Question Answering over Knowledge Bases

Wanyun Cui
SHUFE & Fudan

Backgrounds

- Question Answering (QA) systems answer questions posed by humans in a **natural language**.

IBM Watson

Watson prevailed over the human competitors and received the first place prize of $1 million in 2011.

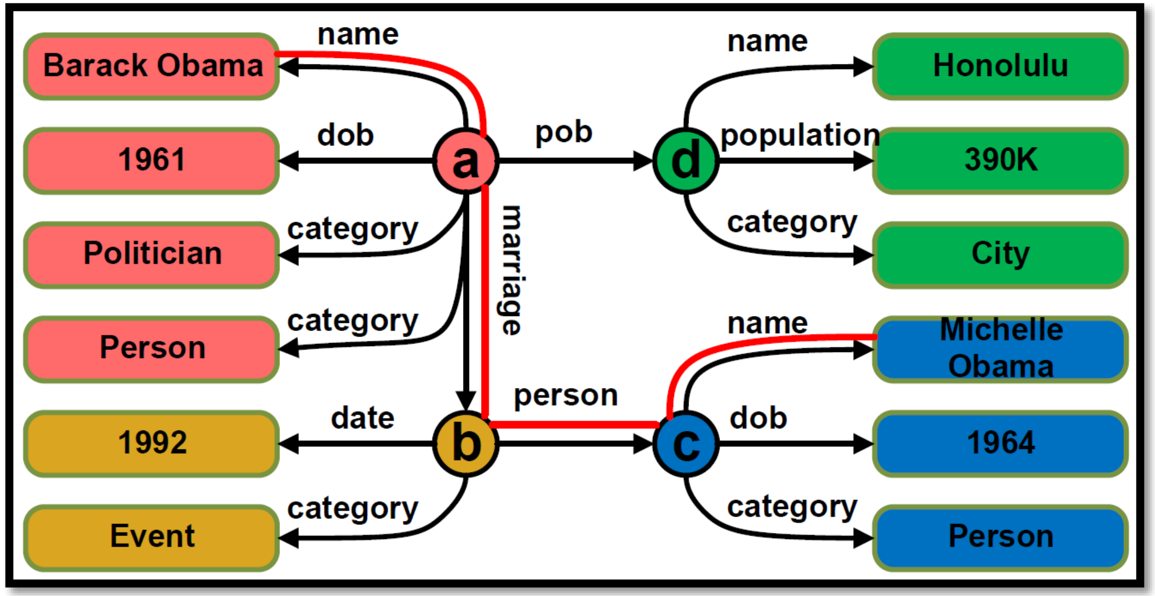
Significance of QA

- QA applications:
  - Lower the bar of human-computer interaction
  - New entrance for knowledge of internet

- QA provides:
  - **Application** of different semantic understanding models’ integration
  - **Real needs** of different models’ association analysis, data sharing, parameter transfer.
  - **Technical vision** of different NLP models’ breakthrough

A toy knowledge base

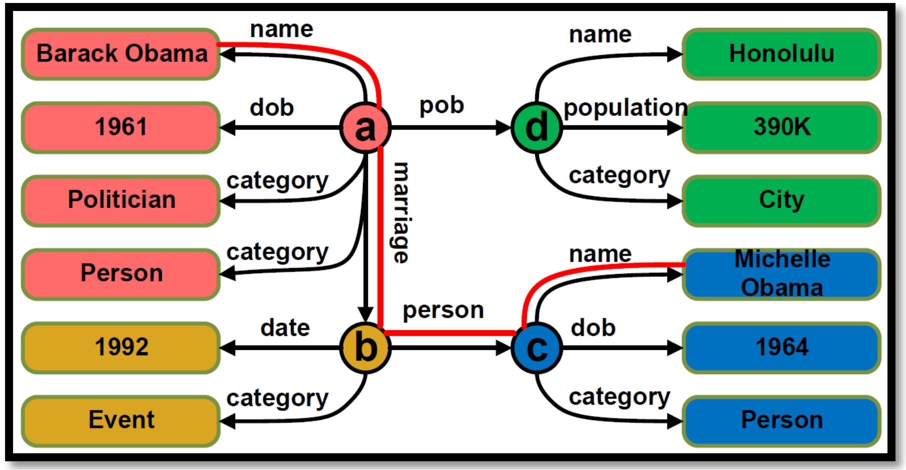
- Structured & linked data
- Each edge represents a knowledge
  - (d, population, 390k)



How KB-based QA works?

- Convert natural language questions to structured queries over knowledge bases.

How many people live in Honolulu?



SPARQL

Select ?number
Where {
Res:Honolulu
dbo:population ?num
}

SQL

Select value
From KB
Where subject='d' and
predicate='population'

- **Key: predicate inference**

Why knowledge base?

- Linked data – knowledge representation
  - Plain text: similarity between question and sentence.
  - KB: relational data provide semantics for question understanding
- Data quality – QA precision
  - Plain text: errors or contradictions in different texts
  - KB: high quality data from human labeling or table in web.
- Structured data – query efficiency
  - Plain text: inverted index
  - KB: stored in database, indexed by subject

