

Multi-Evidence Question Answering with Free-text Knowledge Graph and Multi Hop Attention

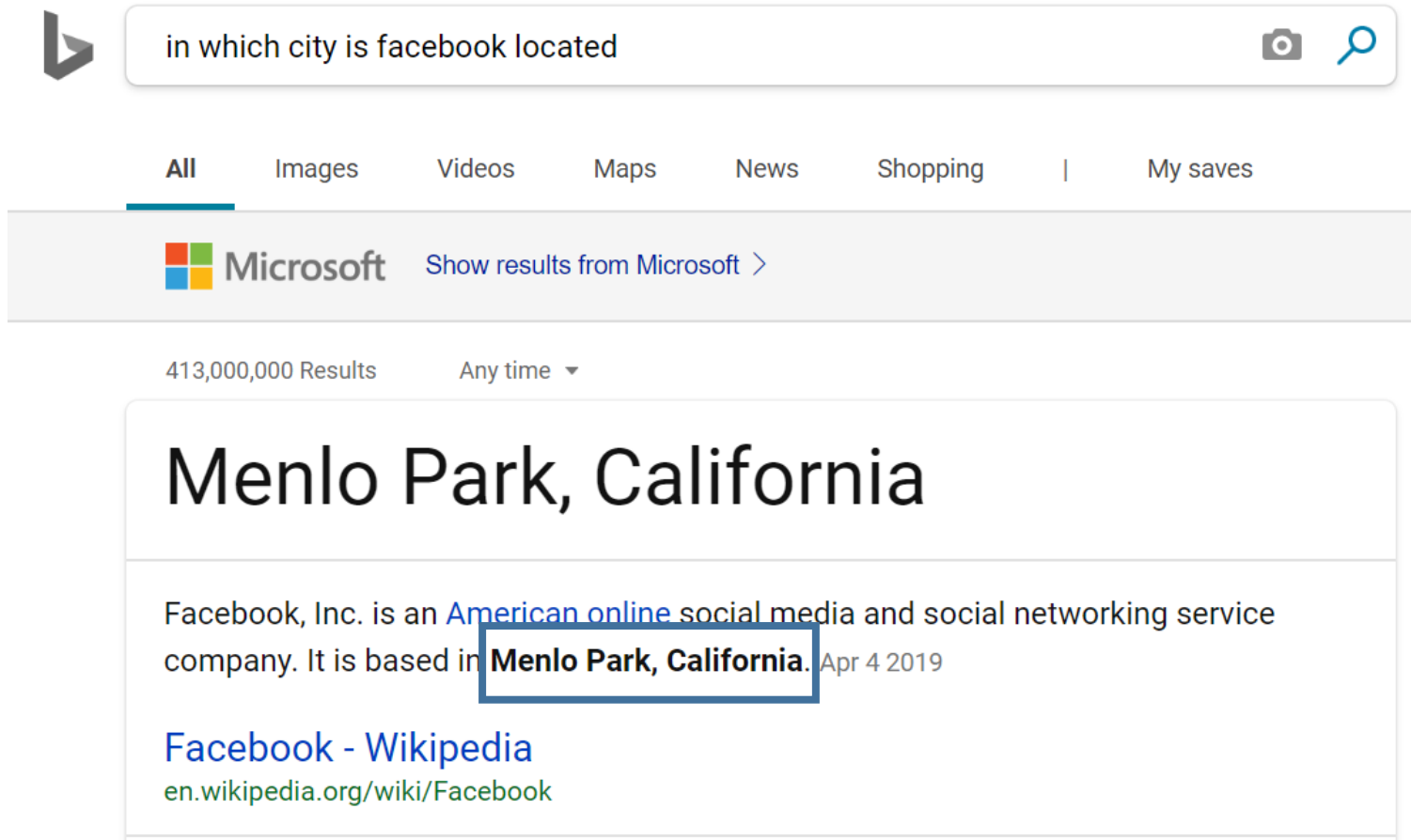
Chen Zhao

University of Maryland

chenz@cs.umd.edu

Machine Reading Style Question Answering


Find the *span* in the text to answer natural language question



The image shows a search engine interface. At the top, there is a search bar with the text "in which city is facebook located" and icons for image search and search. Below the search bar are navigation tabs: "All", "Images", "Videos", "Maps", "News", "Shopping", and "My saves". The "All" tab is selected. Below the tabs is a Microsoft logo and a link "Show results from Microsoft >". Below that, it says "413,000,000 Results" and "Any time ▾". The main result is a snippet for "Menlo Park, California". The text of the snippet is: "Facebook, Inc. is an [American online](#) social media and social networking service company. It is based in **Menlo Park, California**. Apr 4 2019". The words "Menlo Park, California" are highlighted with a blue box. Below the snippet is a link "Facebook - Wikipedia" and the URL "en.wikipedia.org/wiki/Facebook".

in which city is facebook located

All Images Videos Maps News Shopping | My saves

 Microsoft [Show results from Microsoft >](#)

