candidate Entity Generation

" what is the name of justin bieber brother ? "



Candidate Entity Generation

- “ where is rome italy located on a map ? ”
  - Generate candidates
  - Calculate entity linking score

m.06c62
Rome

m.06k176
Rome (TV series)

......

m.03rjj
Italy

m.0218mm
Italian Peninsula

......

......

Candidate Entity Generation

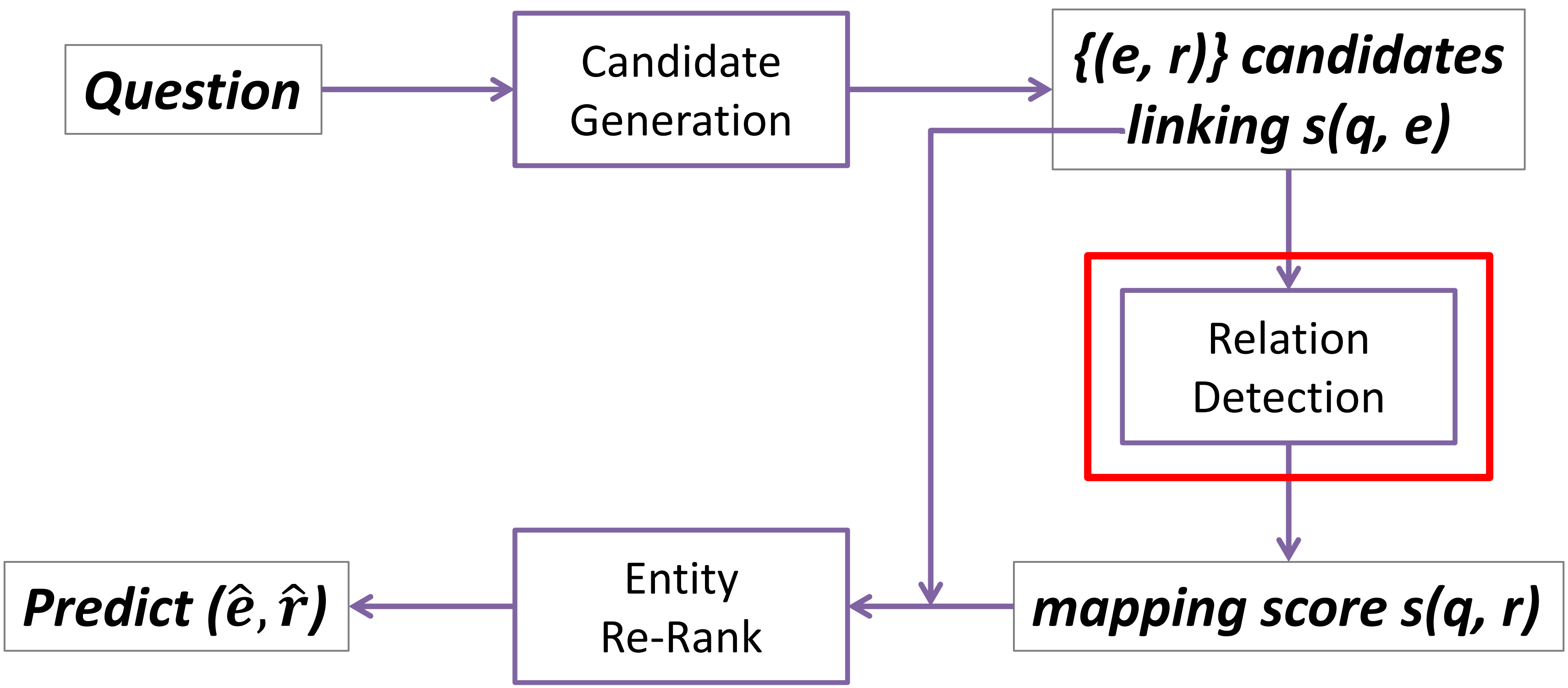
- External Module
  - Lexical Overlap Score [Yin et al., 2016] ***COLING16***
  - SMART [Yang and Chang, 2015] ***ACL15***
  - BiLSTM-CRF [Dai et al., 2016] ***ACL16*** [Yin et al., 2016] ***COLING16***

- Internal Module
  - Learn the score through entity linking model

Candidate Relation Generation

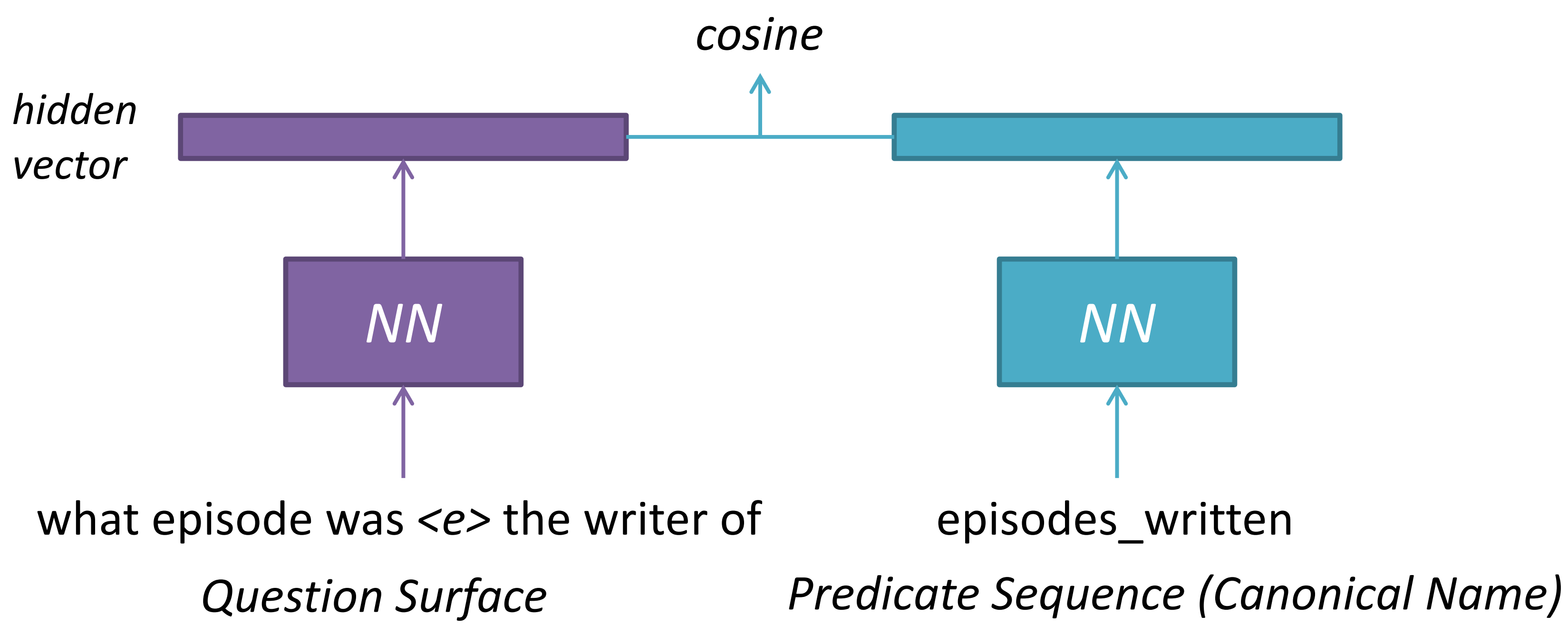
- Main Path
  - from the focus entity, search all paths in 1~3 hops
  - pruning tricks (answer type)
- Constraints
  - entity type
  - time
  - ordinal

KBQA Framework



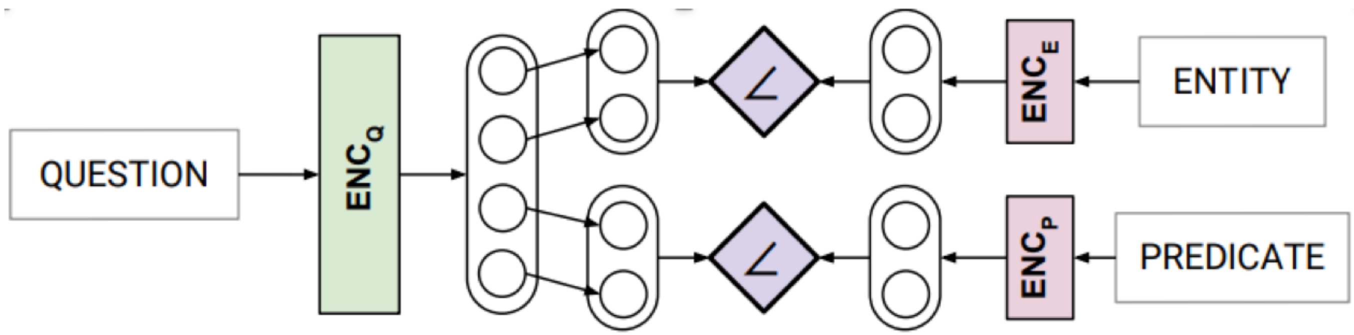
Relation Detection

- Intuition: Calculate the similarity between the question and candidate predicate.