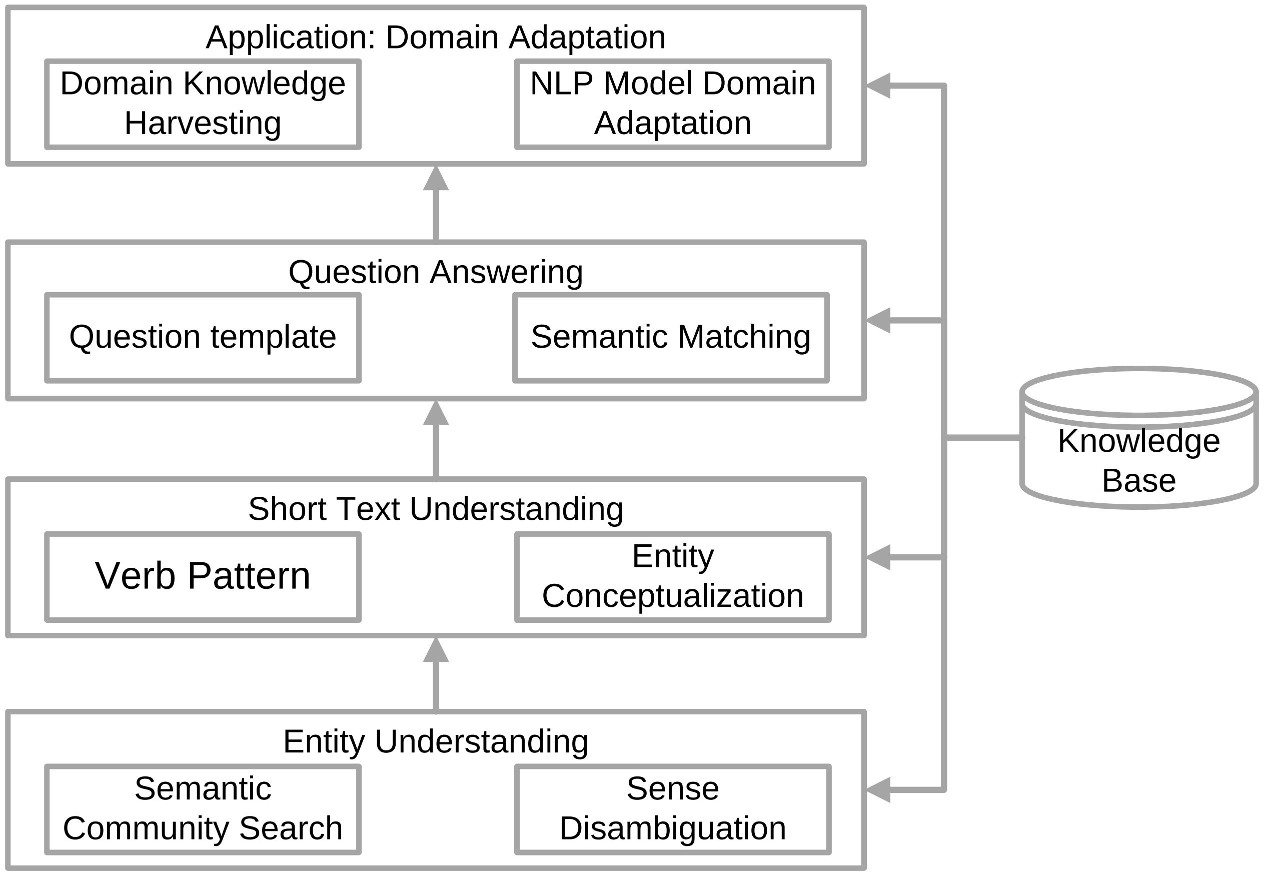
QA is a systematic work

Adapt QA system to specific domains

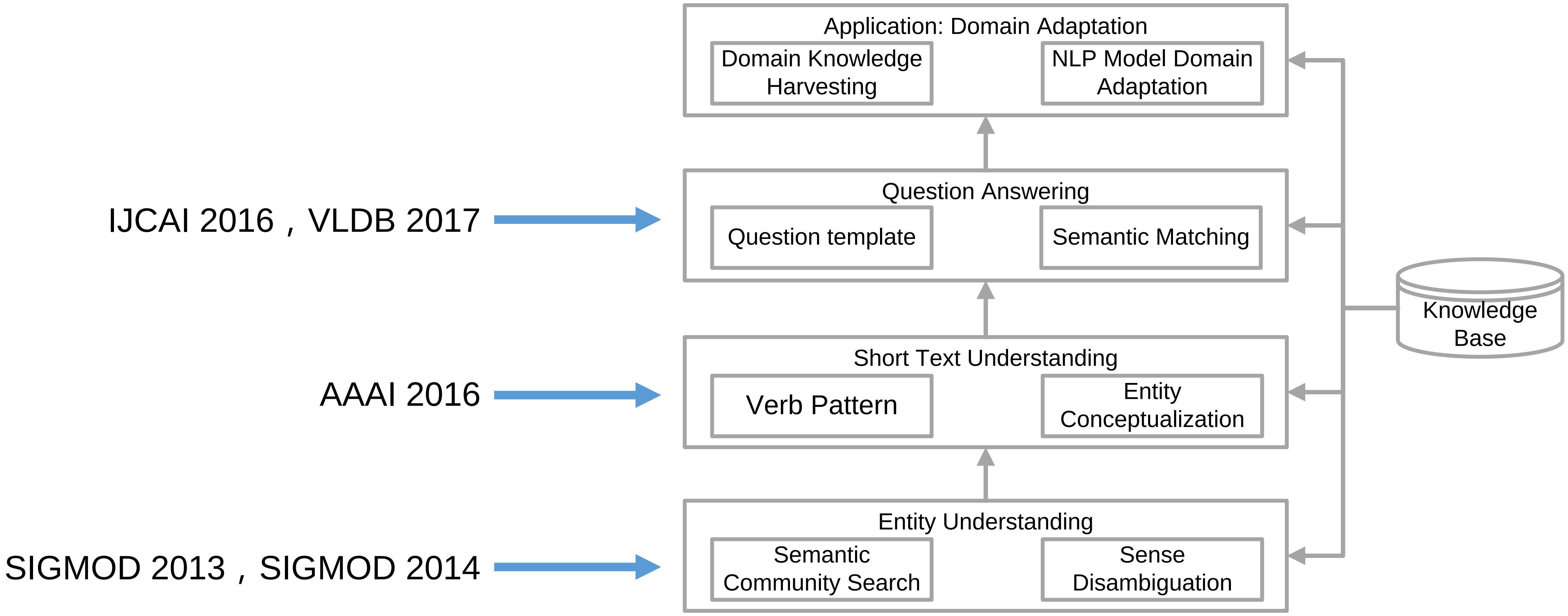
Core level: question answering

Short text connects entities/words and questions

Provides the basic semantic computing for entities in questions



QA is a systematic work



Question Answering

Weakness of previous works

- Template / rule based approaches
  - Represent sentences by templates
  - By human labeling
  - PROs:
    - User-controllable
    - Applicable to industry use
  - CONs:
    - Relies on manpower. Too costly.
    - Cannot handle the diversity of questions.

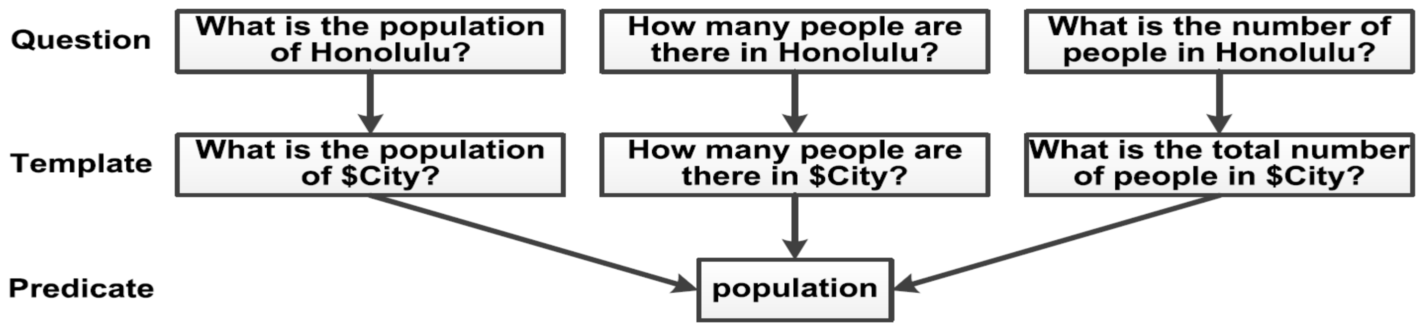
- Neural network based approaches
  - Represent sentences by embeddings
  - By learning from corpus
  - PROs:
    - Feasible to understand diverse questions
  - CONs:
    - Poor interpretability
    - Not controllable. Unfriendly to industrial application.

Our approach

- Represent natural language questions by templates.
  - E.g.
    - How many people are there in $city?
  - Interpretable
  - User-controllable

- Learn templates from QA corpus.
  - Understand diverse questions
  - 27 million templates, 2782 intents

QA by templates



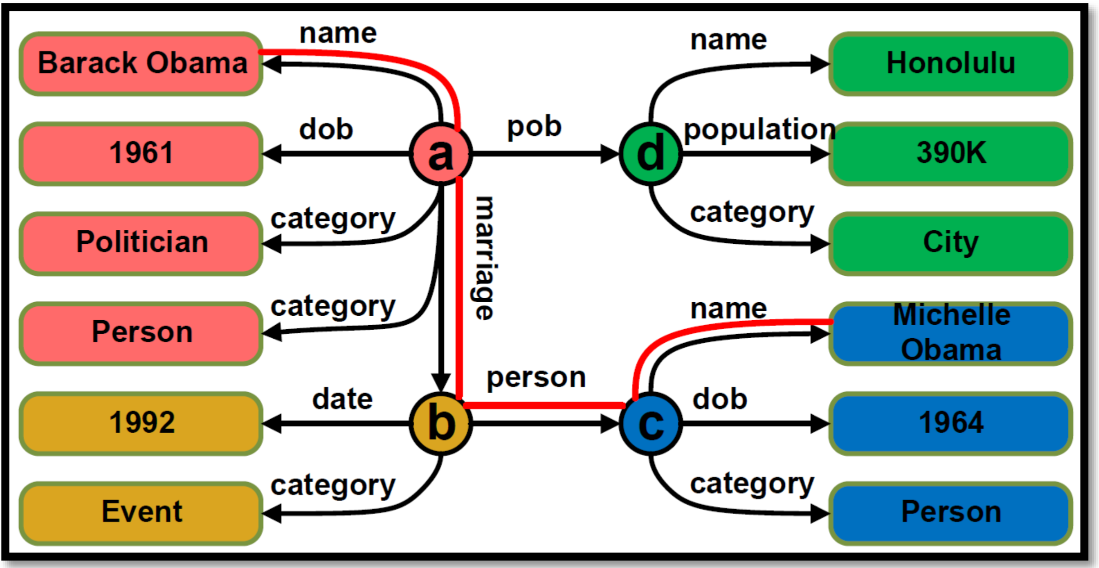
Template: replace the entity of question by its concept

A template represents complete intent of the question.

Key problem: collect templates and identify their corresponding predicates

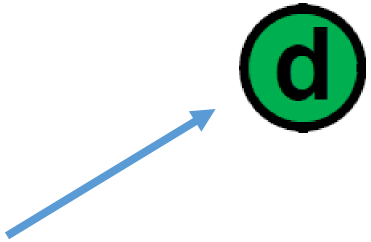
question2answer: a generative process

- A qa pair
  - Q: How many people live in Honolulu?
  - A: It's 390K.

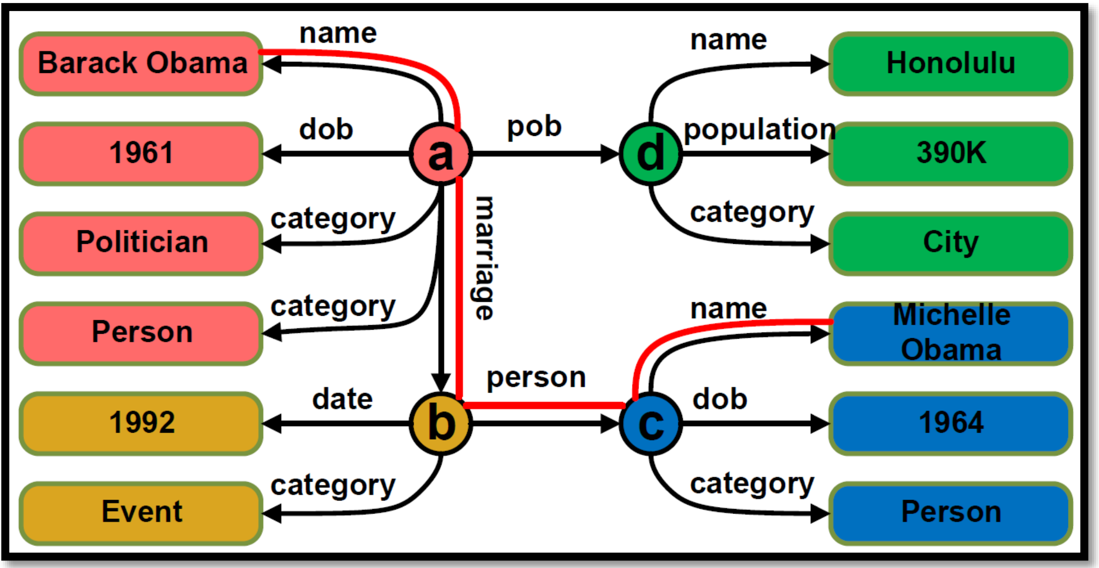


kw.fudan.edu.cn/qa

question2answer: entity linking



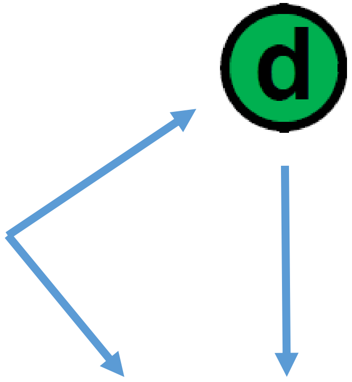
How many people live in Honolulu?