413,000,000 Results Any time ▾

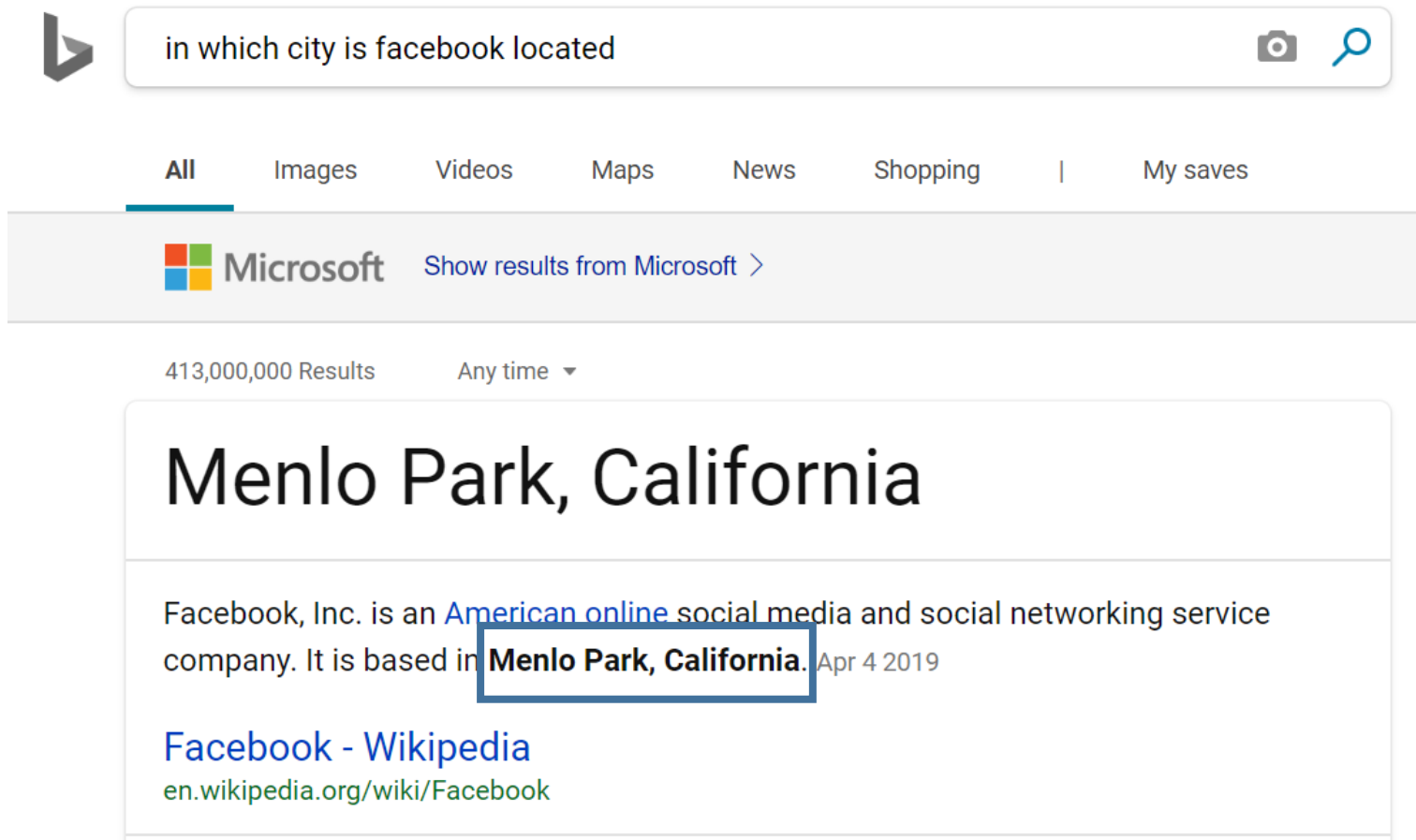
Menlo Park, California

Facebook, Inc. is an [American online](#) social media and social networking service company. It is based in **Menlo Park, California**. Apr 4 2019

[Facebook - Wikipedia](#)
en.wikipedia.org/wiki/Facebook

Machine Reading Style Question Answering

Find the *span* in the text to answer natural language question



The screenshot shows a search engine interface. At the top, a search bar contains the query "in which city is facebook located". Below the search bar, there are tabs for "All", "Images", "Videos", "Maps", "News", "Shopping", and "My saves". A "Microsoft" logo and a link to "Show results from Microsoft" are visible. The search results show "413,000,000 Results" and "Any time" filter. The main result is titled "Menlo Park, California". Below the title, the text reads: "Facebook, Inc. is an [American online](#) social media and social networking service company. It is based in **Menlo Park, California**. Apr 4 2019". A blue box highlights the text "Menlo Park, California". Below this, there is a link to "Facebook - Wikipedia" with the URL "en.wikipedia.org/wiki/Facebook".