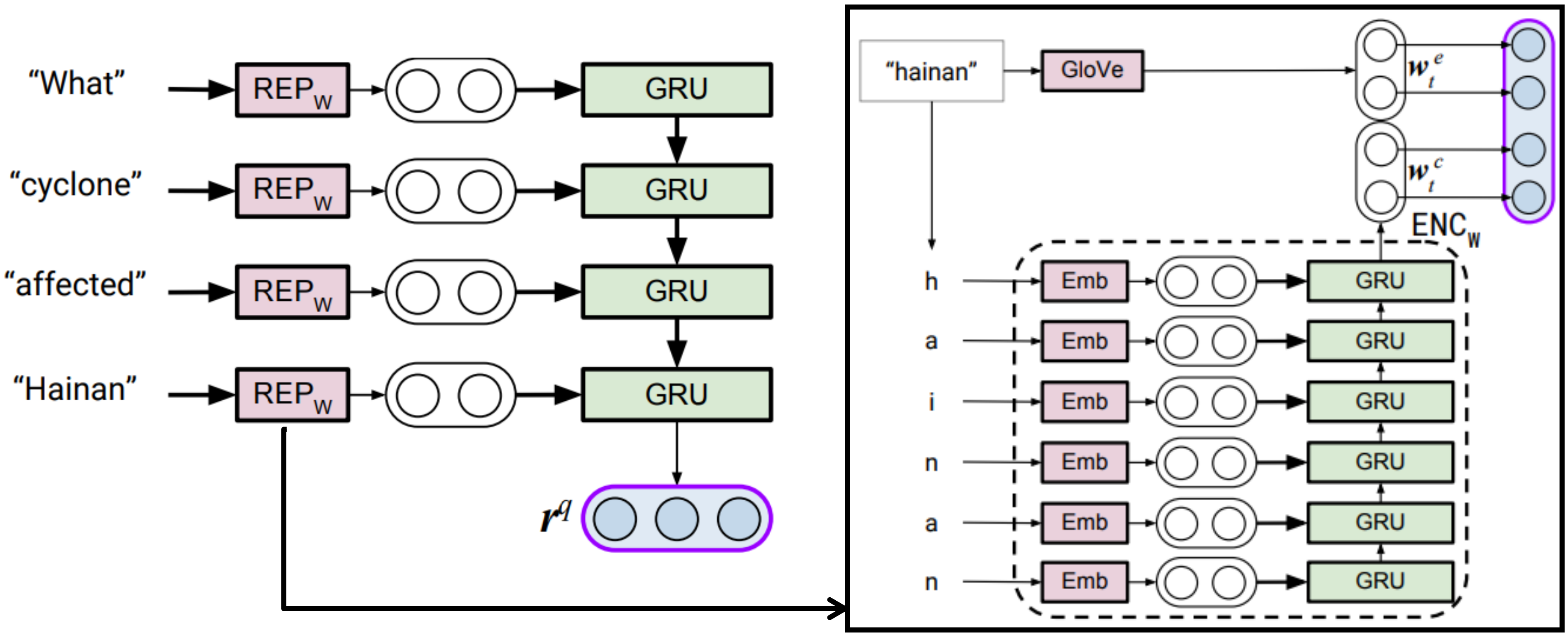
Relation Detection: RNN

- Lukovnikov et al. [2017] ***WWW 2017***



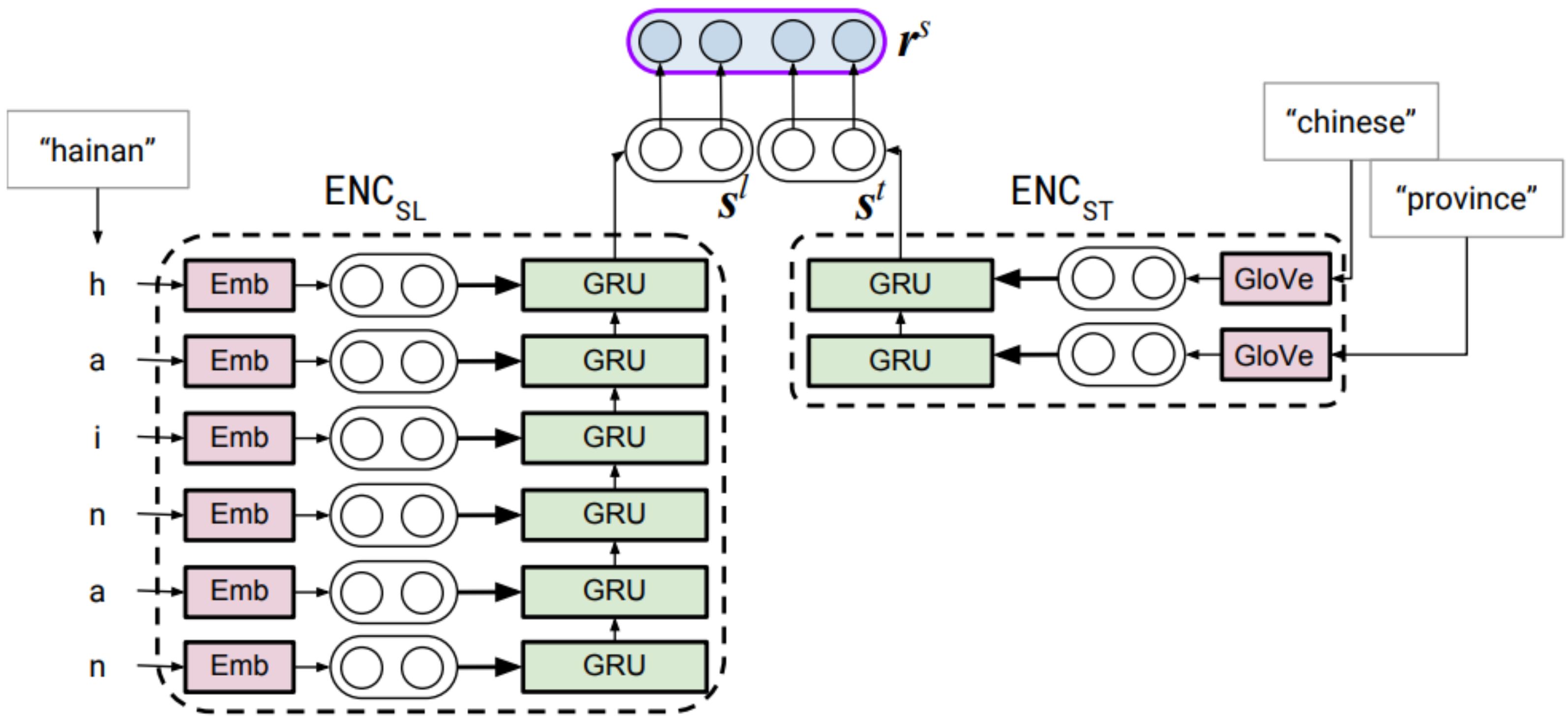
Question encoding

- word and char level GRU

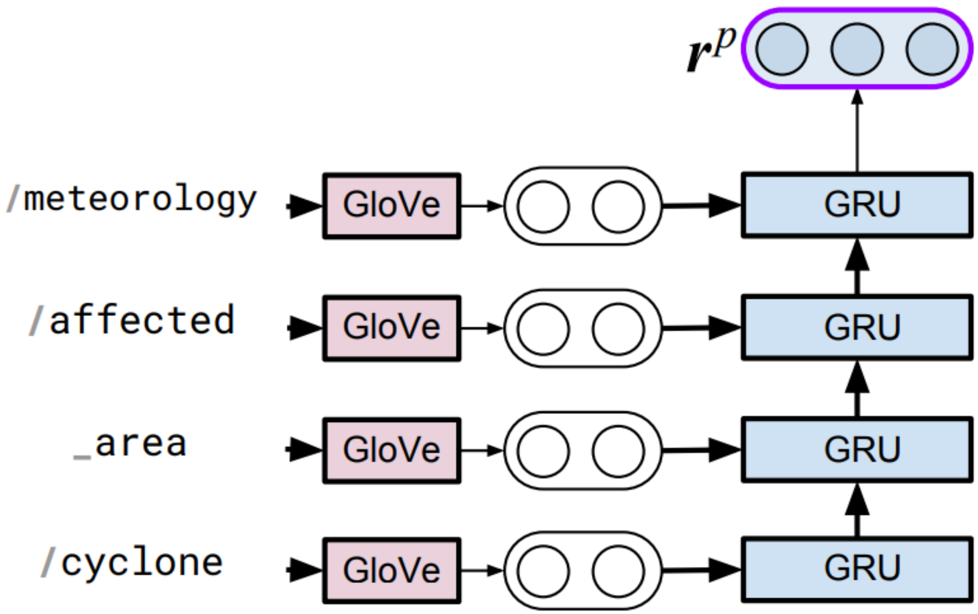


Entity Encoding

- Char level GRU + entity type information

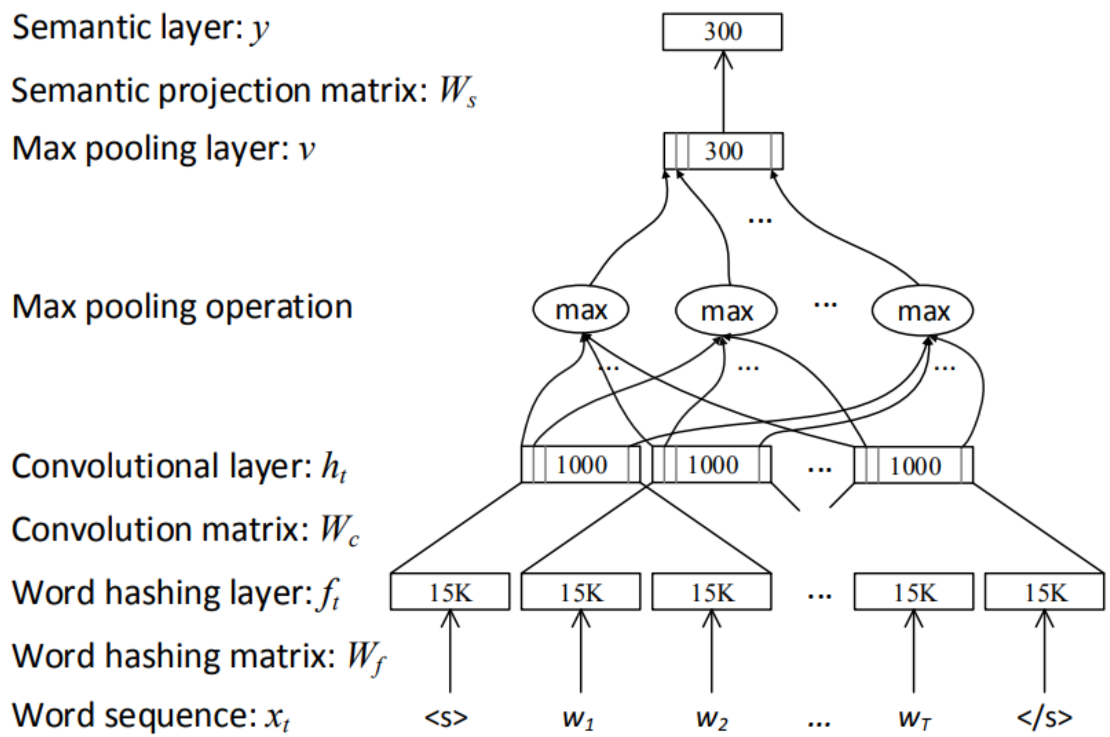


Relation Encoding

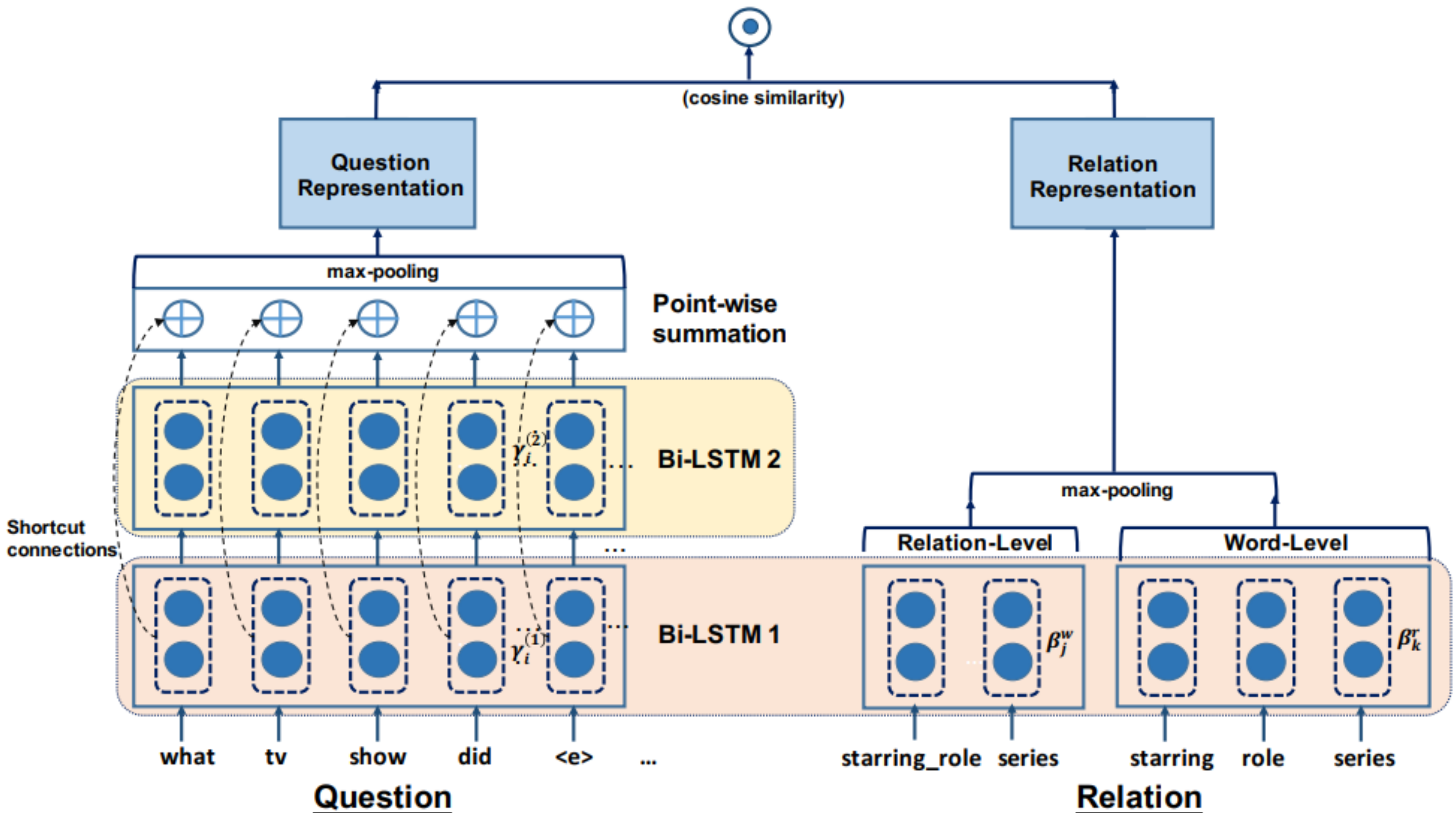


Relation Detection: CNN

- CNN Model [Yih et al., 2015] ***ACL15*** [Bao et al., 2016]***COLING16***
  - Siamese
  - Word Hash