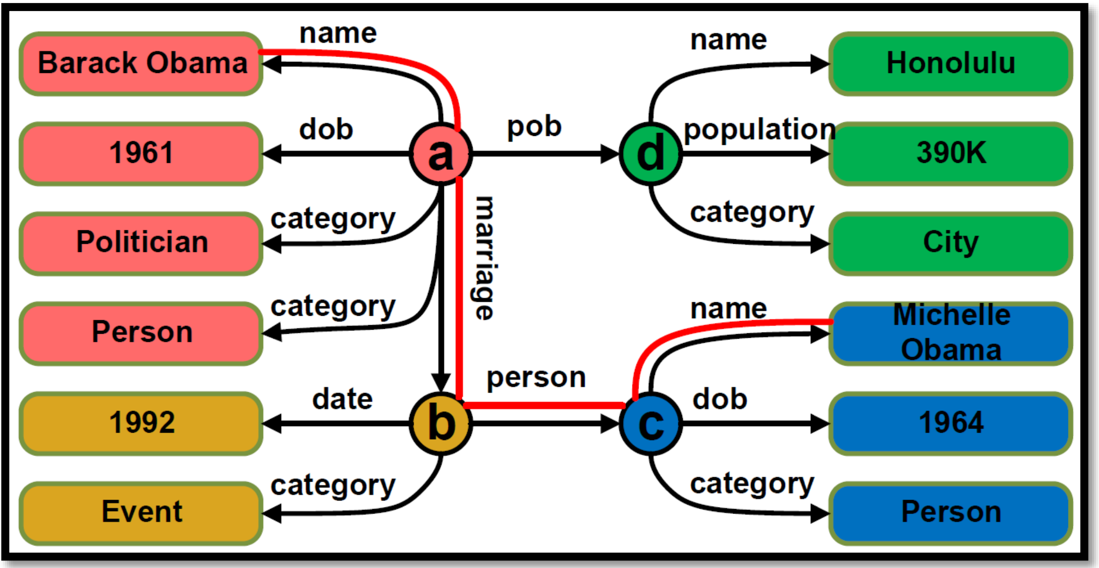
kw.fudan.edu.cn/qa

question2answer: conceptualization

How many people live in Honolulu?



How many people live in $city?

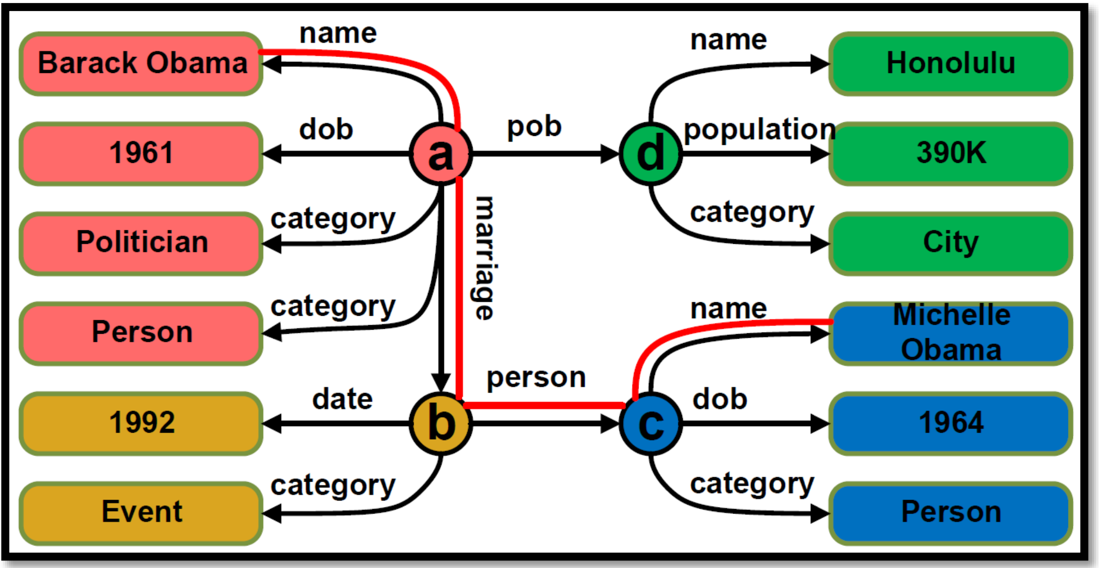


kw.fudan.edu.cn/qa

question2answer: predicate inference

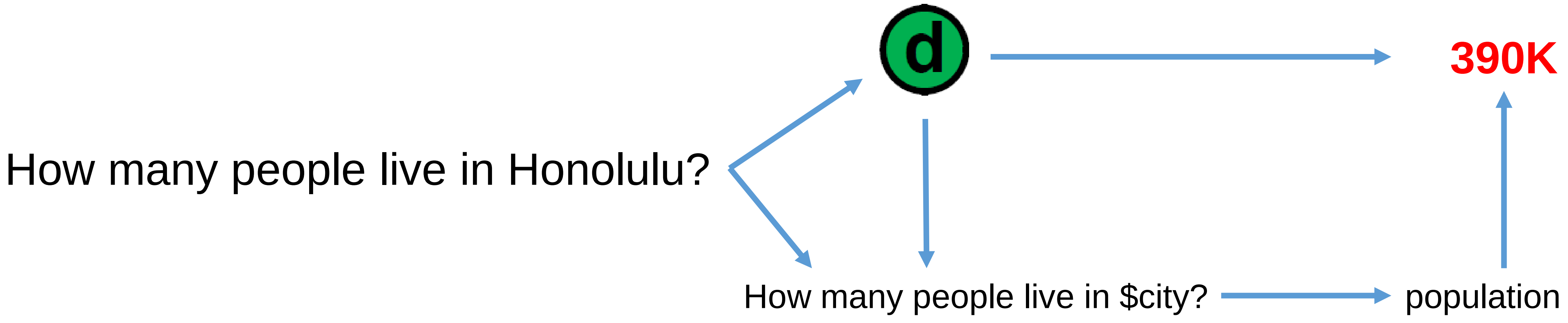
How many people live in Honolulu?

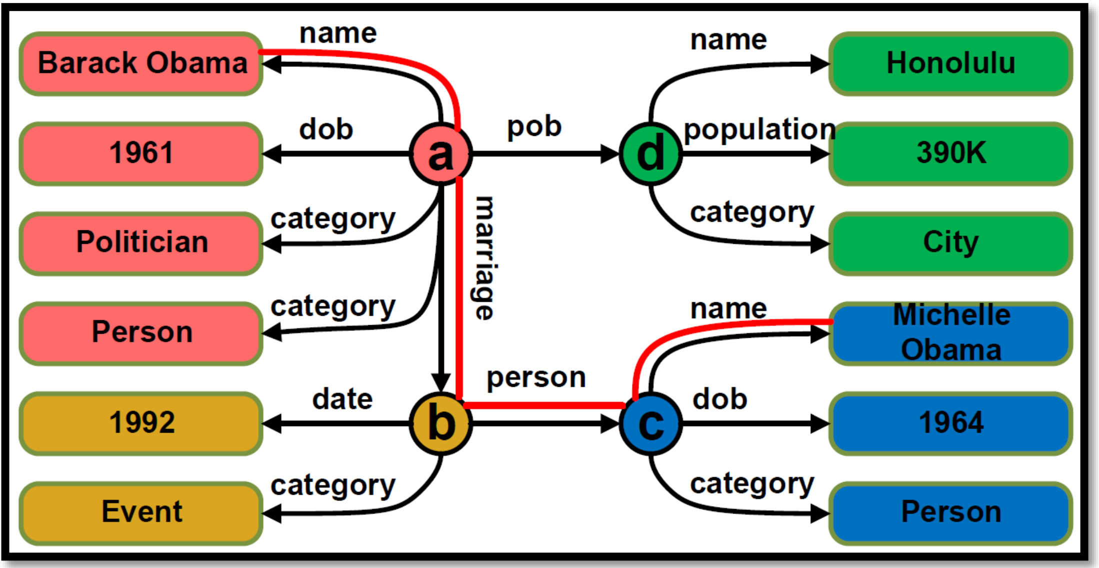
How many people live in $city? → population