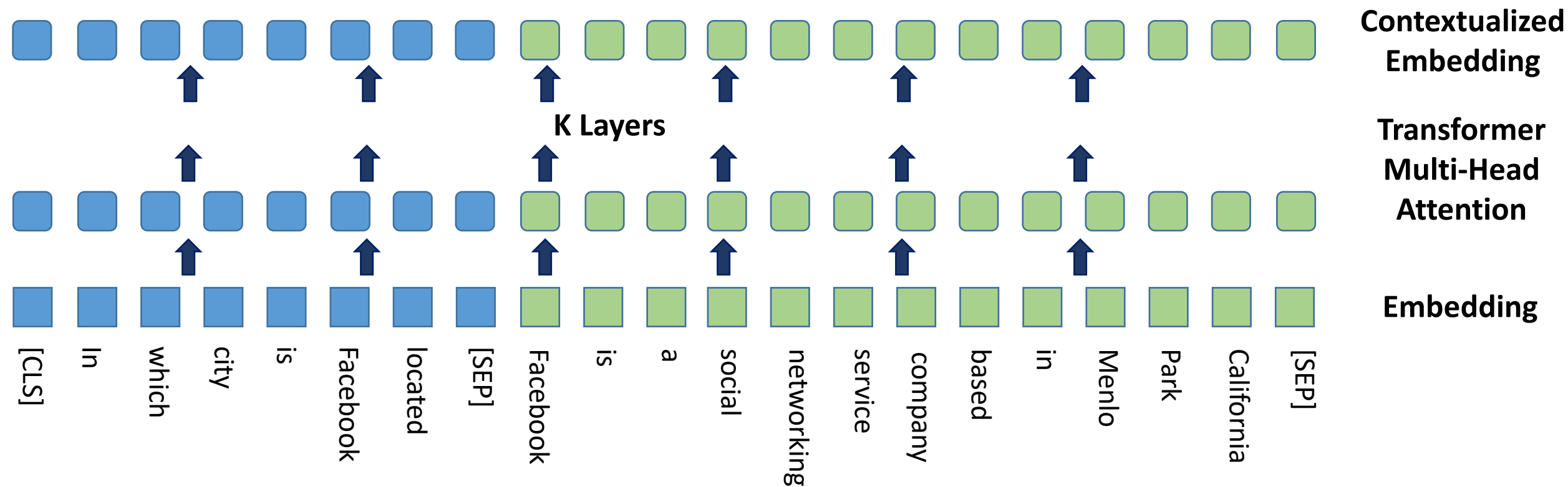


MS MARCO

Natural
Questions

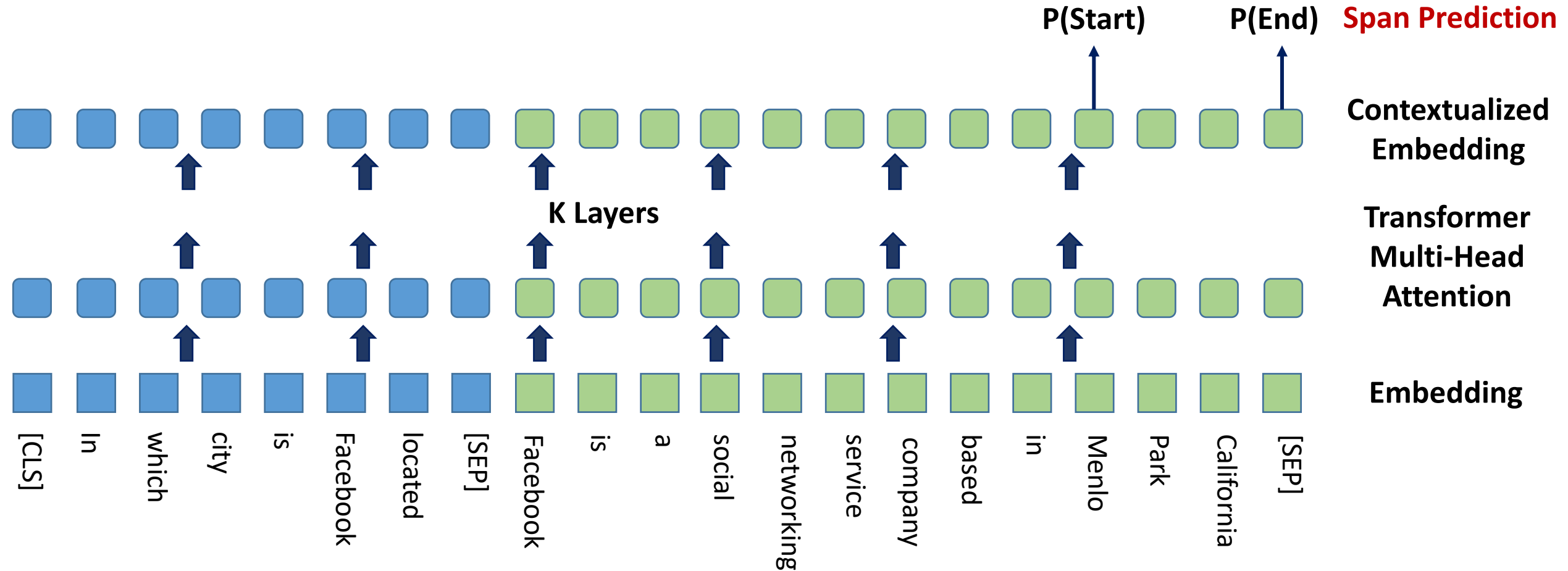
Transformers in Modeling Natural Language

- Input sentence(s) in sequential form
- (New) standard solutions with pre-trained Transformers (e.g., BERT)



Transformer in Machine Reading

- (New) standard solutions with pre-trained Transformers (e.g., BERT)



Multi-Evidence Question Answering

Questions require multiple evidence (in multiple documents) to answer

Vermeer painted a series of cityscapes of this Dutch city, including The Little Street. This city highly influenced the Dutch Golden Age in various aspects.

It is one of only three [Vermeer](#) paintings of views of Delft, the others being [View of Delft](#) and the now lost House Standing in Delft.



Historically, Delft played a highly influential role in the [Dutch Golden Age](#).



Delft

Multi-Evidence Question Answering

Questions require multiple evidence (in multiple documents) to answer



Zuckerberg built a website called "Facemash" in 2003 while attending Harvard University. The site

Harvard University is a private Ivy League research university in **Cambridge,** Massachusetts, with about 6,700

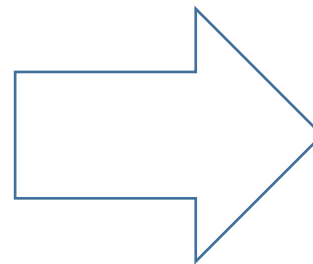
Outline

- DELFT: Multi-evidence QA with a Free-text Knowledge Graph (WWW 2020)
- Transformer-XH: Multi-evidence Reasoning with eXtra Hop Attention (ICLR 2020)

Entity Centric Questions

- Trivia Games (Quizbowl, Jeopardy!)
- Question is entity rich, a multi-sentence description
- Answer is an entity
- Open domain setting

Vermeer painted a series of cityscapes of this *Dutch* city, including *The Little Street*. This city highly influenced the *Dutch Golden Age* in various aspects.



Delft

Knowledge Graph Question Answering (KGQA)

- Search the answer from a Knowledge Graph



Knowledge Graph Question Answering (KGQA)

- Search the answer from a Knowledge Graph
- Can answer complex questions, Example:
 - Q: Who was the [last] [emperor of] China?
 - A: (:?person, Emperor of, China, max:time) → Puyi



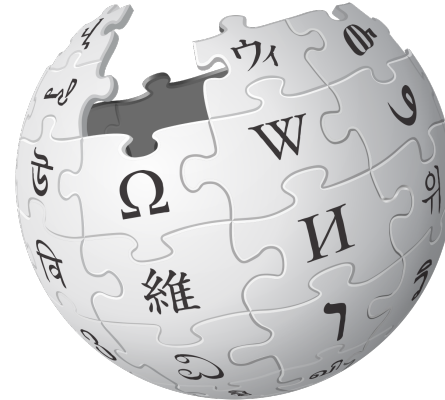
Knowledge Graph Question Answering (KGQA)

- Search the answer from a Knowledge Graph
- Can answer complex questions, Example:
 - Q: Who was the [last] [emperor of] China?
 - A: (?:person, Emperor of, China, max:time) → Puyi
- Knowledge Graph is brittle and incomplete
 - Google Vault 570M entities, 14 times than Freebase
 - 35K relations, similar to Freebase



Machine Reading

- Extract paragraph(s) from Wikipedia, then select a span
- Current Machine Reading approaches only focus on ***single evidence***
 - Question is complex: ***synthesize multiple evidence***



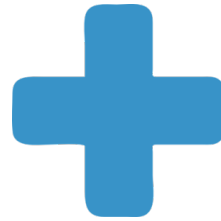
SQuAD2.0
The Stanford Question Answering Dataset

MS MARCO

Natural
Questions

DELFT: Combine the Best of Both World

- Deciphering Entity Links from Free Text



DELFT: Combine the Best of Both World

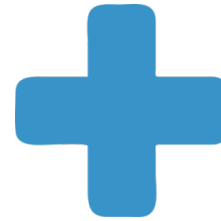
- Deciphering **Entity Links** from **Free Text**
- KGQA style structure reasoning + high Coverage free text

 Freebase™


WIKIDATA


DBpedia


yago
select knowledge



DELFT: Combine the Best of Both World

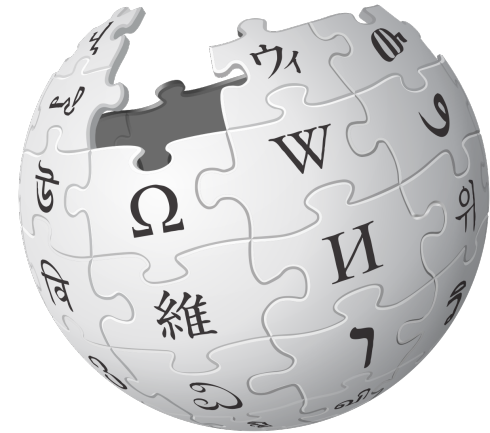
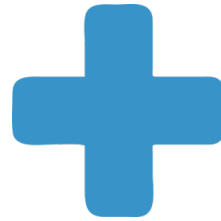
- Deciphering **Entity Links** from **Free Text**
- KGQA style structure reasoning + high Coverage free text
- Graph Construction + Graph Modeling