Relation Detection : Residual RNN



[Yu et al., 2017] ***ACL 2017***

Relation Detection: AR-SMCNN

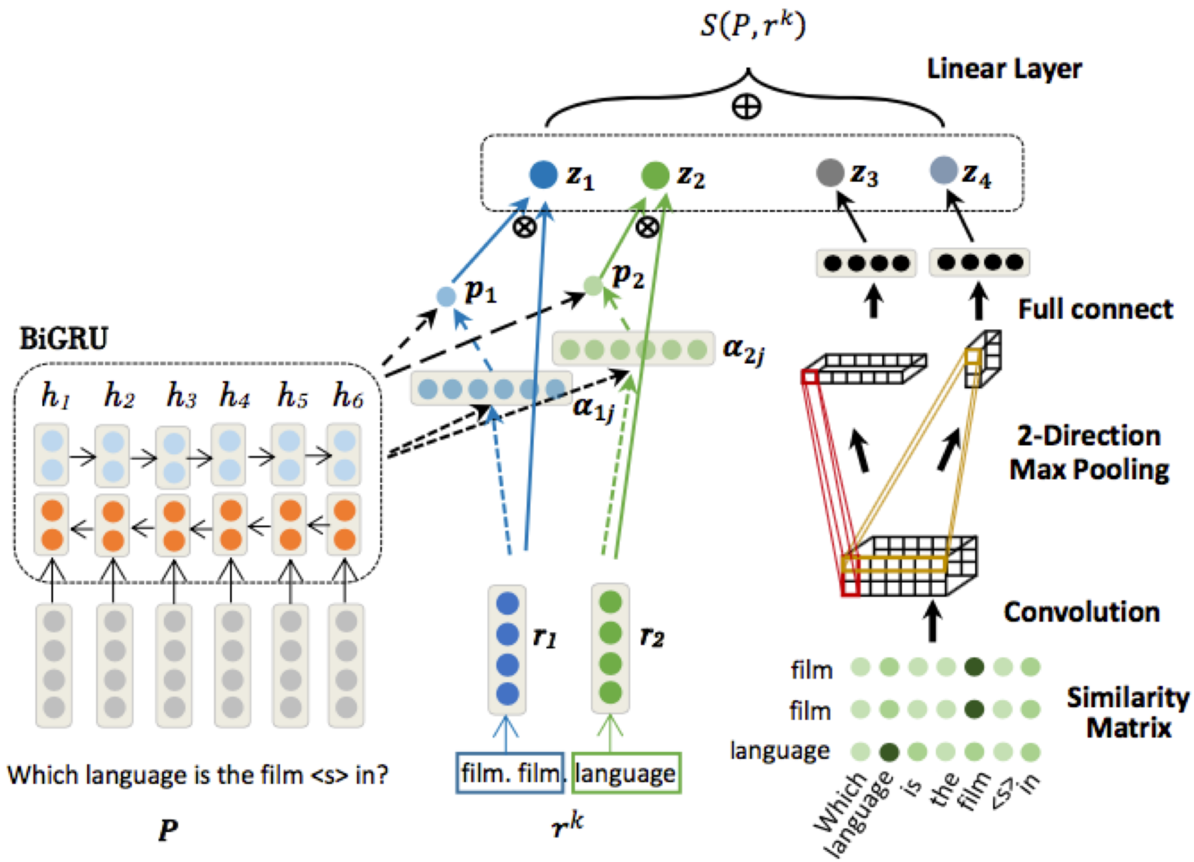
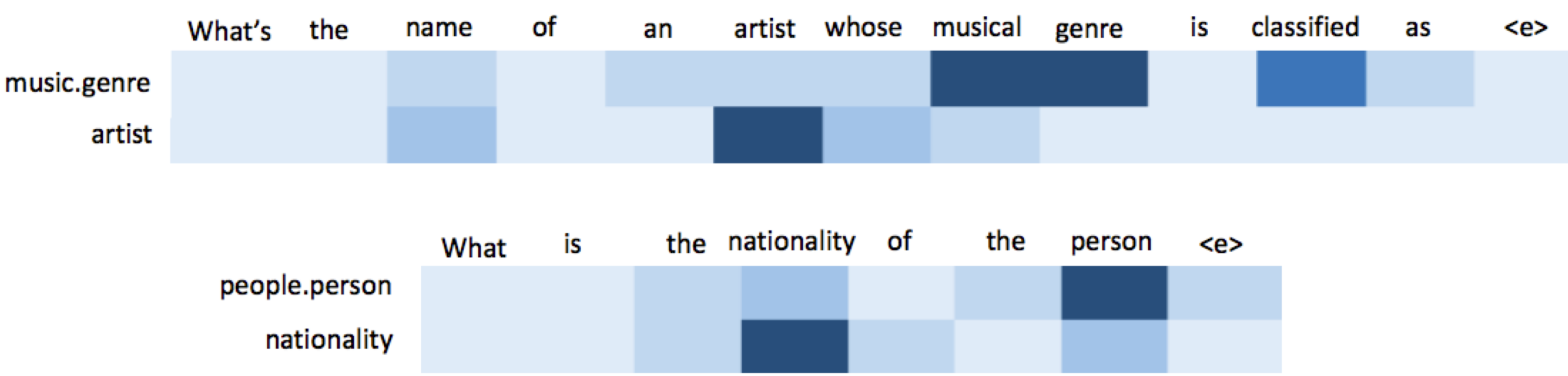


Fig. 3. The proposed AR-SMCNN model. The left part is attentive BiGRU, and the right part is CNN on similarity matrix. The features z_1, z_2, z_3, z_4 is concatenated together and pass through a linear layer to get final score $S(P, r^k)$.

[Qu et al., 2018] ***arXiv***

Relation Detection: AR-SMCNN



[Qu et al., 2018] ***arXiv***

Model Training

- Hinge Loss (ground truth is directly provided)
  - Lukovnikov et al. [2017] ***WWW 2017***

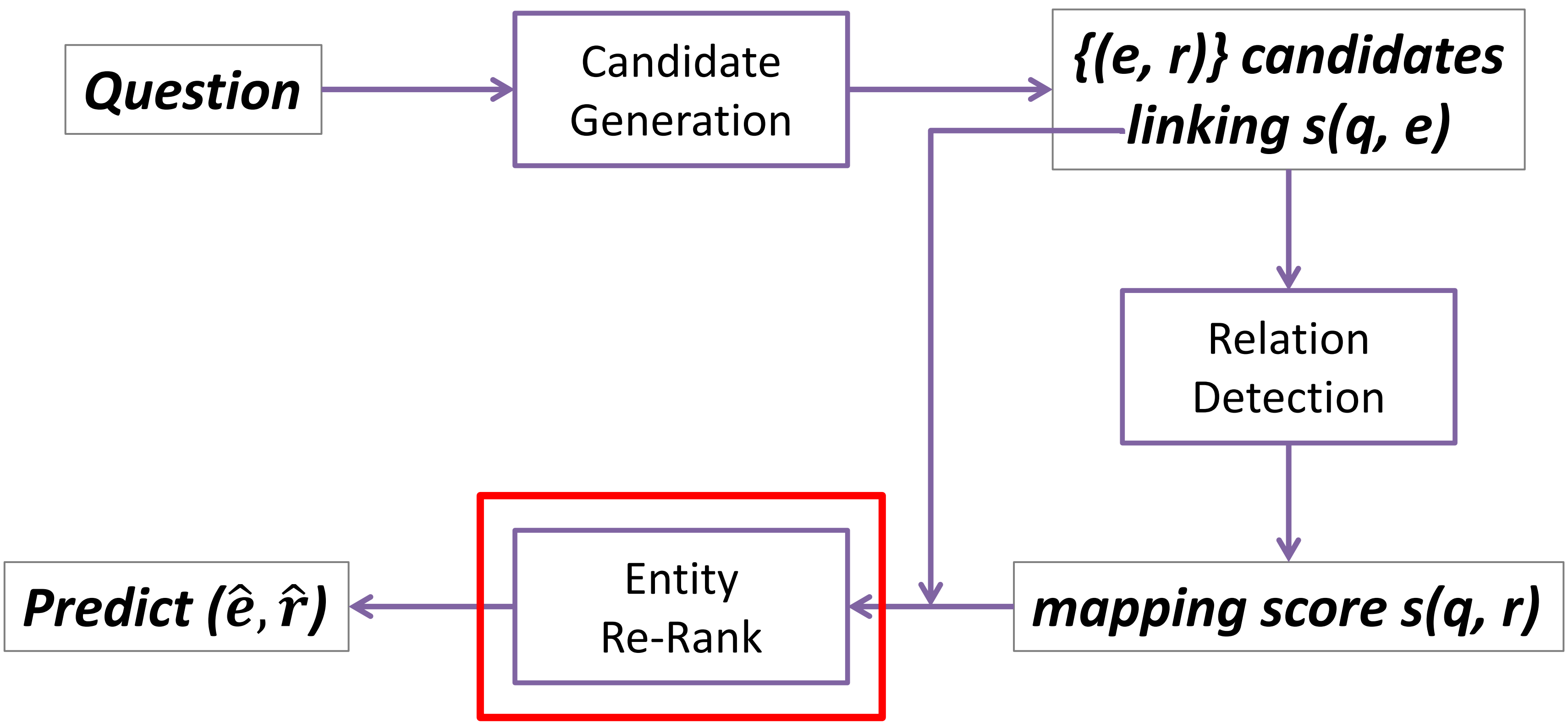
$$-\sum_{(q,s^+,p^+)\in\mathcal{D}} \Big(\max\left(0, S_s(q,s^-) - S_s(q,s^+) + \gamma\right) + \max\left(0, S_p(q,p^-) - S_p(q,p^+) + \gamma\right) \Big) \ .$$

  - Yu et al. [2017] ***ACL 2017***

$$l_{\text{rel}} = \max\{0, \gamma - s_{\text{rel}}(\mathbf{r}^+;\mathbf{q}) + s_{\text{rel}}(\mathbf{r}^-;\mathbf{q})\}$$

- Rank Loss (only answer provided)
  - generate a set of R with different F1 score -> forms a rank

KBQA Framework



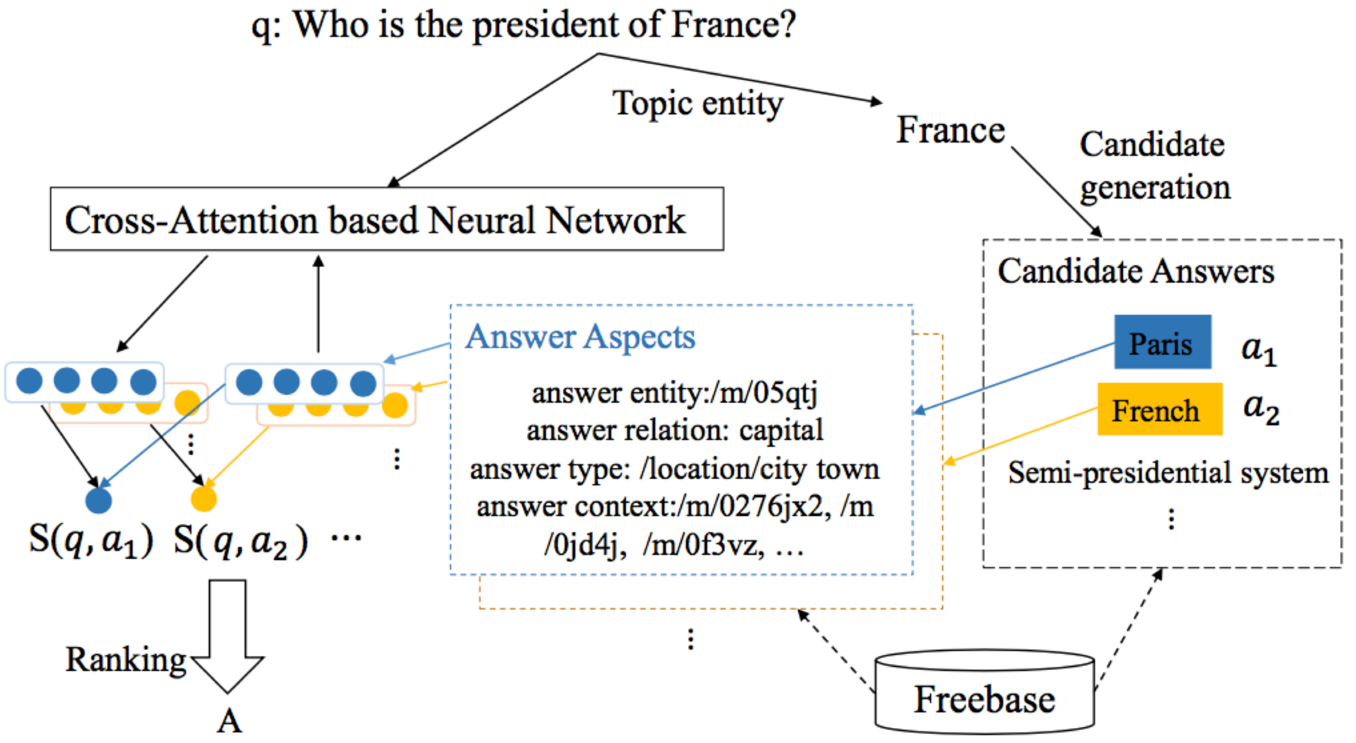
Entity Re-Rank in Prediction

- Re-rank:
  - Relation detection provides a feedback
- Re-rank in Yu et al. [2017] ***ACL 2017***