kw.fudan.edu.cn/qa

question2answer: value lookup

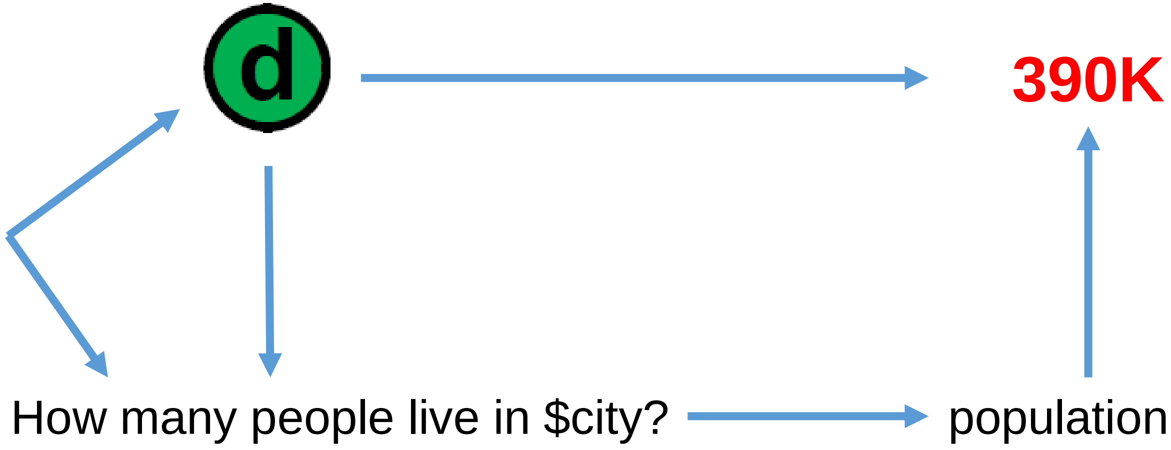


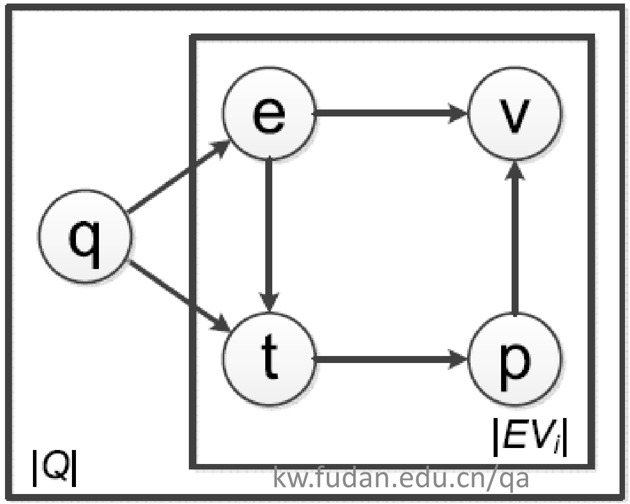


kw.fudan.edu.cn/qa

Probabilistic graph model

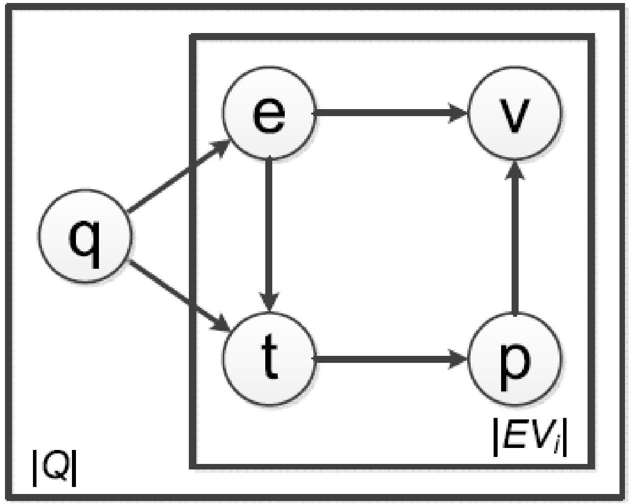
How many people live in Honolulu?





Probabilistic inferencing

- Learning parameters from QA corpora (42M Yahoo! Answers)
  - Intuition: **maximize the likelihood** of observing such QA corpora



问答系统:实验结果

- 27126355个问题模板
- 2782个问题意图

	#pro	#ri	#par	R	R_{BFQ}	R^*	R^*_{BFQ}	P	P_{BFQ}	P^*	P^*_{BFQ}
squall2sparql	96	80	13	.78	.81	.91	.94	.84	.95	.97	.95
SWIP	21	14	2	.14	.24	.16	.24	.67	.77	.76	.77
CASIA	52	29	8	.29	.56	.37	**.61**	.56	.79	.71	.86
RTV	55	30	4	.30	.56	.34	.56	.55	.72	.62	.72
gAnswer [38]	76	32	11	**.32**	.54	**.43**	-	.42	.54	.57	-
Intui2	99	28	4	.28	.54	.32	.56	.28	.54	.32	.56
Scalewelis	70	32	1	**.32**	.41	.33	.41	.46	.50	.47	.5
KBQA+KBA	25	17	2	.17	.42	.19	.46	.68	.68	.76	.76
KBQA+FB	21	15	3	.15	.37	.18	.44	.71	.71	.86	.86
KBQA+DBp	26	25	0	.25	**.61**	.25	**.61**	**.96**	**.96**	**.96**	**.96**

QALD-3上的结果
（KB-based）

微信端界面

Wanyun Cui**, et al., KBQA: Learning Question Answering over QA Corpora and Knowledge Bases, **(VLDB 2017)
***Wanyun Cui**, et al., KBQA: An Online Template Based Question Answering System over Freebase, **(IJCAI 2016)**, demo*

Weakness of previous works

- Template / rule based approaches
  - Represent sentences by templates
  - By human labeling
  - PROs:
    - User-controllable
    - Applicable to industry use
  - CONs:
    - Relies on manpower. Too costly.
    - Cannot handle the diversity of questions.
- Neural network based approaches
  - Represent sentences by embeddings
  - By learning from corpus
  - PROs:
    - Feasible to understand diverse questions
  - CONs:
    - Poor interpretability
    - Not controllable. Unfriendly to industrial application.

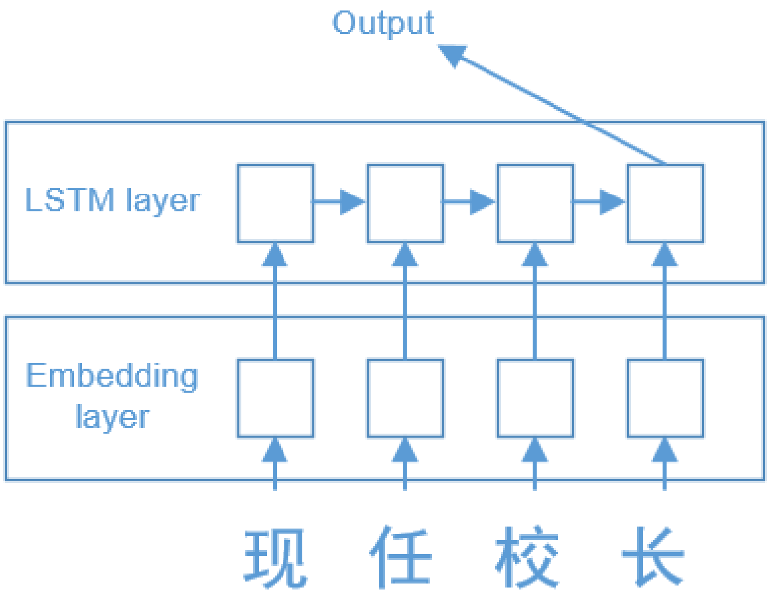
Multi-granularity deep neural network

Problem analysis: features

- From questions:
  - Question
    - 复旦大学的校长是谁
- From knowledge bases:
  - Value
    - 许宁生
  - Entity description
    - 中国科学院院士
  - Predicate
    - 现任校长

Model: 3 granularities for 1 feature

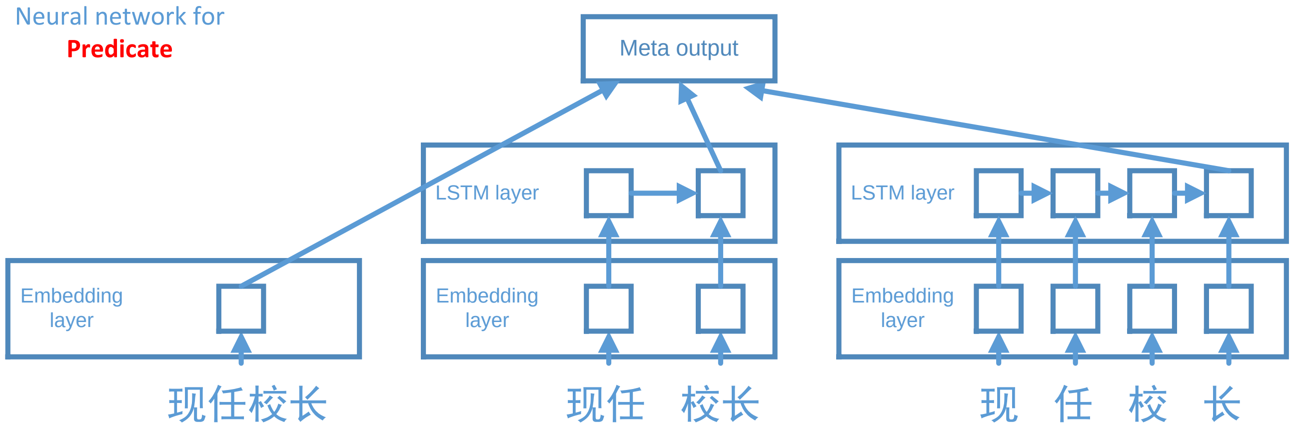
- 3 granularities for 1 feature
  - 现任校长
  - One hot:
    - 现任校长
  - Word list:
    - 现任 校长
  - Char list:
    - 现 任 校 长
- LSTM for char list