 Freebase™


WIKIDATA


DBpedia


yago
select knowledge



DELFT: Combine the Best of Both World

- Free-text Knowledge Graph Construction
- Question Grounding
- Graph Modeling

Free-Text Knowledge Graph Construction

- Build knowledge graph from Wikipedia
 - Nodes are Wikipedia entities



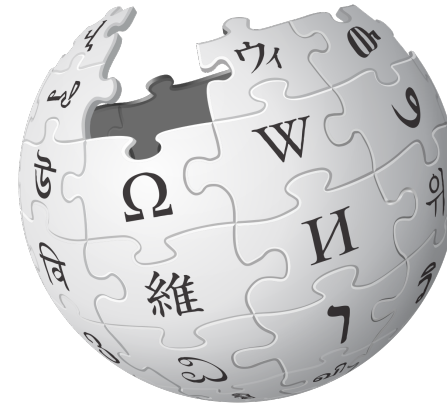
Free-Text Knowledge Graph Construction

- Build knowledge graph from Wikipedia
 - Nodes are Wikipedia entities
 - Free-text edges connect nodes
 - Sentences that mention pairs of entities



Free-Text Knowledge Graph Construction

- Build knowledge graph from Wikipedia
 - Nodes are Wikipedia entities
 - Free-text edges connect nodes
 - Sentences that mention pairs of entities
- High coverage, but noisy



Question Grounding – Question Entity Nodes

- Identify entities in the question by entity linking tools

Question: **Vermeer** painted a series of cityscapes of this Dutch city, including **The Little Street**. This city highly influenced the **Dutch Golden Age** in various aspects.

Answer: *Delft*



Vermeer



The
Little
Street



Dutch
Golden
Age

Question Grounding – Candidate Entity Nodes

- Candidate Entity Nodes directly connect to the entities related to the question

Question: **Vermeer** painted a series of cityscapes of this Dutch city, including **The Little Street**. This city highly influenced the **Dutch Golden Age** in various aspects.

Answer: *Delft*



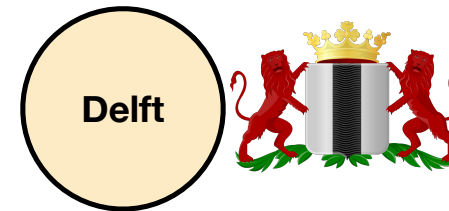
Vermeer



The
Little
Street



Dutch
Golden
Age



Question Grounding – Node Gloss

- Use gloss (First Wikipedia sentence) as the node features

Question: **Vermeer** painted a series of cityscapes of this Dutch city, including **The Little Street**. This city highly influenced the **Dutch Golden Age** in various aspects.

Answer: *Delft*

Vermeer was a Dutch Baroque Period painter who specialized in ...



Vermeer

The Little Street is a painting by the Dutch painter ...

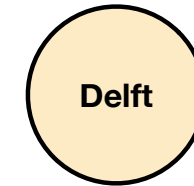


The Little Street

The **Dutch Golden Age** was a period in the history of the Netherlands ...



Dutch Golden Age



Delft is a city in the province of South Holland, Netherlands ...



Amsterdam is the Netherlands' capital ...

Question Grounding – Evidence Edges

- Evidence signals to find the correct answer

Question: **Vermeer** painted a series of cityscapes of this Dutch city, including **The Little Street**. This city highly influenced the **Dutch Golden Age** in various aspects.

Answer: *Delft*

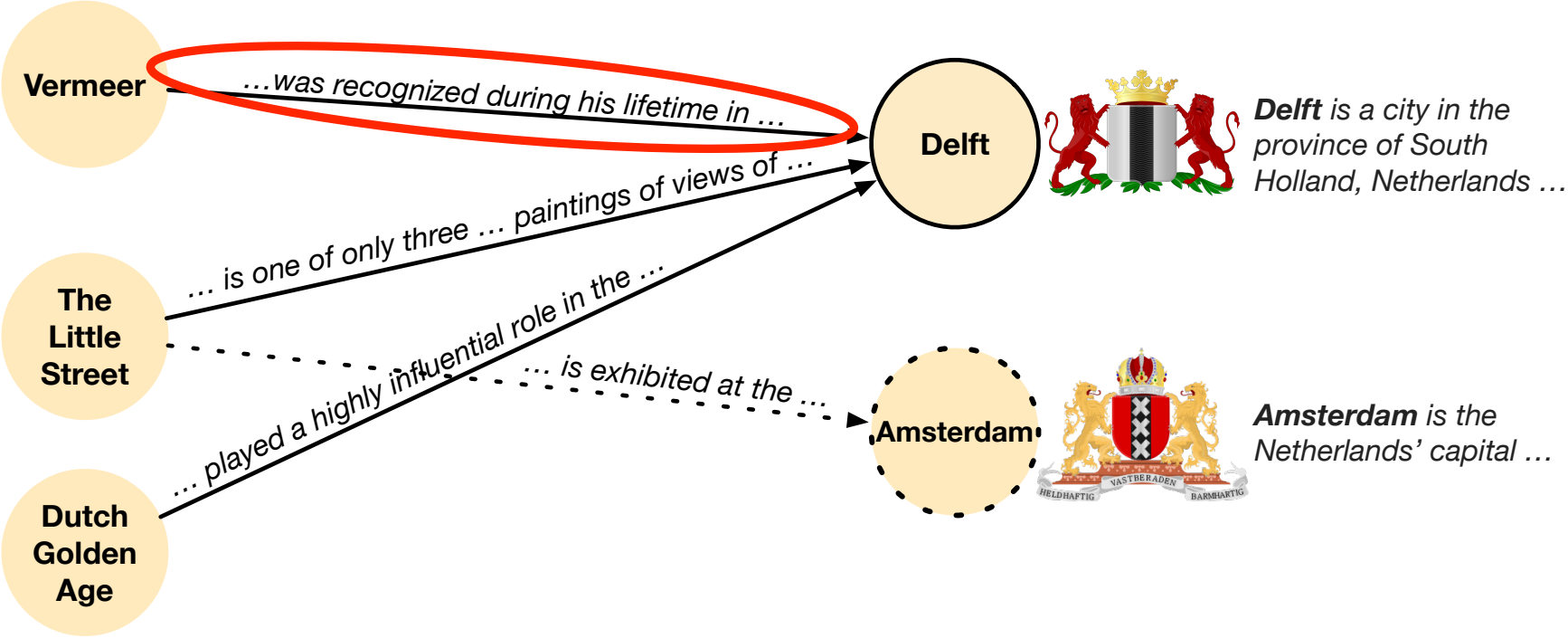
Vermeer was a Dutch Baroque Period painter who specialized in ...



The Little Street is a painting by the Dutch painter ...



The **Dutch Golden Age** was a period in the history of the Netherlands ...