$$
\begin{aligned}
s_{\text{rerank}}(e;q) =& \alpha \cdot s_{\text{linker}}(e;q) \\
&+ (1-\alpha) \cdot \max_{r \in R_q^l \cap R_e} s_{\text{rel}}(r;q)
\end{aligned}
$$

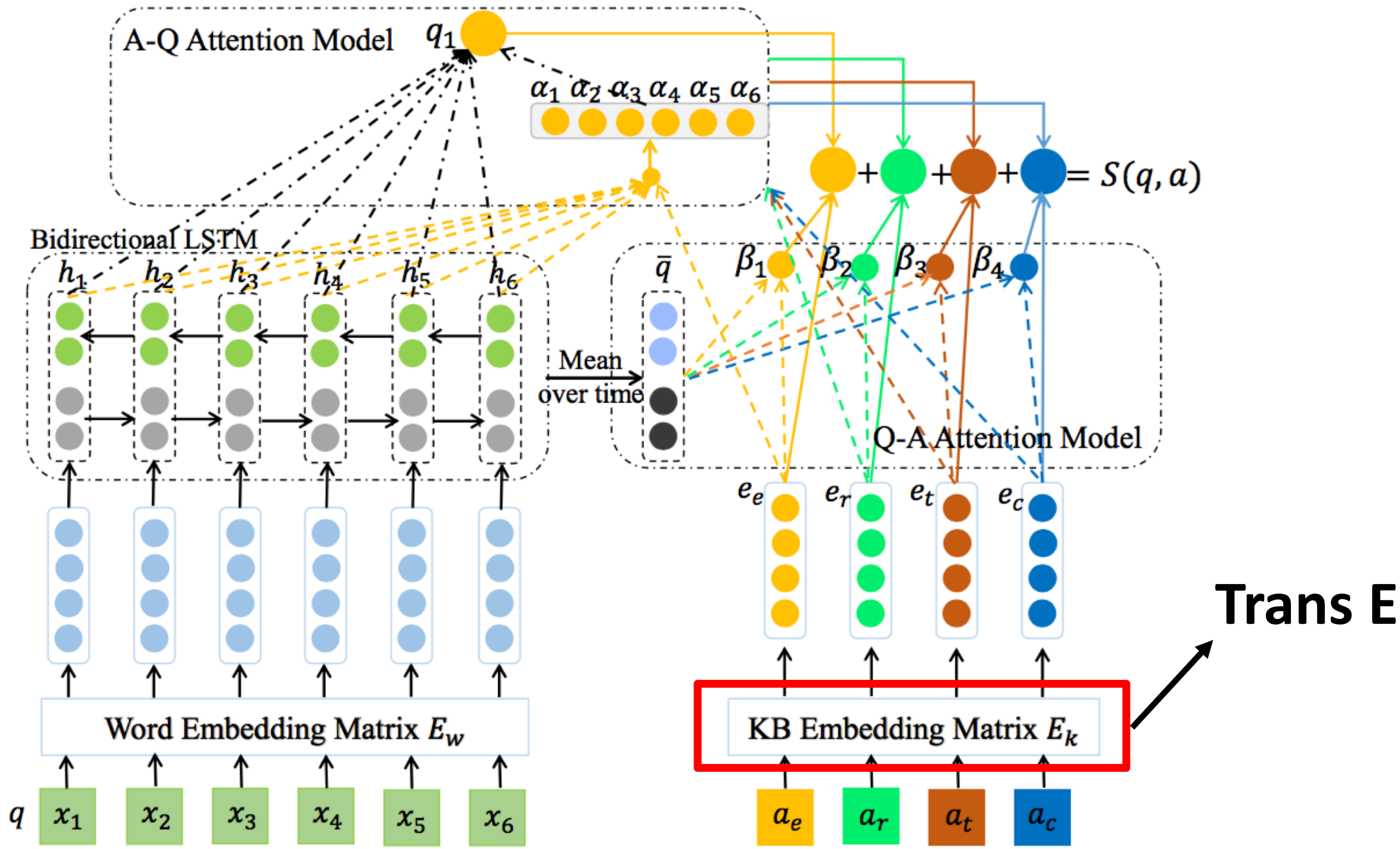
Latest IE Based System

- [Hao et al., 2017] ***ACL 2017***



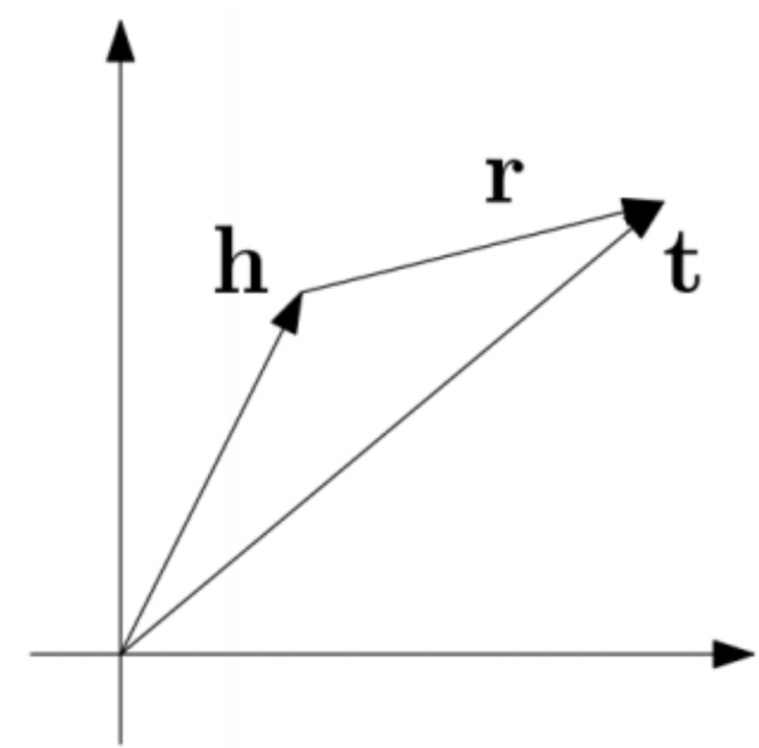
Latest IE Based System

- [Hao et al.. 2017] ***ACL 2017***



KG Embedding

TransE



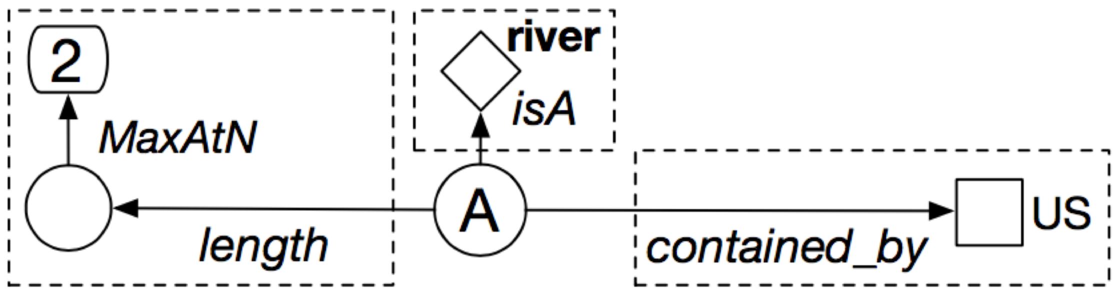
Learning objective: **h + r = t**

*What if Questions r **Complex**?*

q: What is the second longest river in China?

It is easy to encode a question(sentence)

How to find & encode a complex query structure?

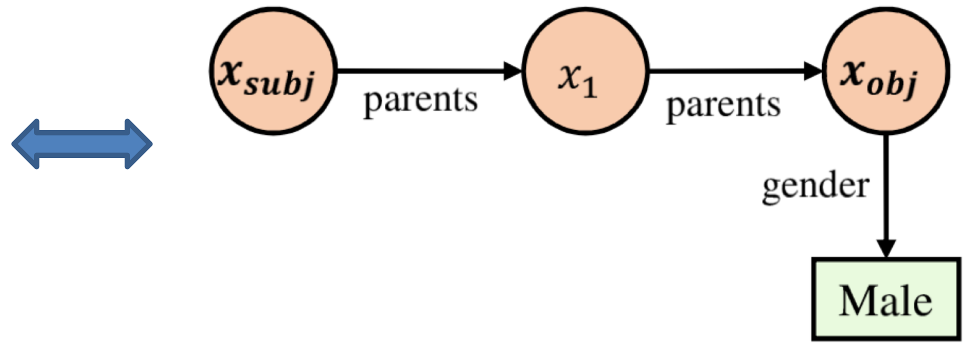


What is the second longest river in the United States?

Natural Language Relation Paraphrasing

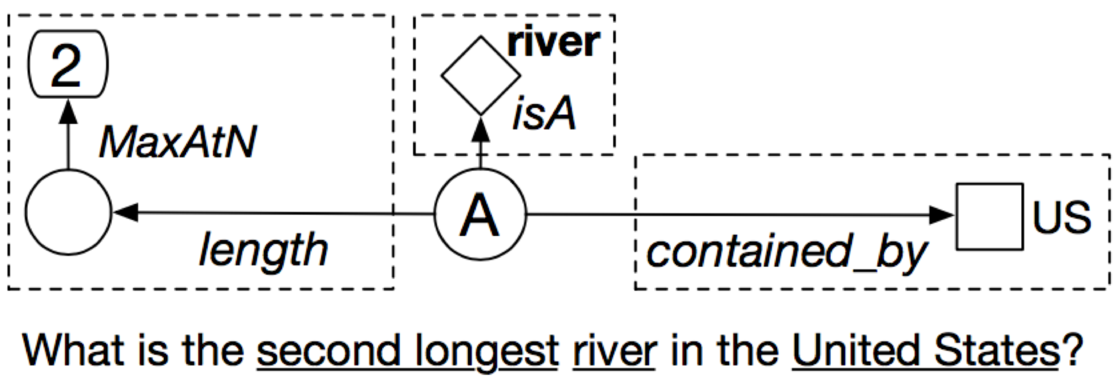
IJCAI 2017

"has grandfather"