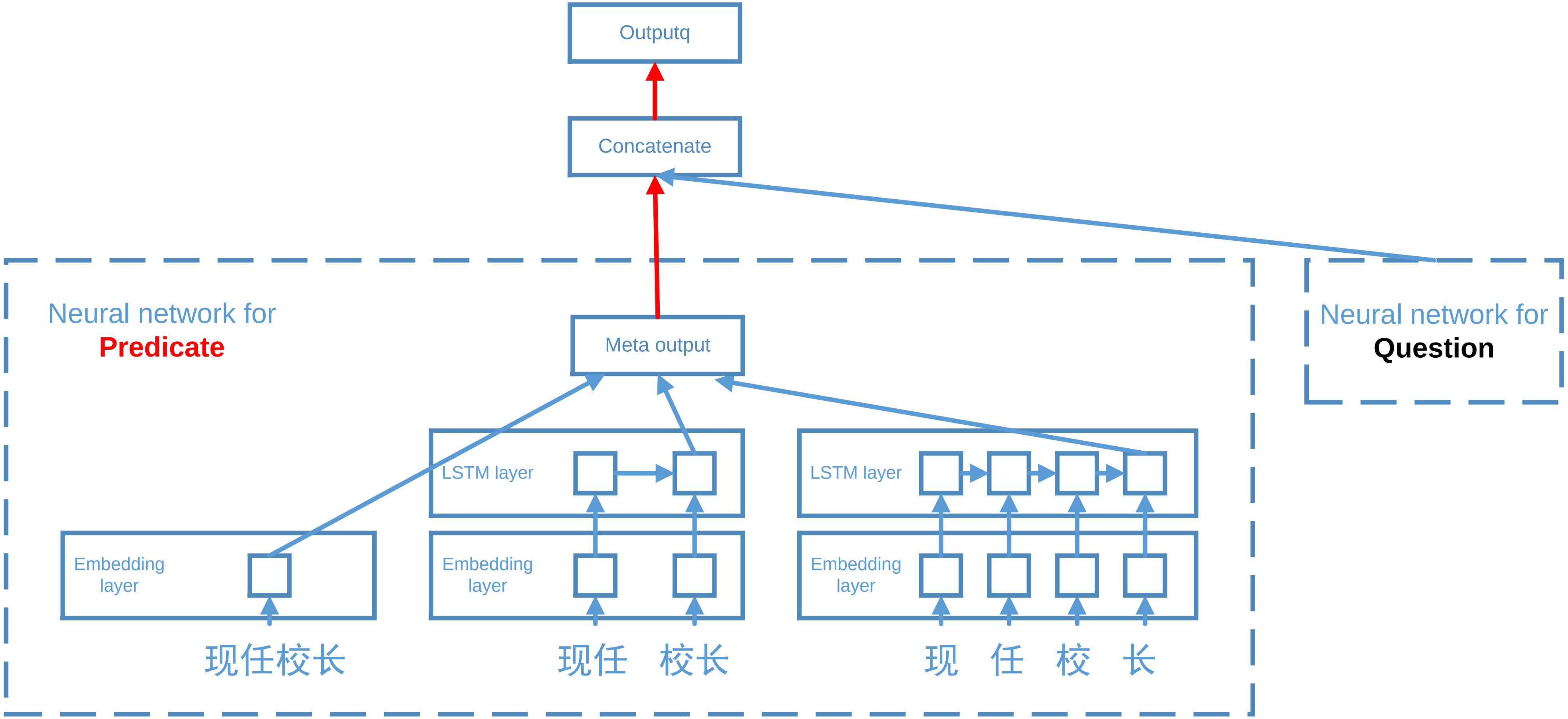
feature2vector: aggregation

- 3 representation granularity
  - 现任校长
  - One hot:
    - 现任校长
  - Word list:
    - 现任 校长
  - Char list:
    - 现 任 校 长