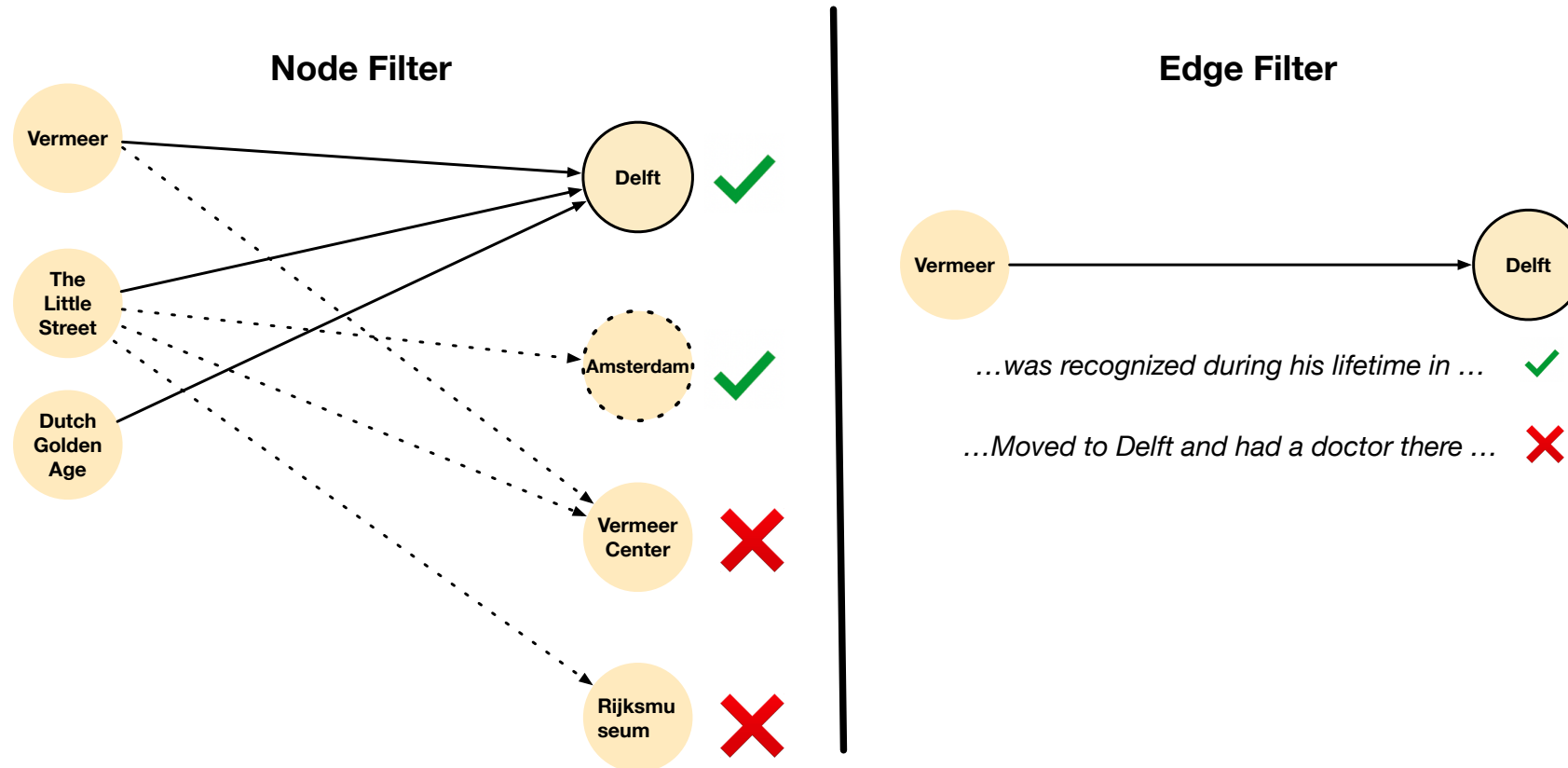


Delft is a city in the province of South Holland, Netherlands ...

Amsterdam is the Netherlands' capital ...

Graph Pruning

- Candidate Node Filter: Fine-tuned BERT Ranker
- Edge Sentence Filter: TFIDF similarity to question

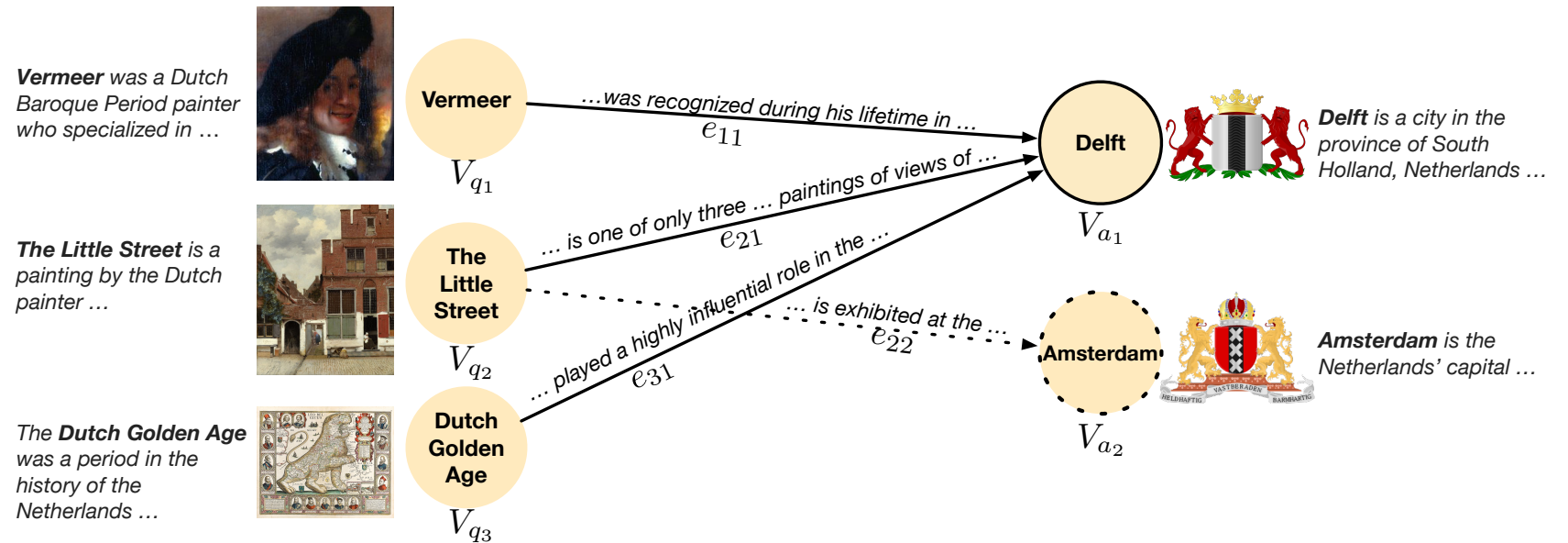


Graph Modeling

- Goal: find the correct answer node from the Candidate Entity Nodes
- Motivations for model design:
 - Graph Connectivity
 - Edge Relevance
 - Node Relevance