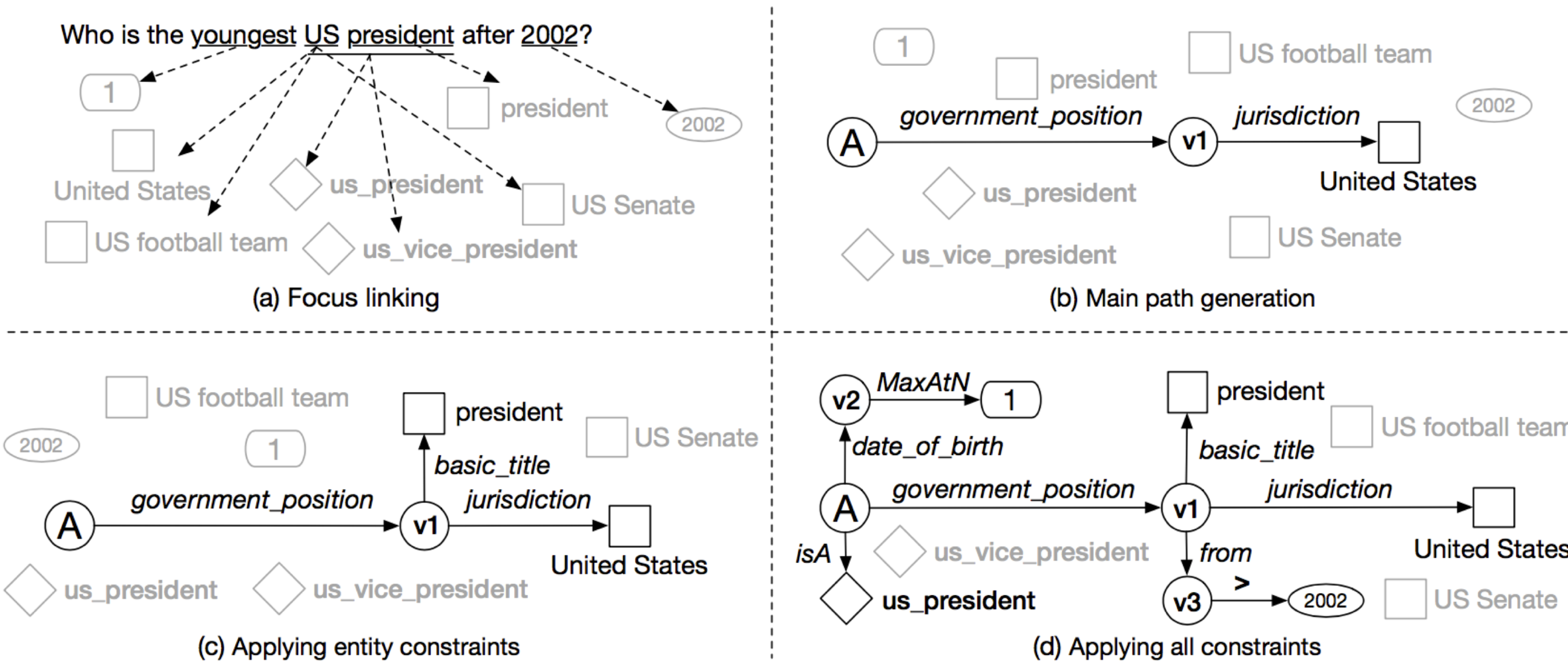
KBQA via Encoding of Complex Query Graphs

EMNLP 2018



Query Graph Generation

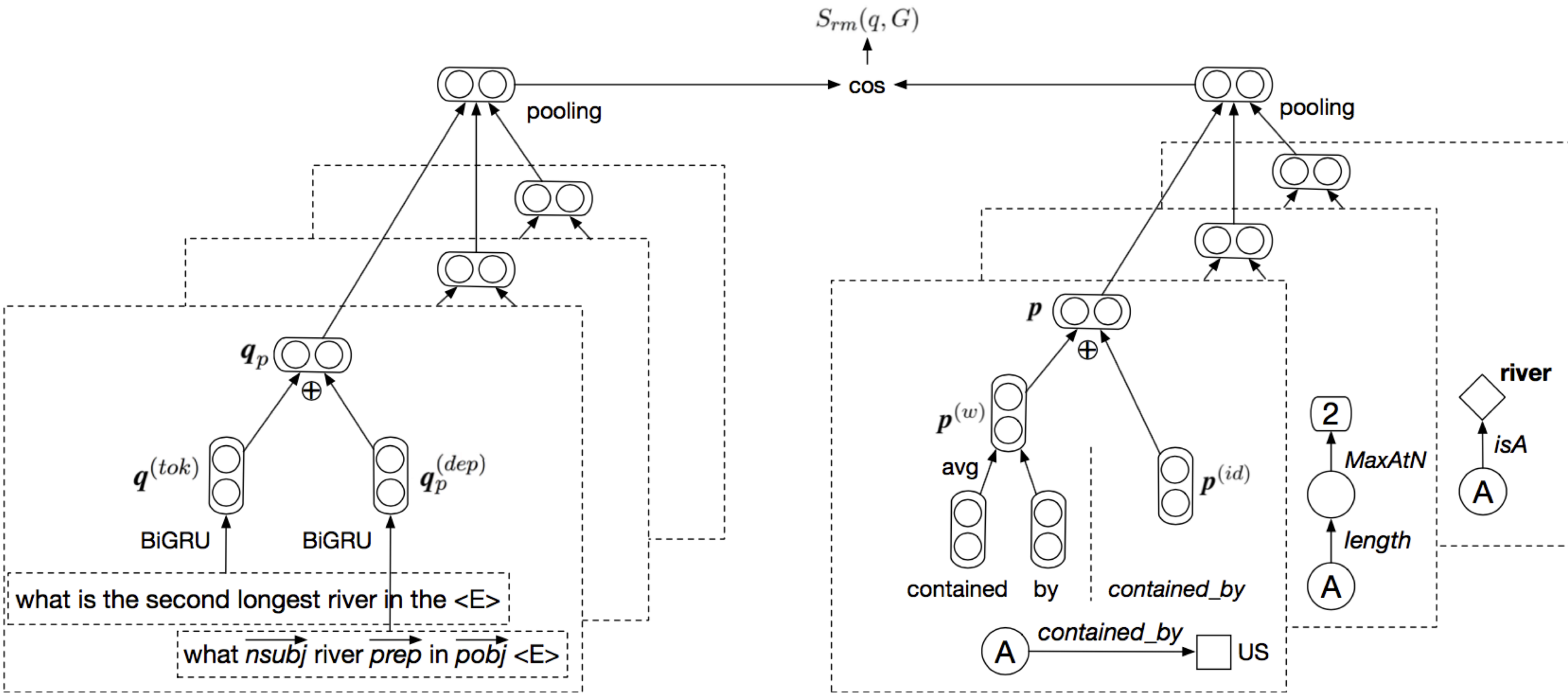
- Entity / Type / Time / Ordinal



Query Graph Generation

- Focus Entity Linking
  - SMART, uni-bi-tri-gram, top10 <m, e> pairs
  - time: year regex, ordinal: superlative word list ("largest", "highest"... 20+-)
- Main Path Generation
  - connect answer and focus entity within 1 or 2 hops in FB
- Entity Constraints
  - DFS
- Type Constrains
  - Implicit types, pre-calculate type similarity in FB
- Time and Ordinal
  - 2-hops, second is virtual (before, after, in/ maxAtN, minAtN)

Semantic Matching



Entity Linking +

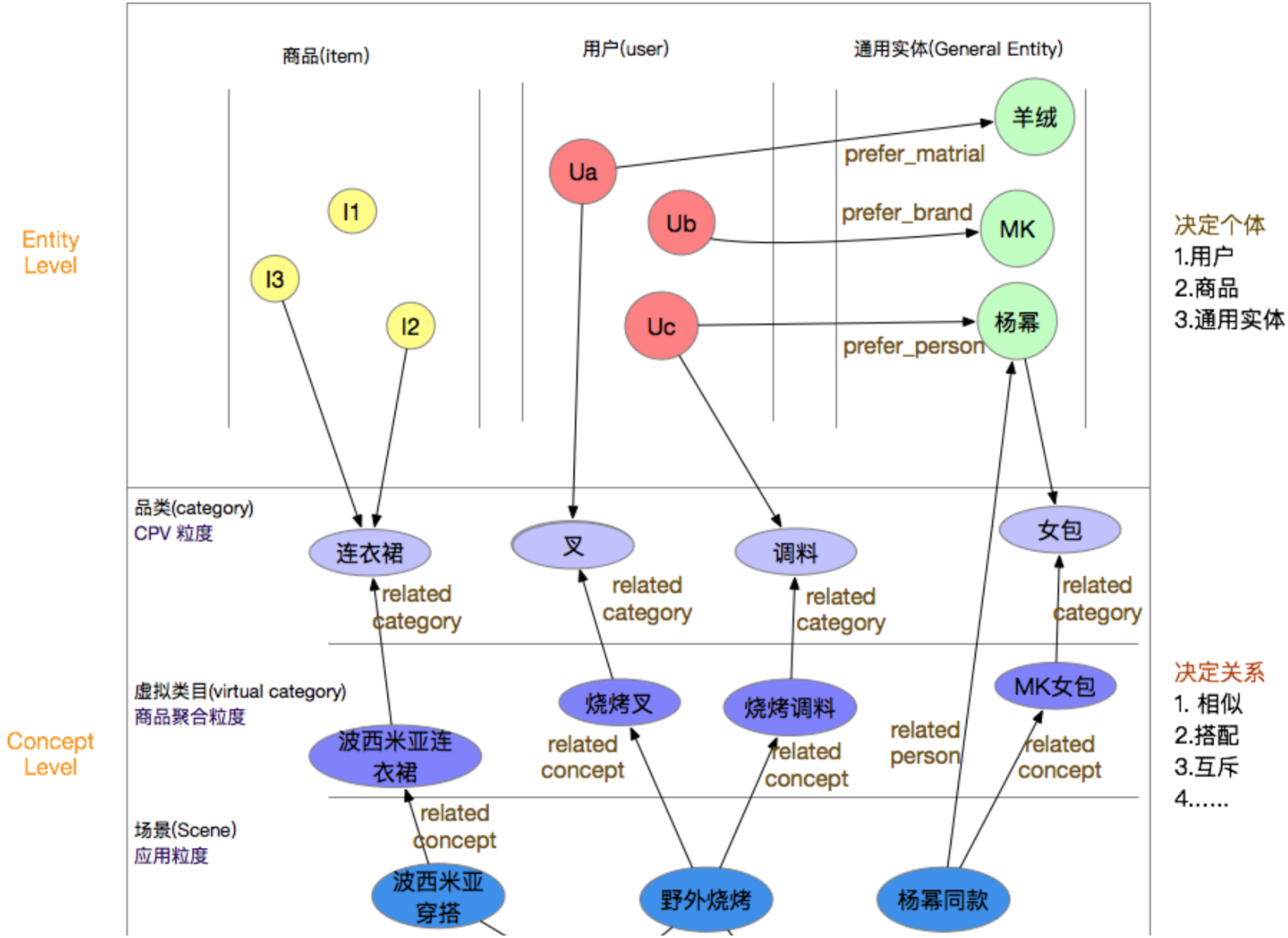
- External lexicon: Wikipedia
  - article titles
  - anchor text
  - redirects
  - disambiguate pages
- Generate features for each <m, e> pair
  - popularity
  - jaccard similarity
  - ...
- LR model to fit SMART scores
- Enrich SMART at entity linking step in QA