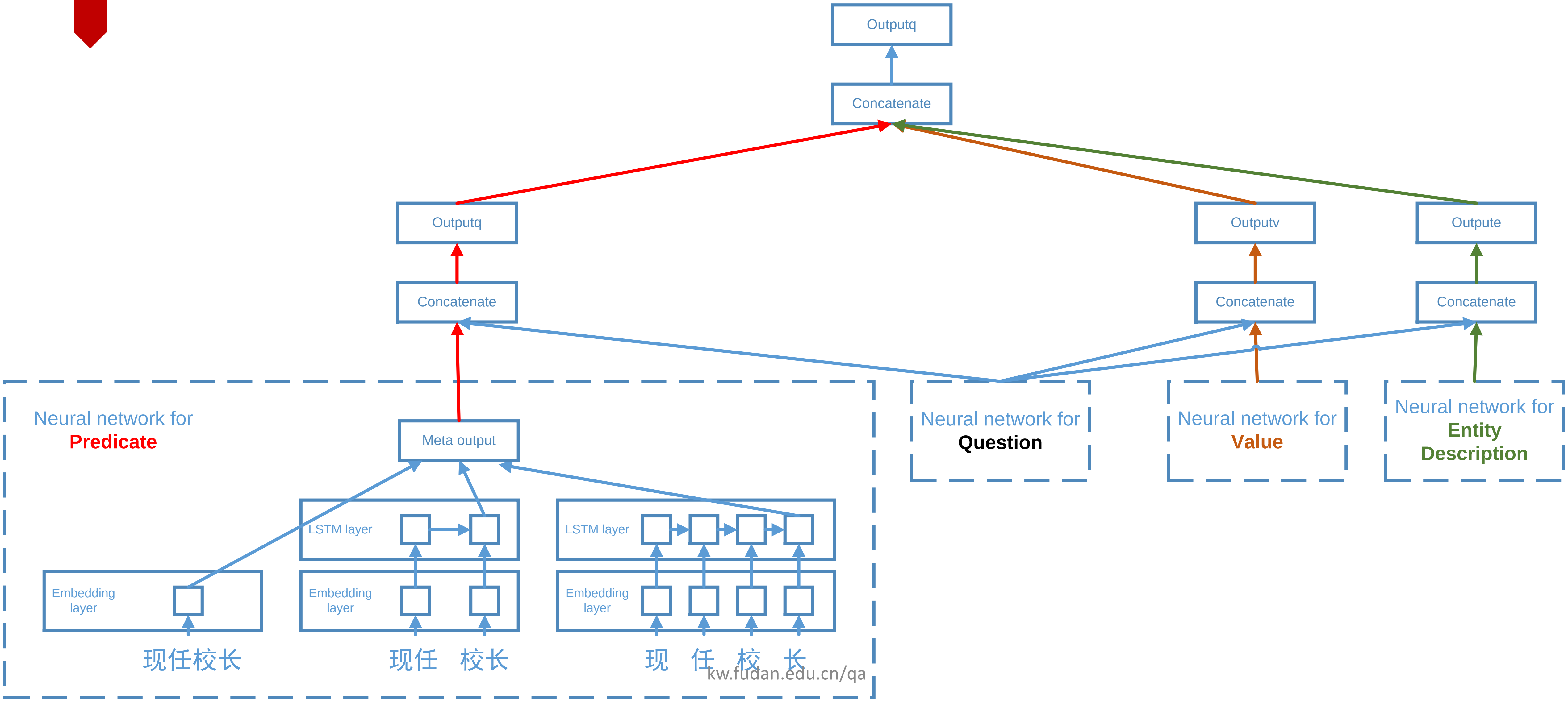
- Aggregation



predicate - question similarity

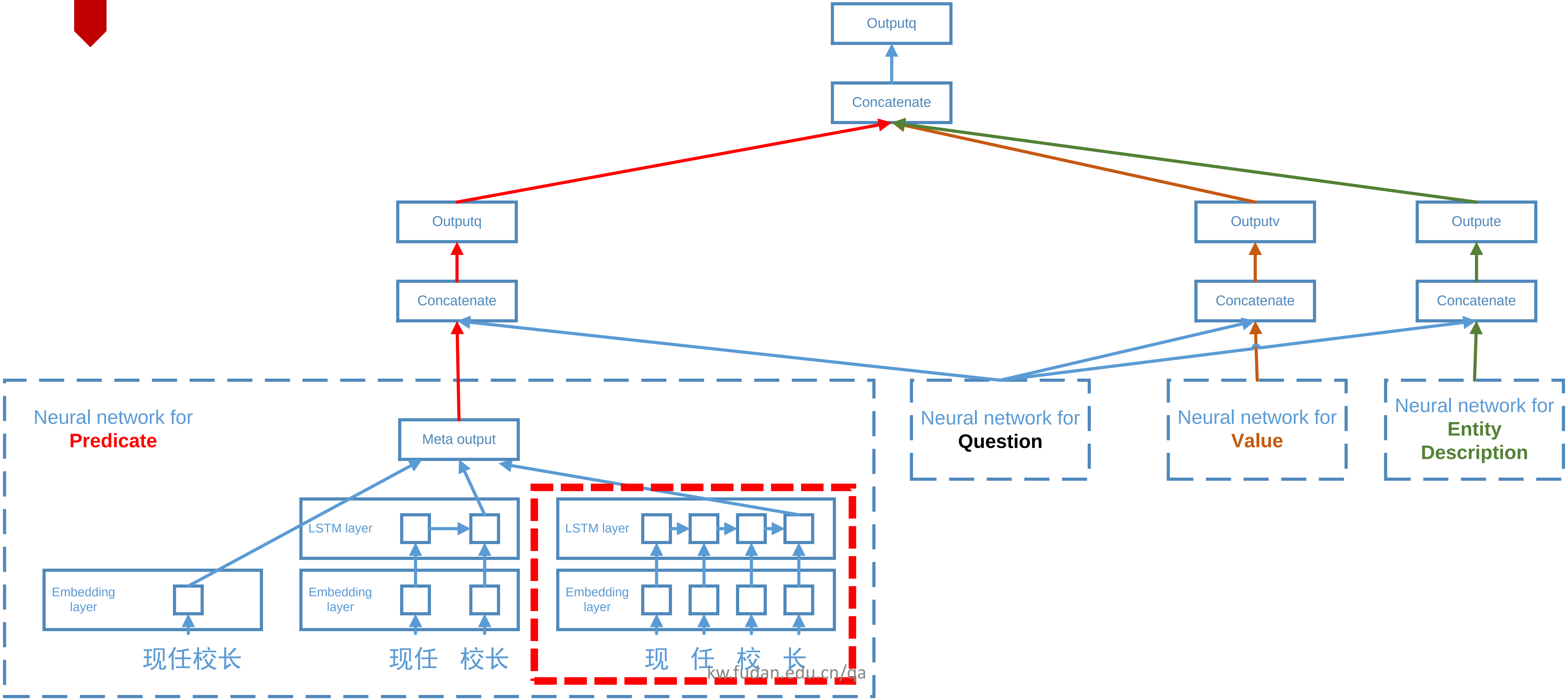


Multi-granularity deep neural network

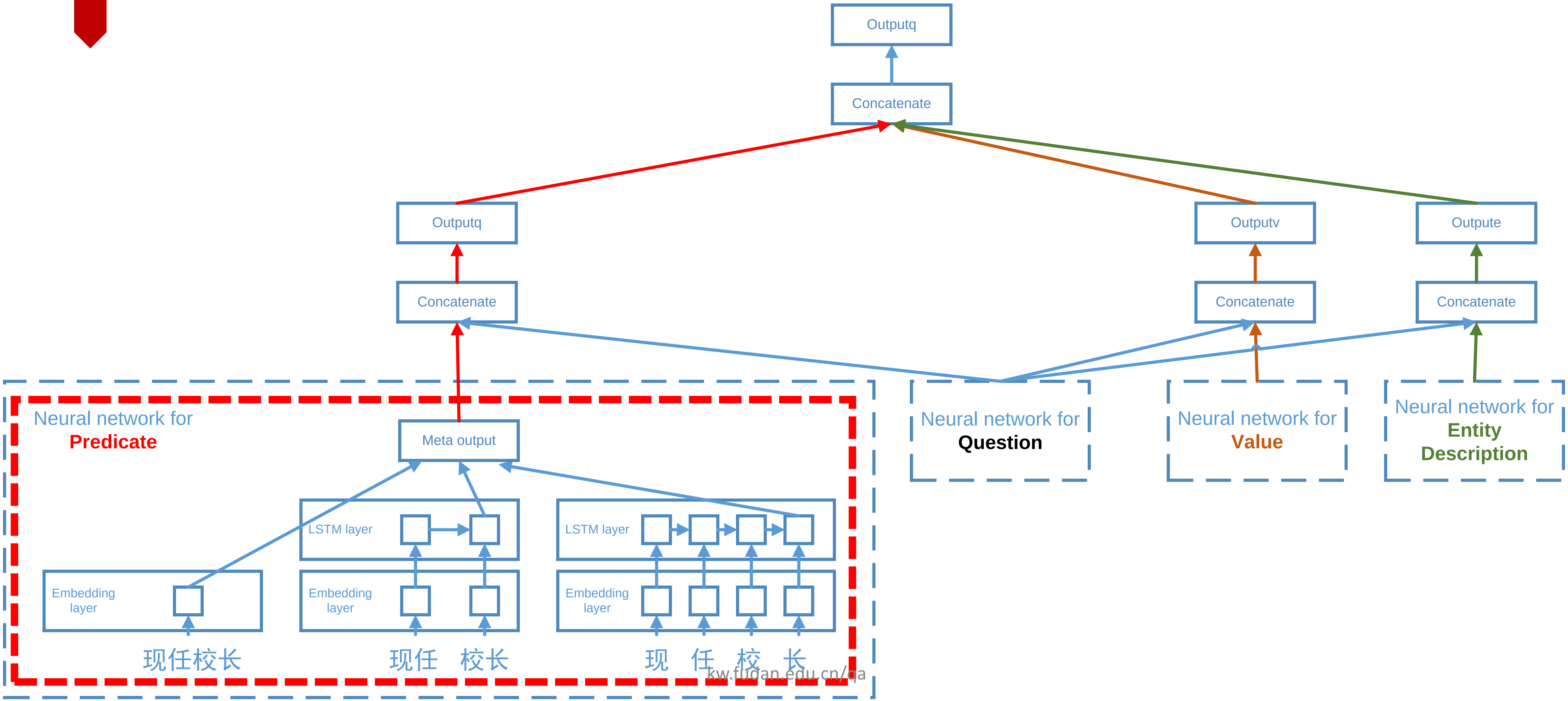


kw.fudan.edu.cn/qa

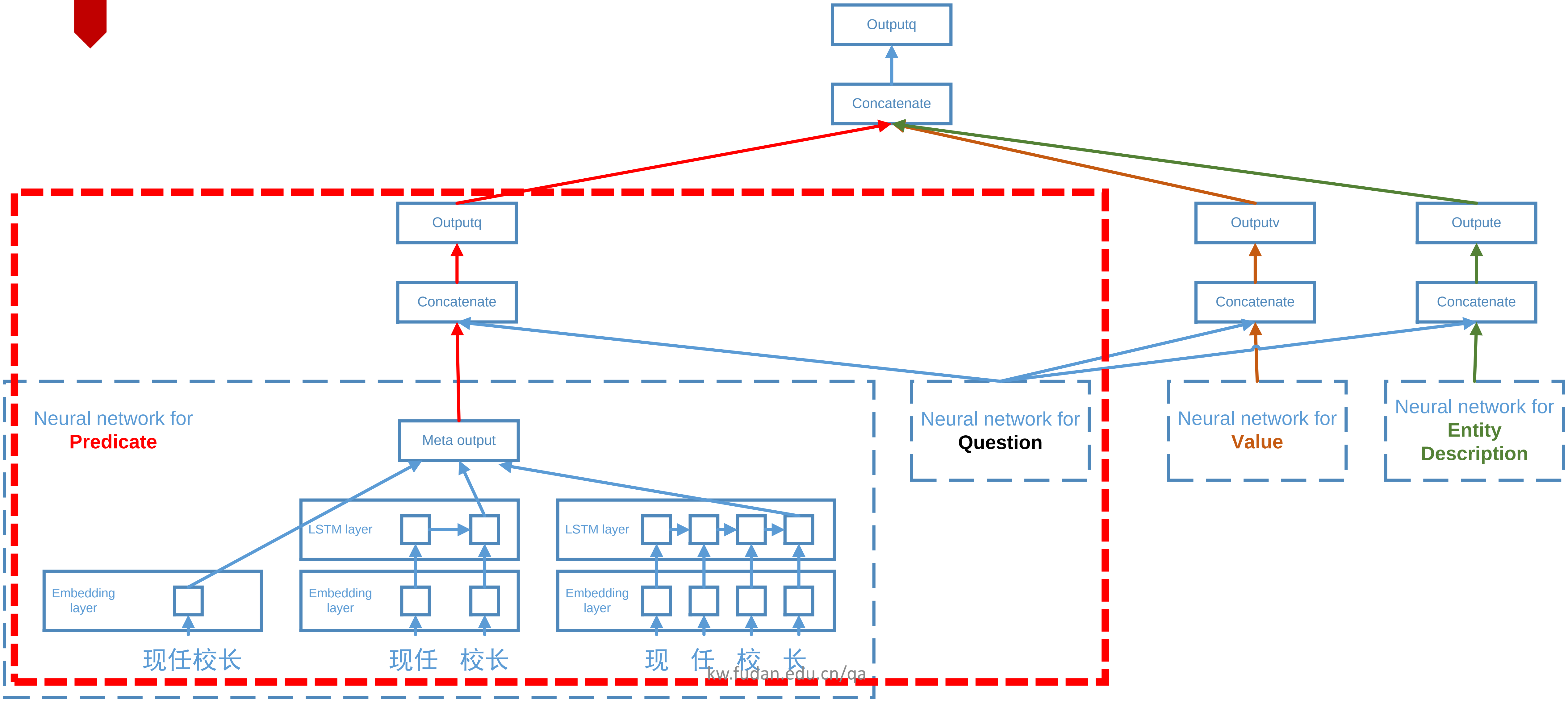
Multi-granularity deep neural network: One granularity