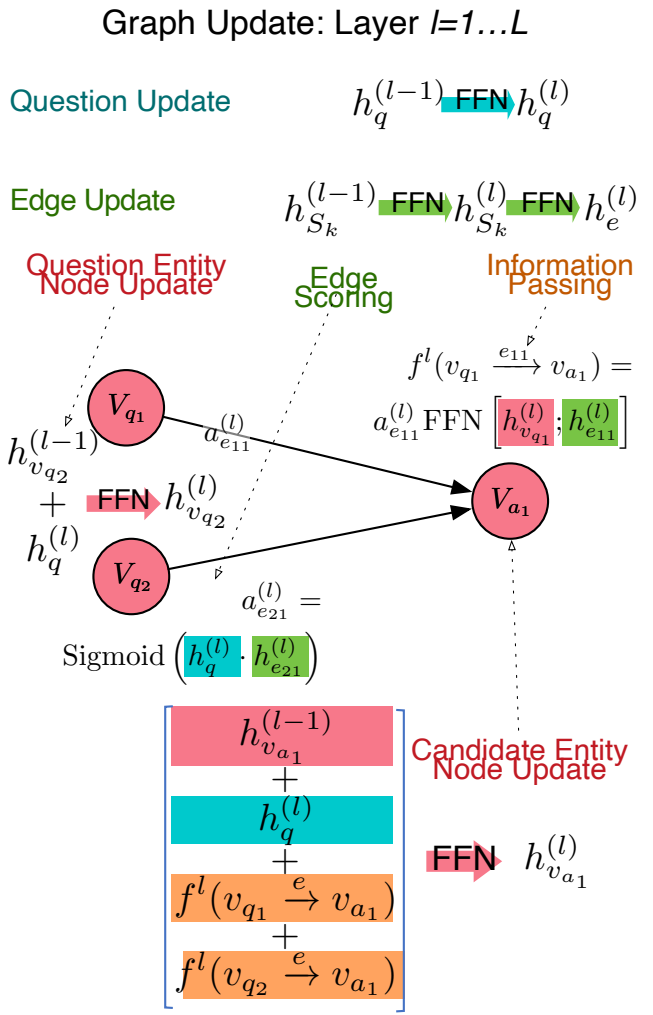
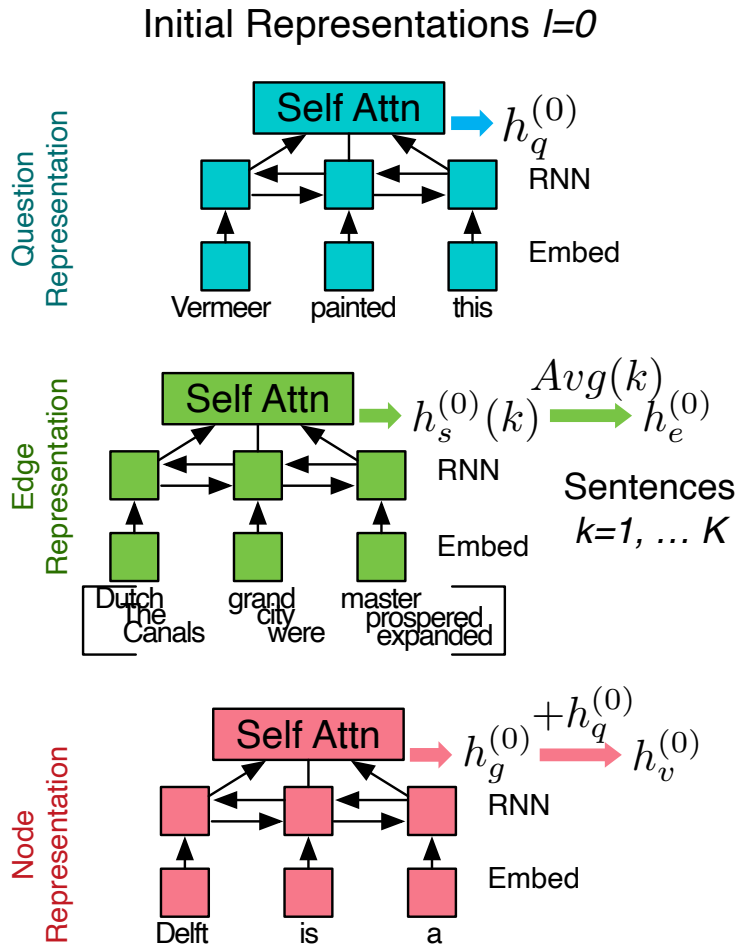
Question: **Vermeer** painted a series of cityscapes of this Dutch city, including **The Little Street**. This city highly influenced the **Dutch Golden Age** in various aspects.

Answer: *Delft*



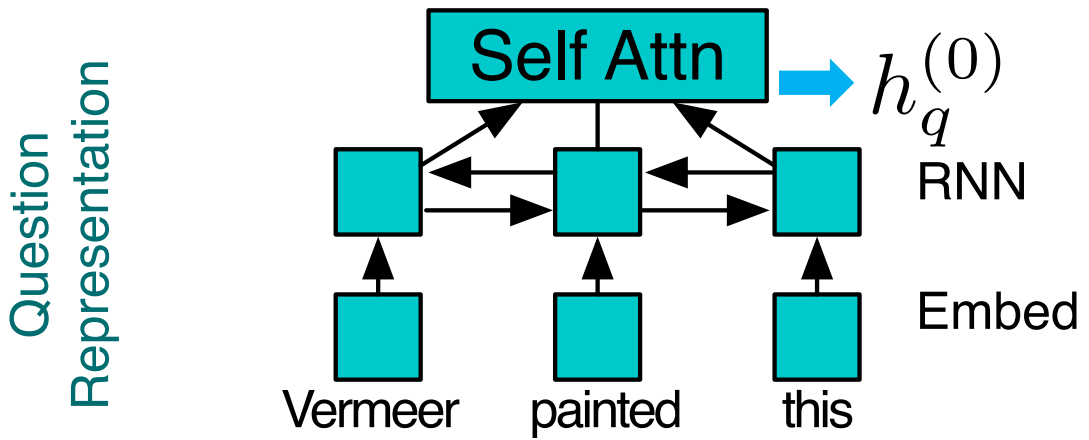
Graph Modeling – A Special Graph Neural Network