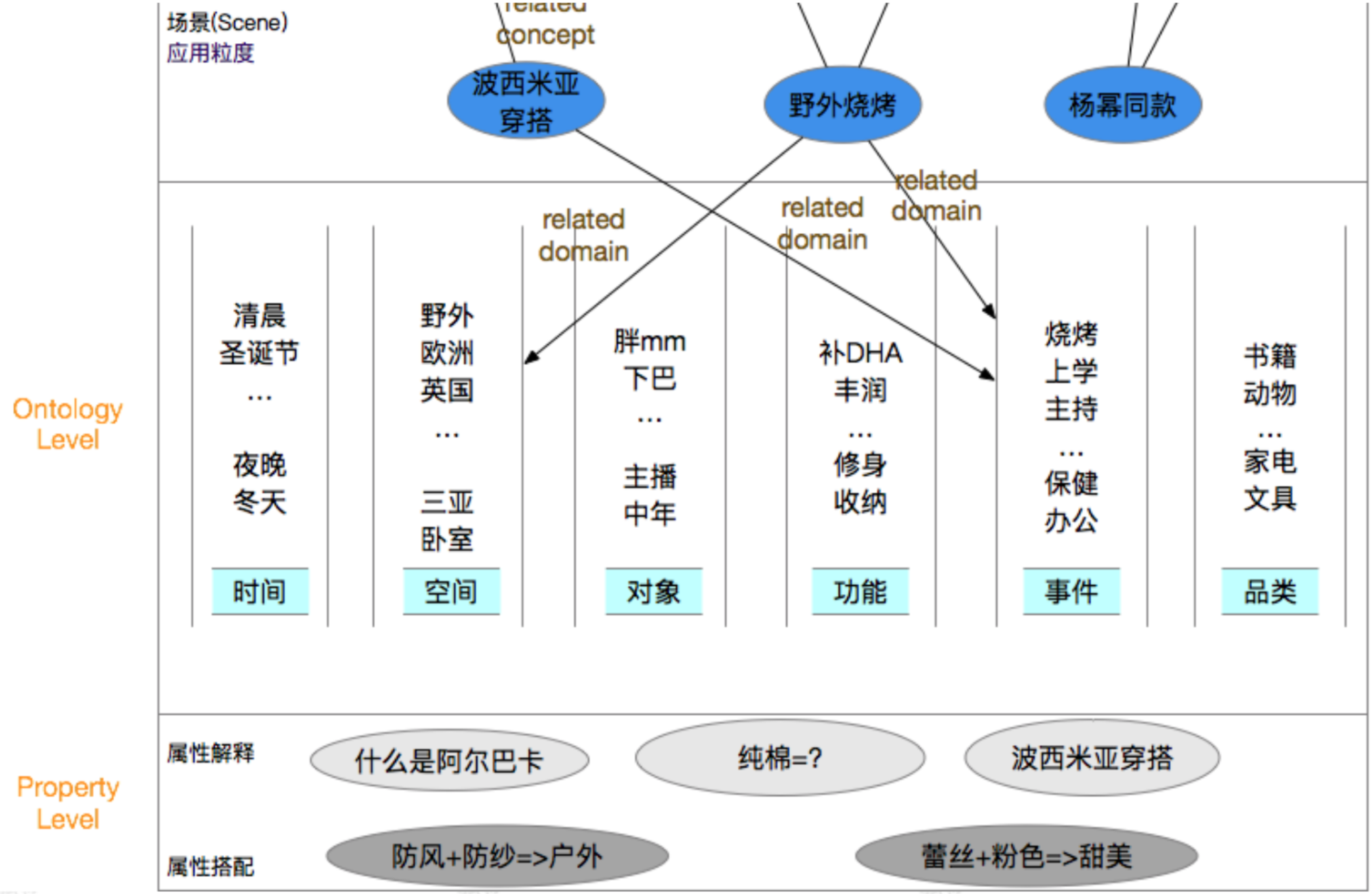
End2end Result

Method	CompQ	WebQ
Dong et al. (2015)	-	40.8
Yao (2015)	-	44.3
Bast and Haussmann (2015)	-	49.4
Berant and Liang (2015)	-	49.7
Yih et al. (2015)	36.9	52.5
Reddy et al. (2016)	-	50.3
Xu et al. (2016) (w/o text)	-	47.0
Bao et al. (2016)	40.9	52.4
Jain (2016)	-	**55.6**
Abujabal et al. (2017)	-	51.0
Cui et al. (2017)	-	34.0
Hu et al. (2018)	-	49.6
Talmor and Berant (2018)	39.7	-
Ours (w/o linking enrich)	42.0	52.0
Ours (w/ linking enrich)	**42.8**	52.7

Outline

- Introduction to Knowledge Graphs
- Applications
  - Entity Linking
  - Factoid Question Answering
  - E-Commerce Application
  - Knowledge Based Recommendation

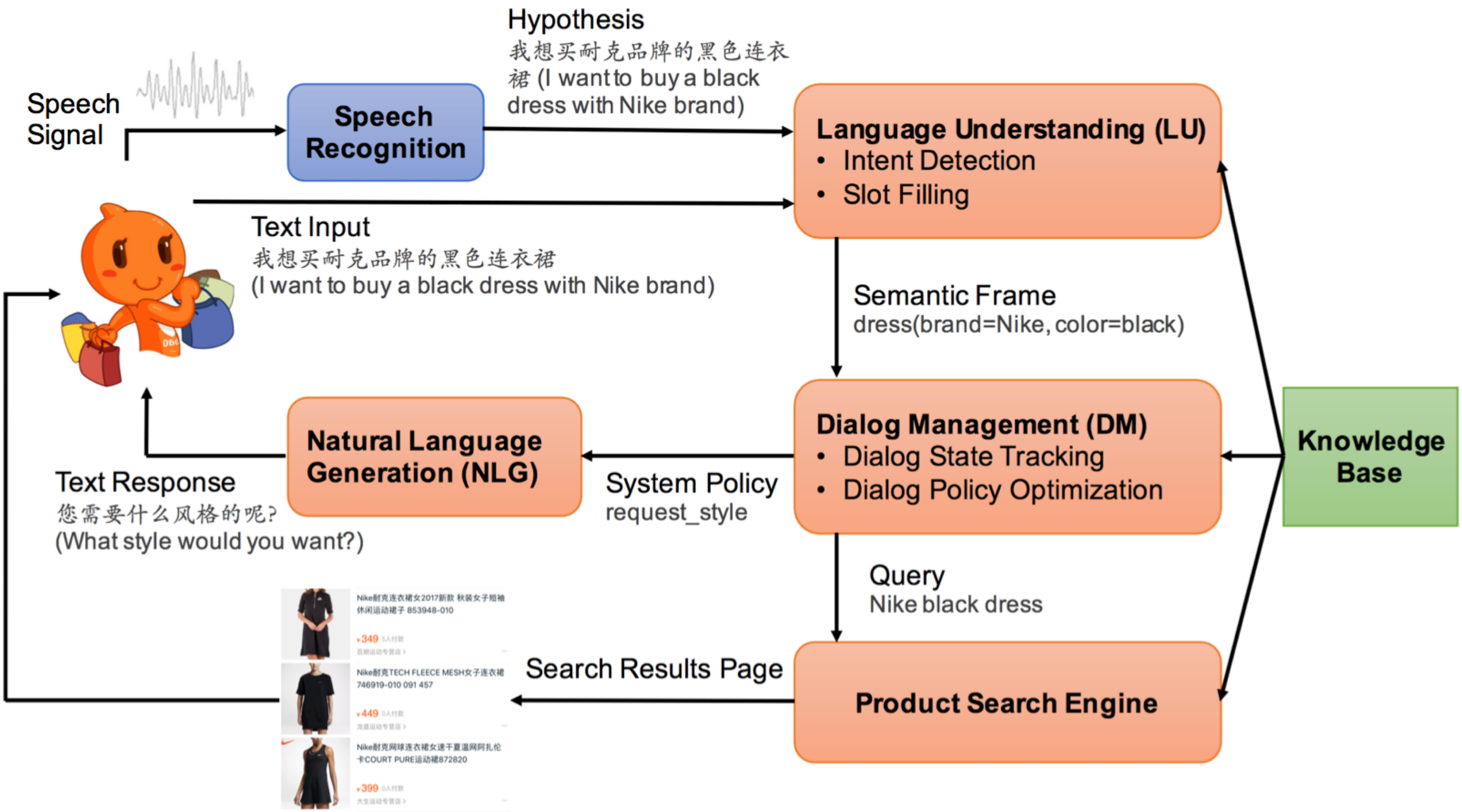

商品(item)
用户(user)
通用实体(General Entity)
Entity Level
I1
I3
I2
Ua
Ub
Uc
羊绒
MK
杨幂
prefer_matrial
prefer_brand
prefer_person
决定个体
1.用户
2.商品
3.通用实体
品类(category)
CPV 粒度
连衣裙
叉
调料
女包
related category
related category
related category
related category
虚拟类目(virtual category)
商品聚合粒度
波西米亚连衣裙
烧烤叉
烧烤调料
MK女包
related concept
related concept
related person
related concept
决定关系
1. 相似
2.搭配
3.互斥
4.……
Concept Level
场景(Scene)
应用粒度
related concept
波西米亚穿搭
野外烧烤
杨幂同款