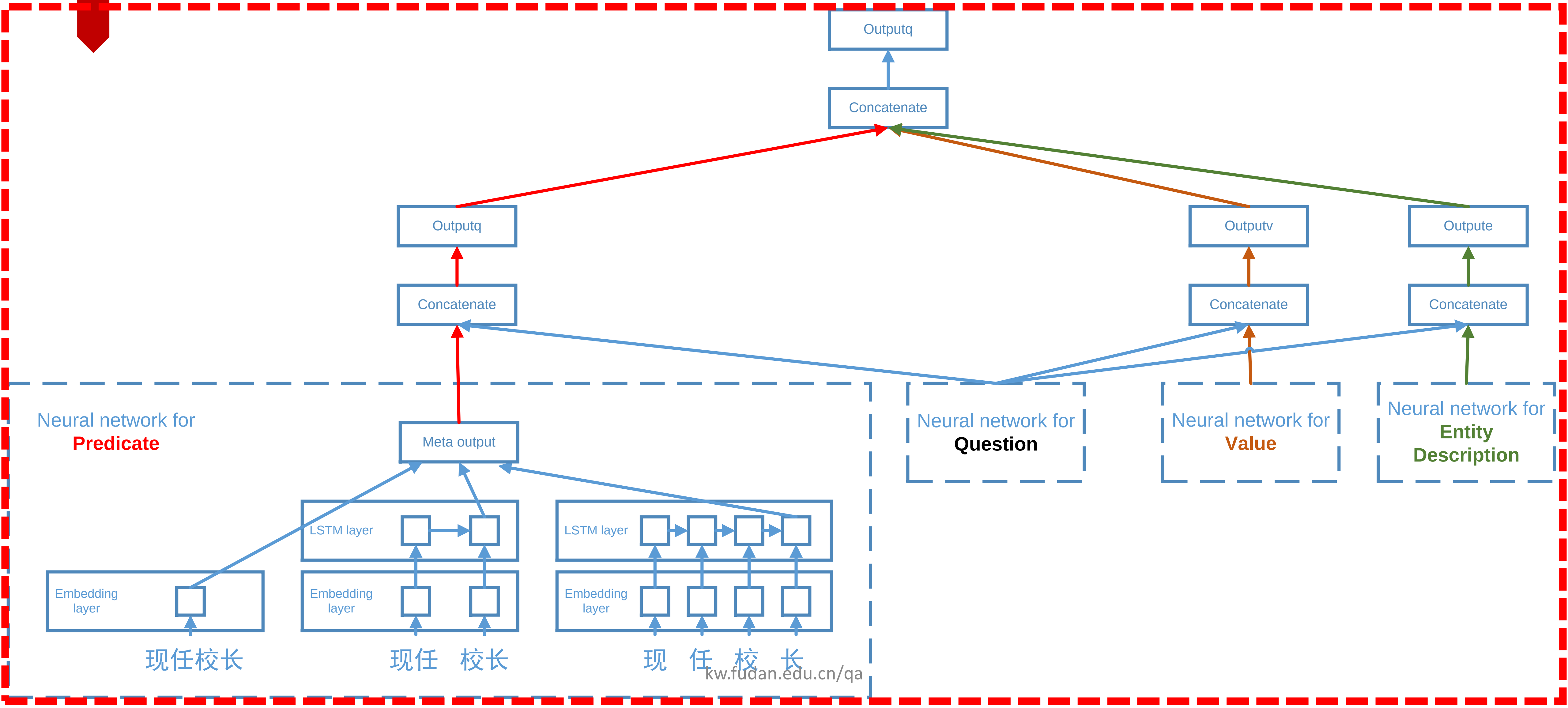
Multi-granularity deep neural network: One feature



Multi-granularity deep neural network: Feature-Question



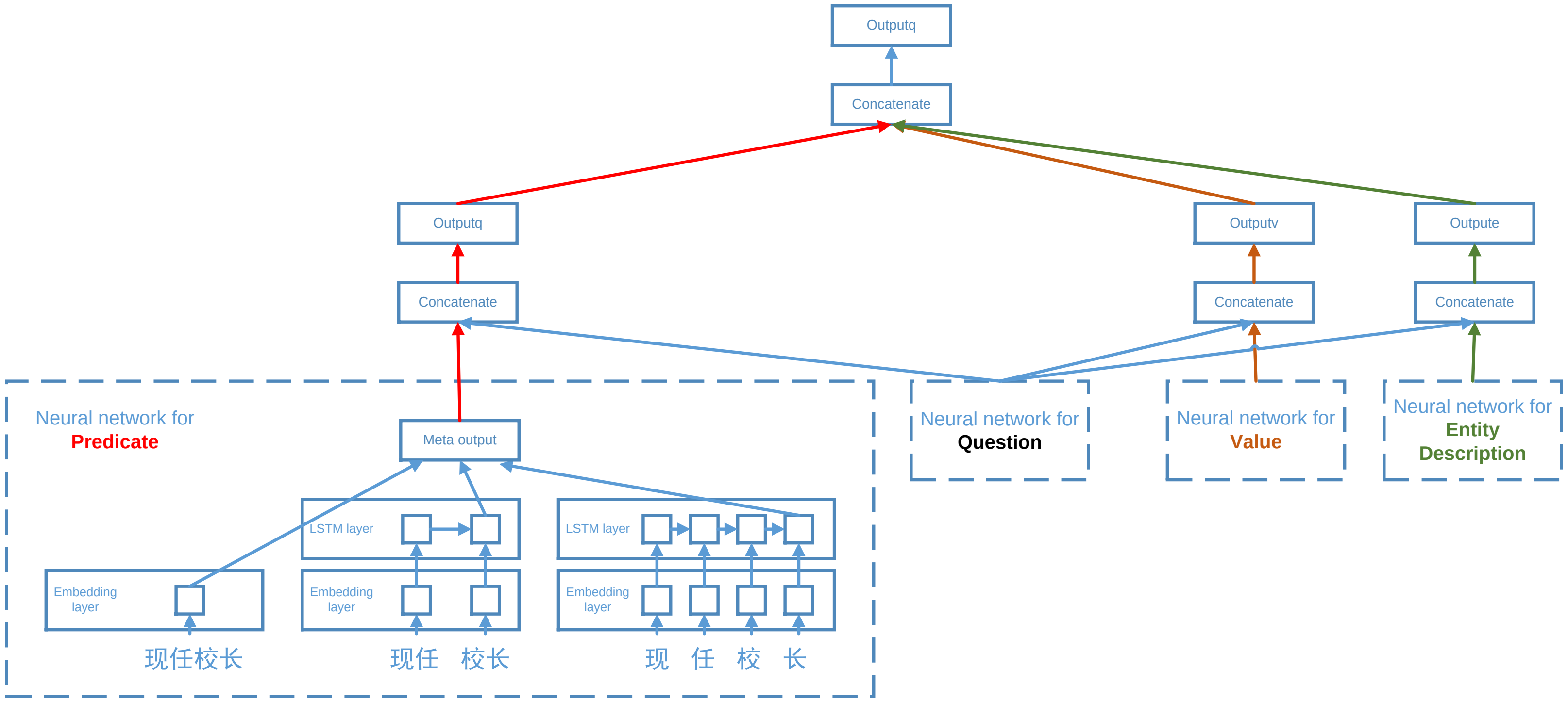
Multi-granularity deep neural network: All Features - Question



Multi-granularity deep neural network

- Fully utilize all the features.

- Reasonable network structure



Results

- NLPCC 2016
- Our score: **0.92** (improving)
- KBQA beats all competitors.

KBQA Submissions	F1 Score
北京大学	**0.8247**
国防科学技术大学	**0.8159**
华中师范大学	**0.7957**
哈尔滨工业大学 (HIT-SCIR)	**0.7914**
东北大学 (自然语言处理实验室)	**0.7272**
Harbin ShenZhi Technology Co., Ltd.	**0.7251**

DAKSE: domain knowledge extraction

Domain knowledge extraction: Domain-awareness

- S1 is a valid knowledge for AI researchers.
- S2 is a valid knowledge for college student.

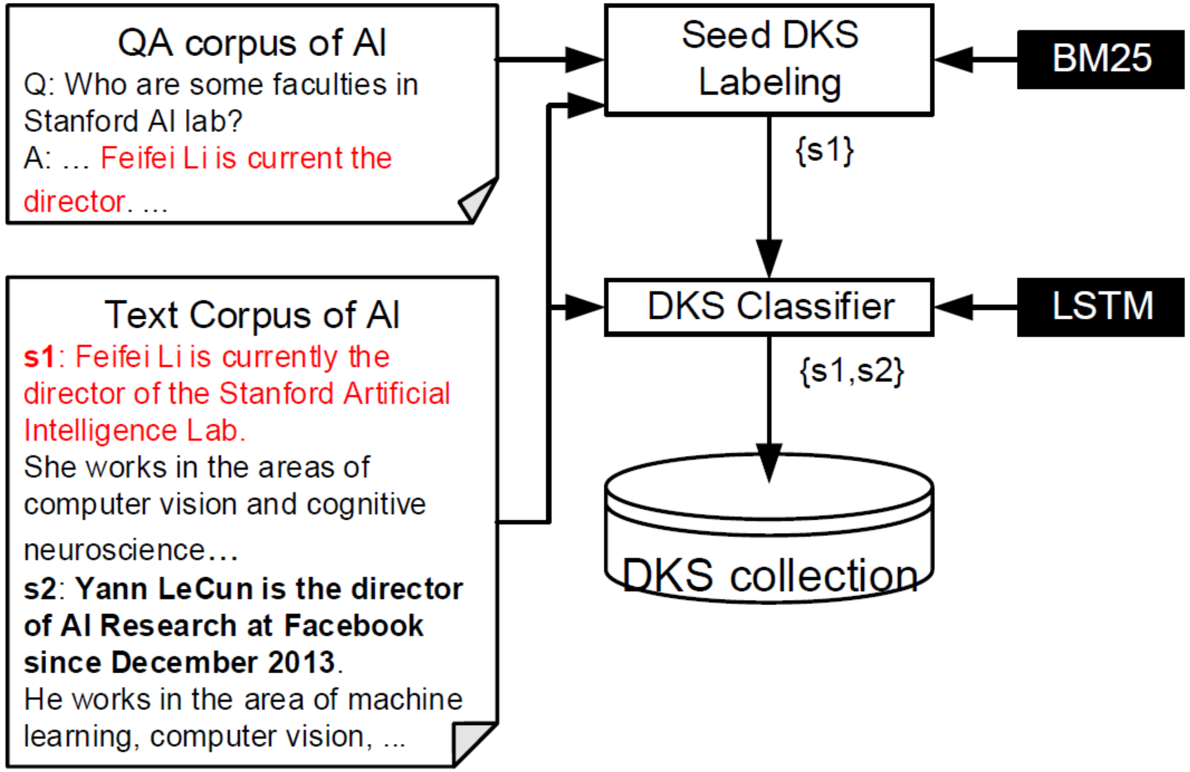
- Whether a sentence being knowledge-rich depends on the ***domain***.