- Initial Representation
- Graph Update
- Node Scoring



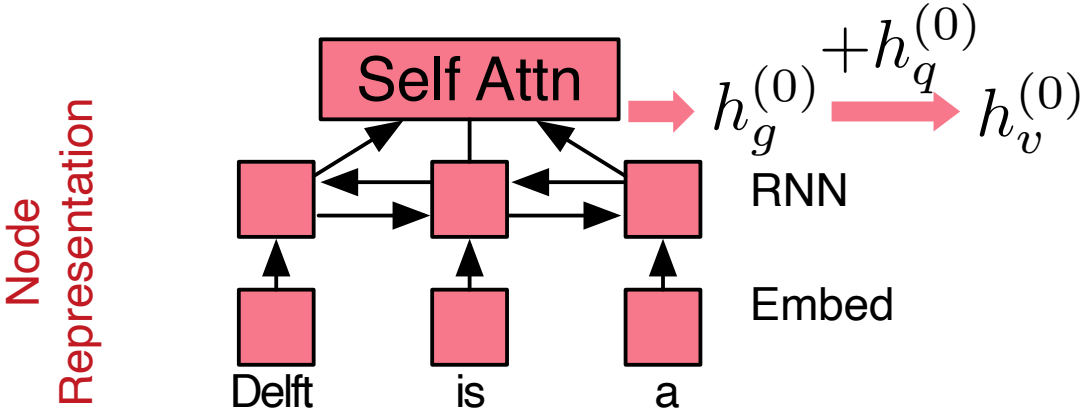
Graph Modeling - Initial Question Representation

Initial Representations $l=0$



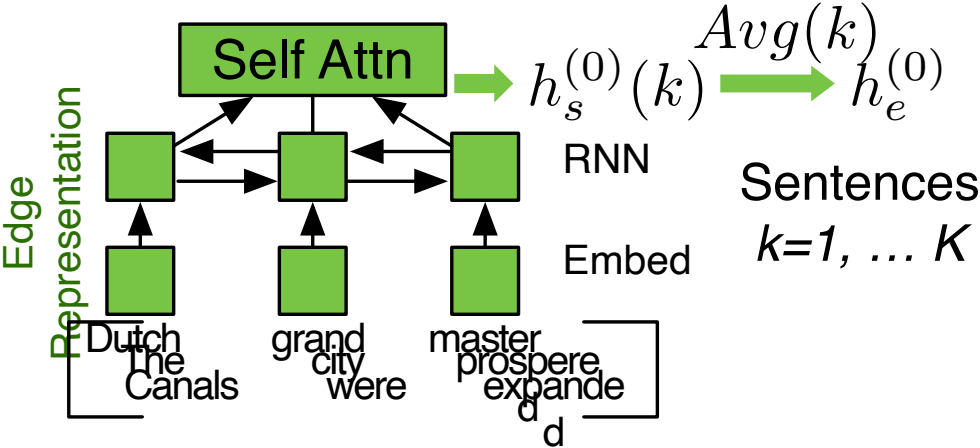
Graph Modeling - Initial Node Representation

Initial Representations $l=0$



Graph Modeling - Initial Edge Representation

Initial Representations $l=0$

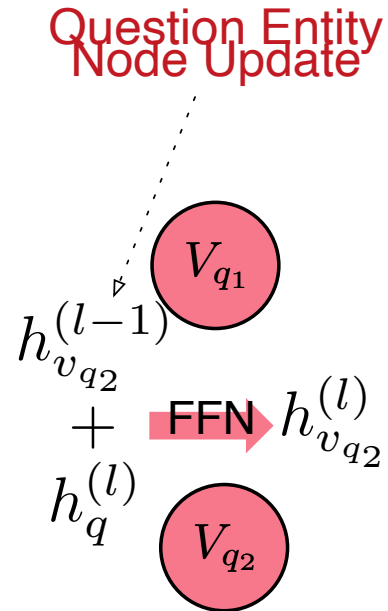


Graph Modeling - Representation Forwarding

Graph Update: Layer $l=1 \dots L$

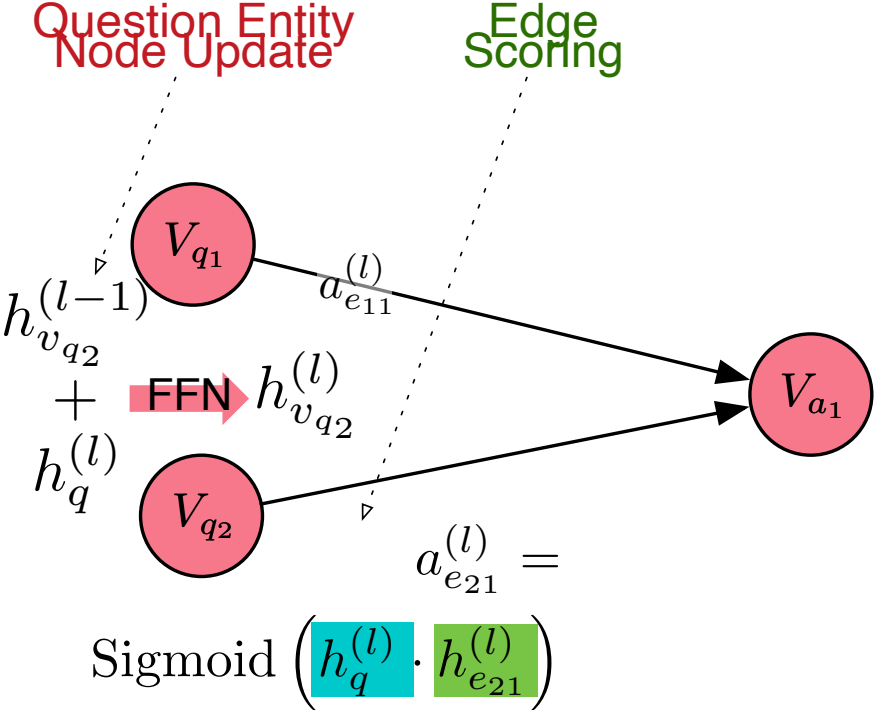
Question Update $h_q^{(l-1)} \xrightarrow{\text{FFN}} h_q^{(l)}$

Edge Update $h_{S_k}^{(l-1)} \xrightarrow{\text{FFN}} h_{S_k}^{(l)} \xrightarrow{\text{FFN}} h_e^{(l)}$