决定层级

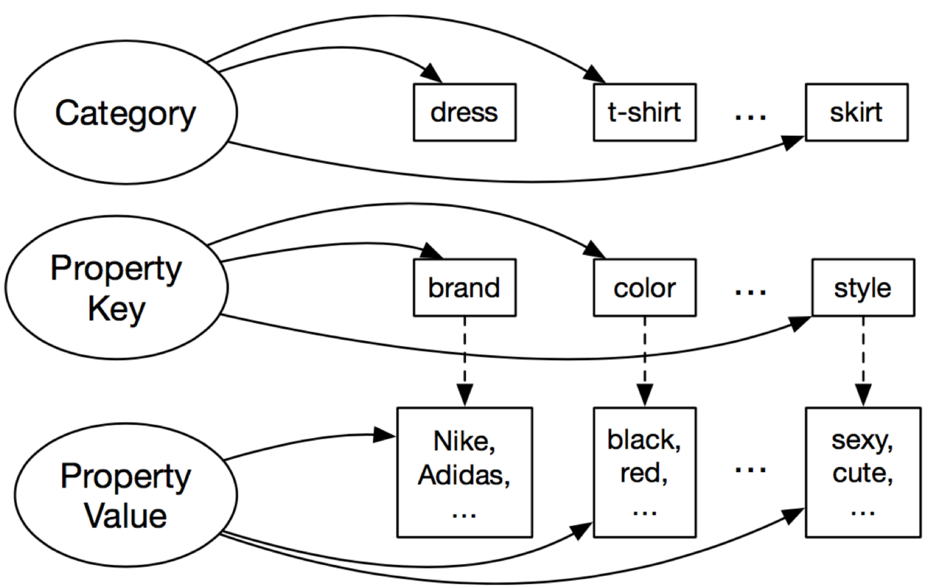
1. 上下位
2. 发散性
3. ……

Task-oriented E-Commerce dialog system



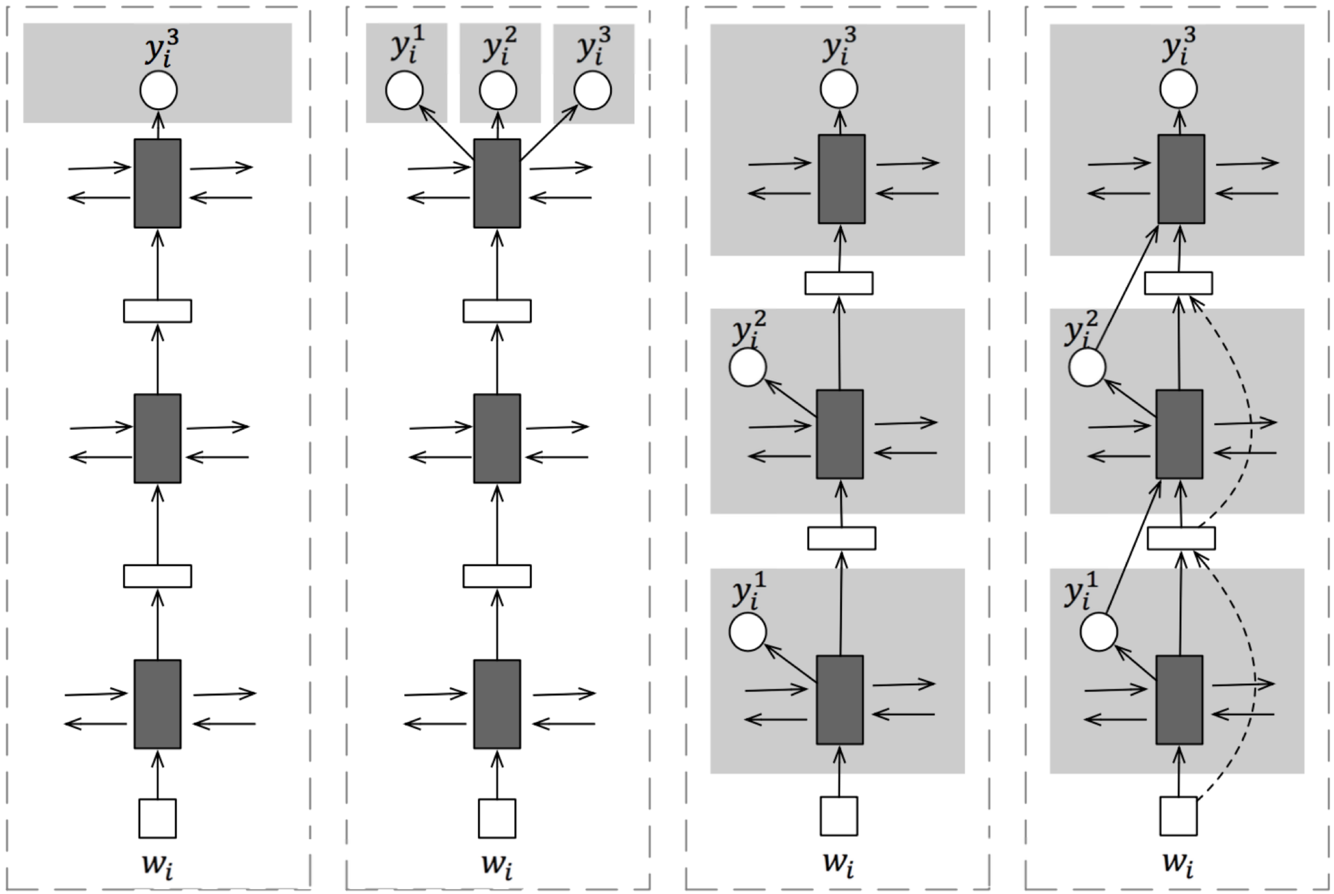
Slot Filling

Utterance	我	想	买	耐	克	品	牌	的	黑	色	连	衣	裙
	I	want	buy	Nike		brand		\	black		dress		
Slot Label	**O**	**O**	**O**	**B-Brand**	**I-Brand**	**B-PK**	**I-PK**	**O**	**B-Color**	**I-Color**	**B-CG**	**I-CG**	**I-CG**
Named Entity Label	**O**	**O**	**O**	**B-PV**	**I-PV**	**B-PK**	**I-PK**	**O**	**B-PV**	**I-PV**	**B-CG**	**I-CG**	**I-CG**
Segment Label	**O**	**O**	**O**	**B**	**I**	**B**	**I**	**O**	**B**	**I**	**B**	**I**	**I**



Category-Property-Value (CPV) as a KG

Deep Cascade Multi-task Slot Filling



(a) Basic BiLSTM-CRF

(b) Vanilla Multi-task

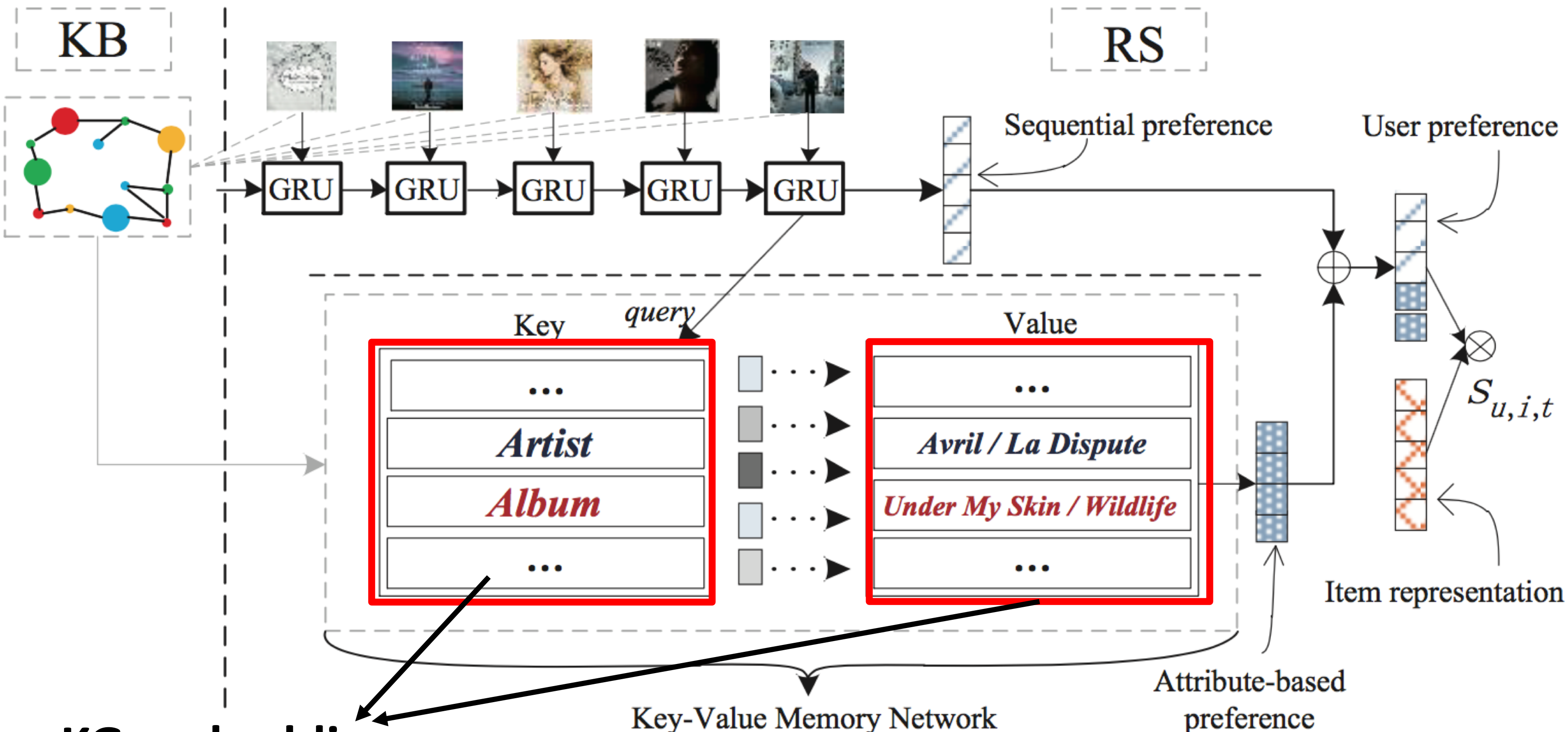
(c) Hierarchy Multi-task

(d) Deep Cascade Multi-task

Outline

- Introduction to Knowledge Graphs
- Applications
  - Entity Linking
  - Factoid Question Answering
  - E-Commerce Application
  - Knowledge Based Recommendation

KG enhanced Recommendation

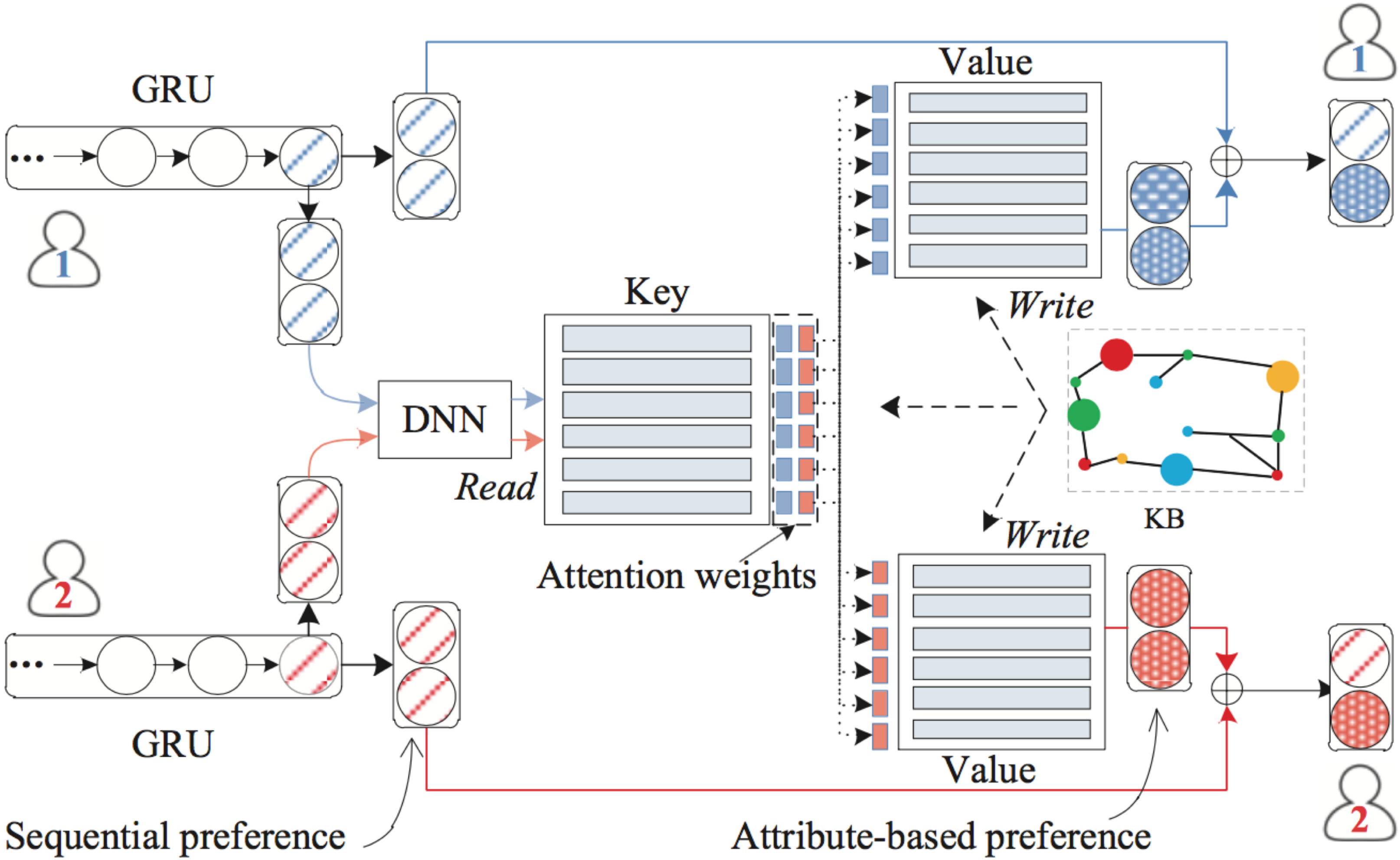


KG embedding

TransE: Item(h) + Key(r) = Value(t)