Corpus for Stanford:

- S1: Feifei Li is currently the director of the Stanford Artificial Intelligence Lab. . . .
- S2: Full-time undergraduate tuition was $42,690 for 2013-2014 in Stanford.

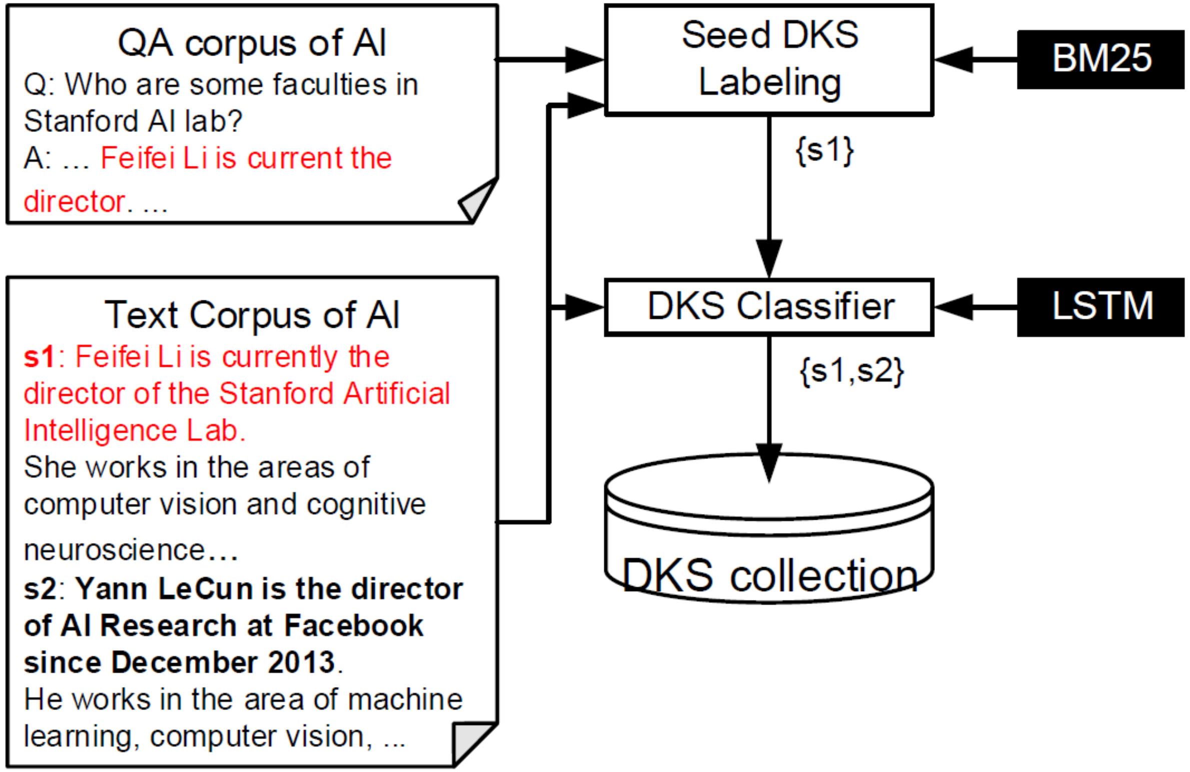
Domain knowledge extraction: Approach

- Intuition
  - For QA corpus of a given domain, its answers are knowledge-rich for that domain.
  - Learn the domain-aware knowledge representations in answers.



Domain knowledge extraction: Approach

- Seed knowledge labeling:
  - Leverage QA corpora to label seed dks for further training
  - s1 is extracted
    - s1 is similar to the answer
- DK Identifier:
  - Identify patterns in seed dks
  - Extract more dks in text corpora
  - s1, s2 are extracted
    - s2 is similar to s1



Domain knowledge extraction: Results

Method	Correct	Irrelevant	Incorrect
Topic Model + SVM	52%	9%	39%
Language Model	56%	5%	39%
DAKSE	**68%**	3%	29%

Evaluation for China Mobile Customer Service

Domain knowledge extraction: Results

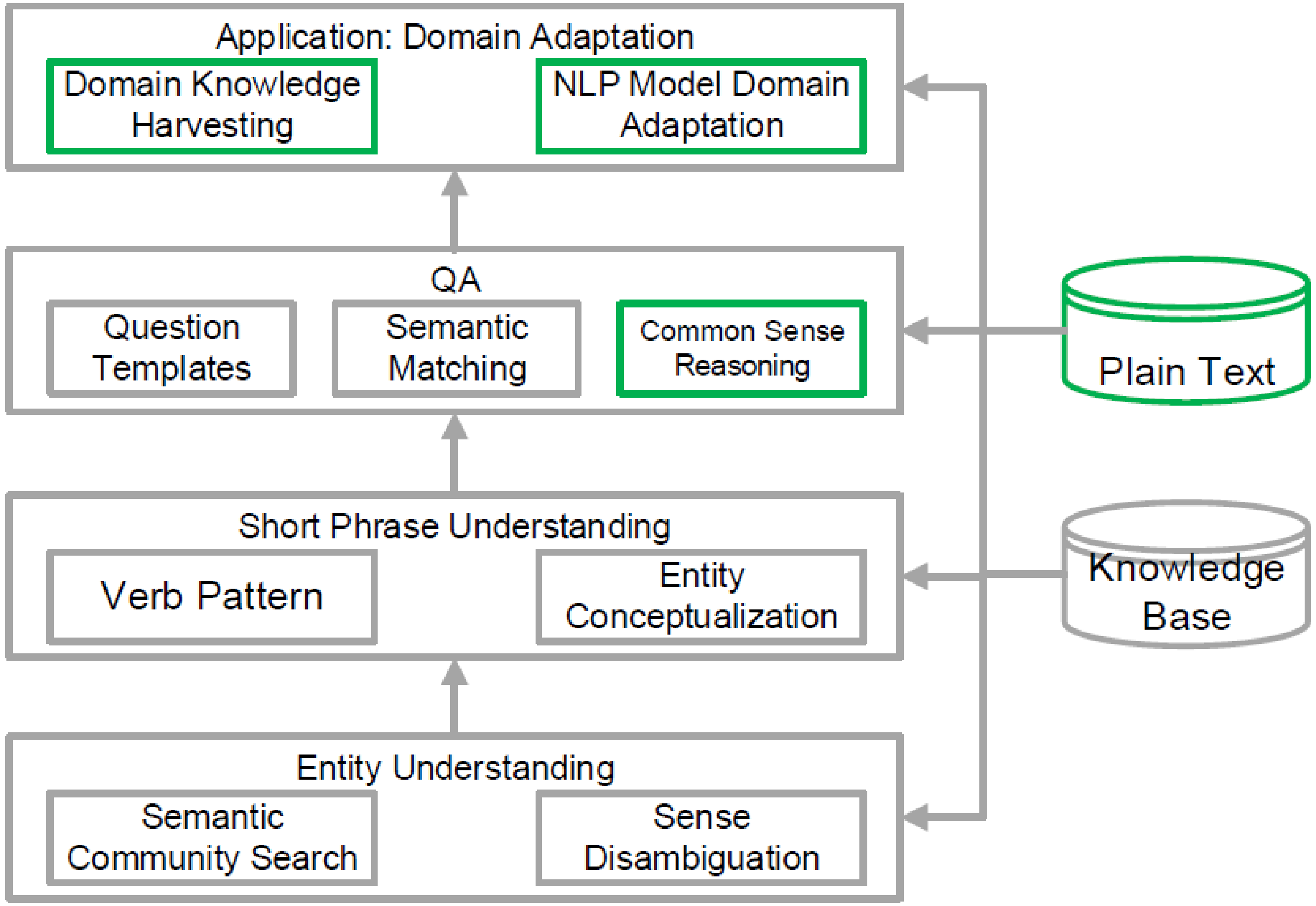
DKSs	can go through, can visit, contain, mainly contain, can visit link for, please send, reply to, welcome to, directly reply to, be divide into
original	will, can, have, see, can get, will receive, can partic-ipate in, can use, can enjoy, can apply for

Top 10 predicates in tuples for China Mobile Customer Services

With DAKSE, tuples by open IE are more focused.

Future works

- Domain adaptation
  - More applicable
- Common sense reasoning
  - Deeply understand the question
- Hybrid QA
  - Plain text + KB
  - Improve the effectiveness



Thank you!

Wechat QR code for our system.