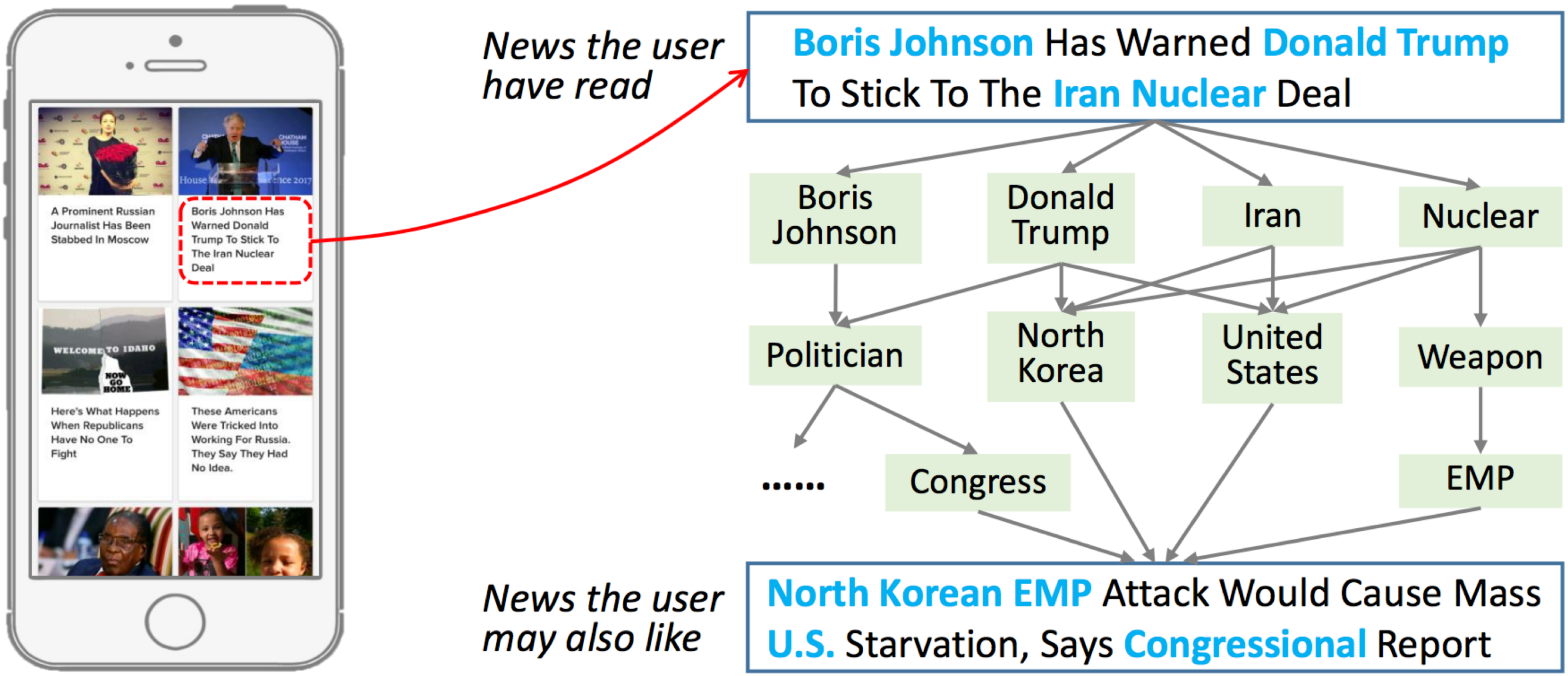
[Huang et al., 2018] ***SIGIR 2018***

KG enhanced Recommendation



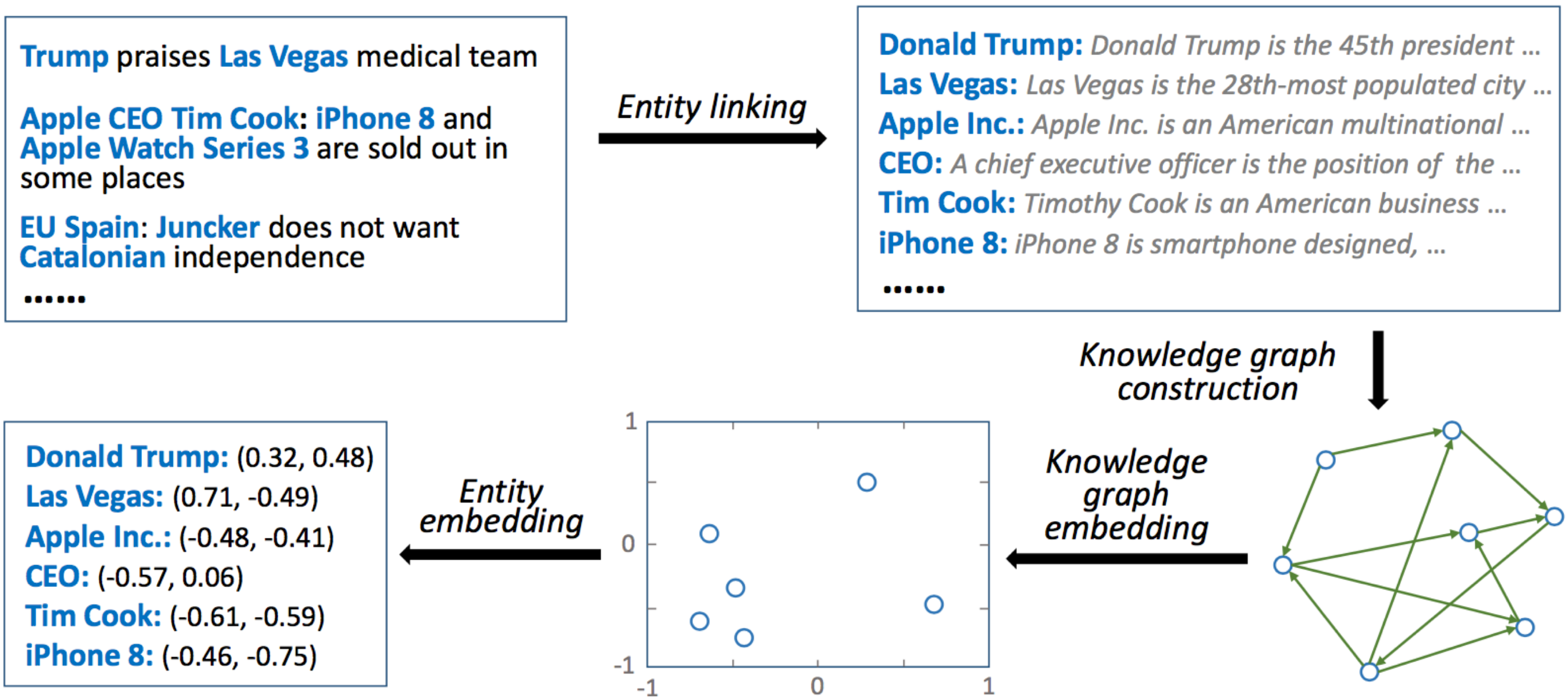
[Huang et al., 2018] ***SIGIR 2018***

KG enhanced Recommendation



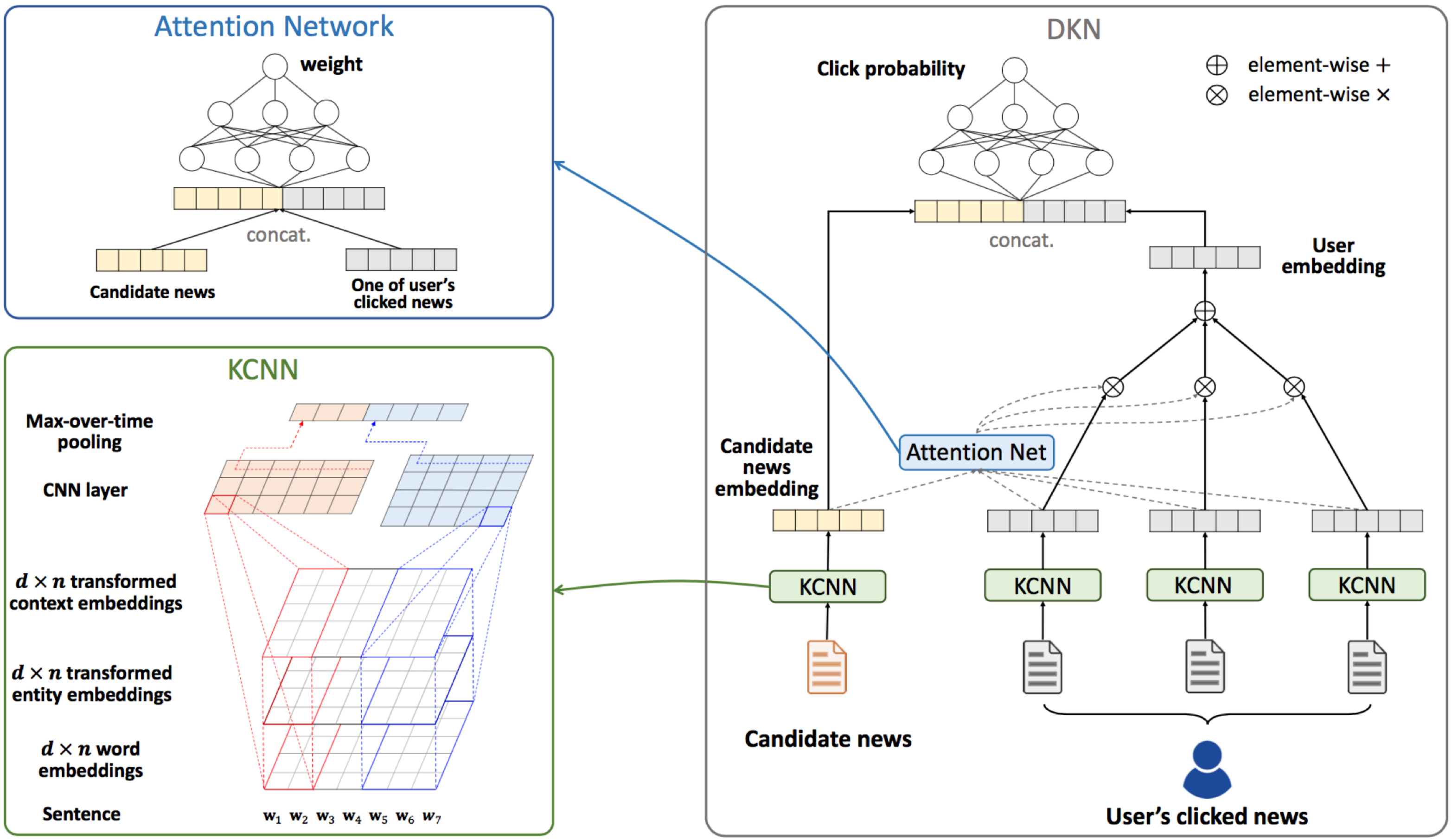
[Wang et al., 2018] ***WWW 2018***

KG enhanced Recommendation



[Wang et al., 2018] ***WWW 2018***

KG enhanced Recommendation



$$\mathbf{W} = \left[[\mathbf{w}_1\, g(\mathbf{e}_1)\, g(\overline{\mathbf{e}}_1)]\, [\mathbf{w}_2\, g(\mathbf{e}_2)\, \overline{g}(\mathbf{e}_2)] \ldots [\mathbf{e}_n\, g(\mathbf{e}_n)\, g(\overline{\mathbf{e}}_n)] \right]$$

[Wang et al., 2018] ***WWW 2018***

Conclusion

- External knowledge beyond training data
  - Text/video/item... understanding
  - entity linking as first step
- KG + DL: TransX embedding
  - memory nets
- Future AI: Common sense knowledge
  - casual: rain->umbrella, flood->damage
  - locatednear: arm & leg, door & room

References

- [Bao et al., 2016] Constraint-Based Question Answering with Knowledge Graph. ***COLING 2016***.
- [Berant et al., 2013] Semantic Parsing on Freebase From Question Answer Pairs. ***EMNLP 2013***.
- [Berant et al., 2014] Semantic Parsing via Paraphrasing. ***ACL2014.***
- [Yao et al., 2014] Information Extraction over Structured Data: Question Answering with Freebase. ***ACL2014.***
- [Bordes et al., 2015] Large-scale simple question answering with memory networks. ***arXiv 2015***.
- [Dai et al., 2016] CFO: Conditional Focused Neural Question Answering with Large-Scale Knowledge Bases. ***ACL 2016***.
- [Golub and He 2016] Character-level question answering with attention. ***EMNLP 2016***.

References

- [Lukovnikov et al., 2016] Neural Network-based Question Answering over Knowledge Graphs on Word and Character Level. ***WWW 2017***.
- [Yih et al., 2015] Semantic Parsing via Staged Query Graph Generation: Question Answering with Knowledge Base. ***ACL 2015***.
- [Yin et al., 2016] Simple Question Answering by Attentive Convolutional Neural Network. ***COLING 2016***.
- [Yu et al., 2017] Improved Neural Relation Detection for Knowledge Base Question Answering. ***ACL 2017***.
- [Yang and Chang, 2015] S-MART: Novel tree-based structured learning algorithms applied to tweet entity linking. ***ACL 2015***.
- [Hao et al., 2017] An End-to-End Model for Question Answering over Knowledge Base with Cross-Attention Combining Global Knowledge. ***ACL 2017***.
- [Qiu et al., 2018] Hierarchical Type Constrained Topic Entity Detection for Knowledge Base estion Answering. ***WWW2018 Poster.***

References

- [Wang et al., 2018] DKN: Deep Knowledge-Aware Network for News Recommendation ***WWW 2018***.
- [Huang et al., 2018] Improving Sequential Recommendation with Knowledge-Enhanced Memory Networks. ***SIGIR 2018***.

THANK YOU

Xusheng Luo (棋落)
Sept 06, 2018