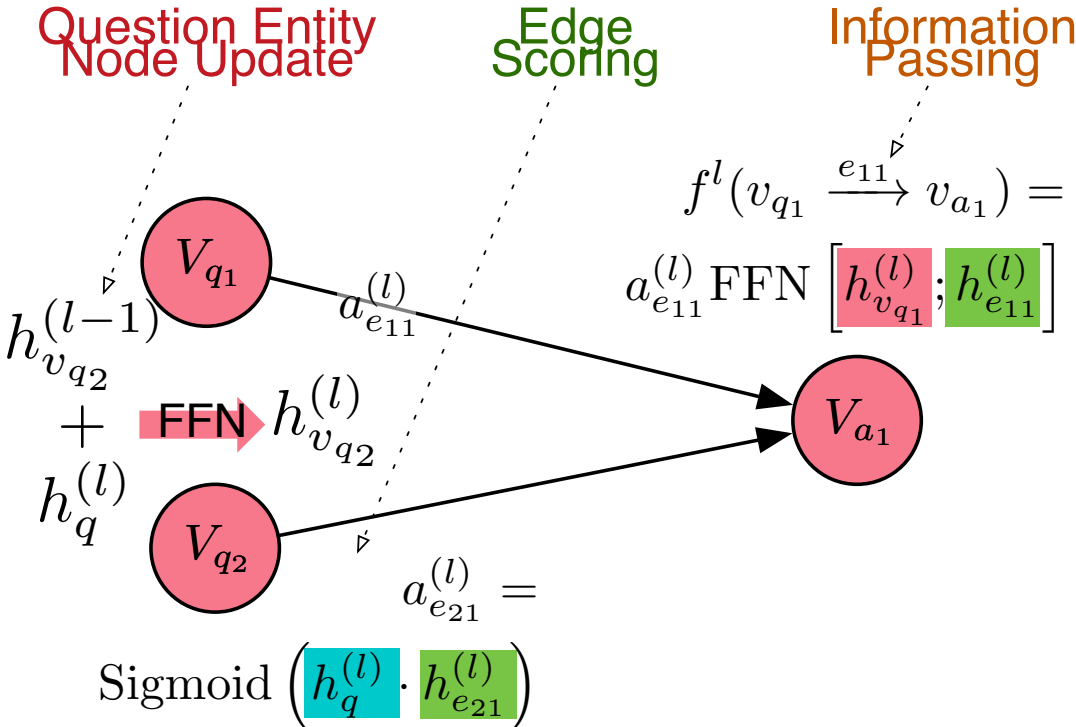
Graph Modeling – Edge Scoring

Graph Update: Layer $l=1 \dots L$



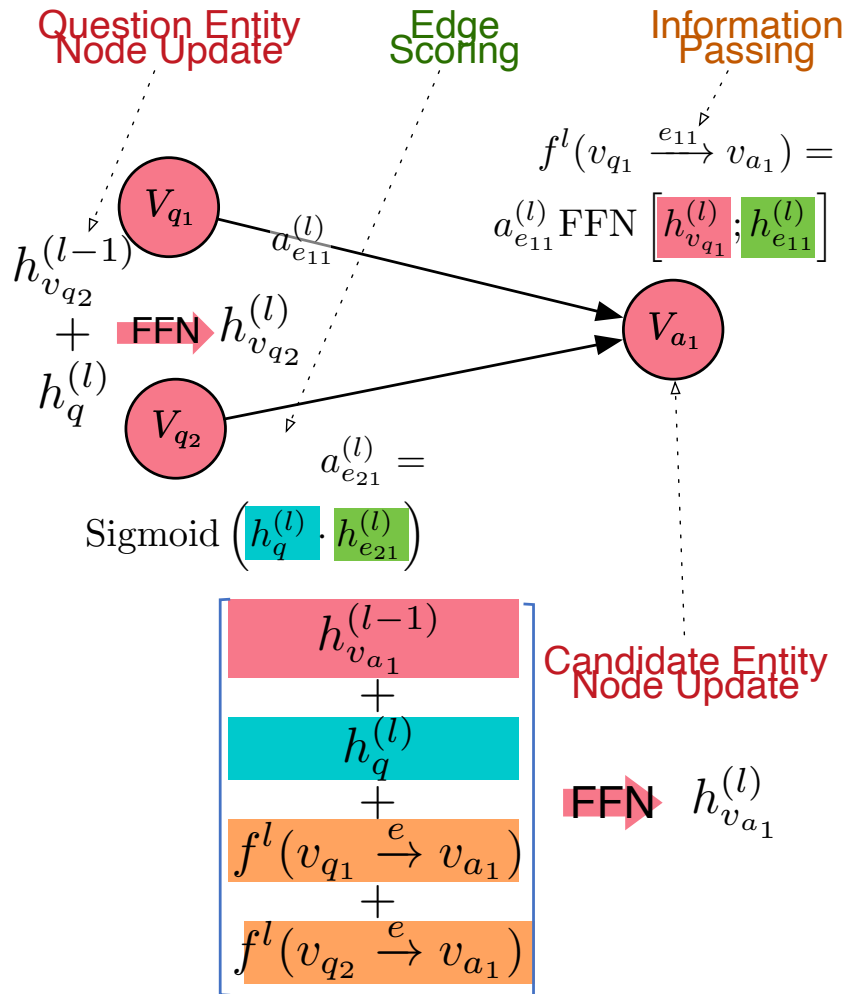
Graph Modeling – Information Passing

Graph Update: Layer $l=1 \dots L$



Graph Modeling – Candidate Nodes Update

Graph Update: Layer $l=1 \dots L$



DELFT Summary

- Graph Construction (Retrieval Step)
- Graph Modeling (Reasoning Step)

Experiments Datasets: QbLink

Example

Q: Name this person ridiculed for the film *Bedtime for Bonzo* by incumbent Pat Brown during an election which he won to become governor of California.

A: Ronald Reagan

Experiments Datasets: QANTA

Example

Q: Following one of these events, the Casa Pia children's education center was established. Trajan and Hadrian experienced one of these events in Antioch and suffered minor injuries. After one of these events, a city rallied around the phrase, "Bury the dead and heal the living." The aftermath of one of these events in 1923 saw the massacre of ethnic Koreans. That one of these events on the Kanto plain of Japan was exacerbated by the ensuing tsunami, a common consequence of these events. For 10 points, name this type of natural disaster that has often affected San Francisco due to the activity of the San Andreas fault.

A: Earthquakes

Experiments Datasets: TriviaQA

- Answerable by Wikipedia entities

Example

Q: What was the occupation of Lovely Rita according to the song by the Beatles?

A: Traffic Warden

	qbLink	QANTA	TriviaQA
Training	42219	31489	41448
Dev	3276	2211	4620
Test	5984	4089	5970
# Tokens	31.7 ± 9.4	129.2 ± 32.0	16.5 ± 8.6
# Entities	6.8 ± 2.4	21.2 ± 7.3	2.2 ± 1.3
% 1-3 Entities	9.6%	0	86.9%
% 4-6 Entities	36.7%	0	13.1%
% 7-9 Entities	36.5%	0	0
% 10+ Entities	17.1%	100%	0

Graph Coverage

- DELFT's graph is dense
- The graph separates the correct answer by its structure

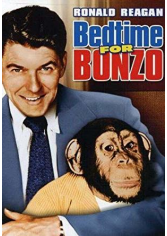
	qbLink	QANTA	TriviaQA
Answer Recall after Filtering	87.6%	83.9%	86.4%
Answer Recall within Two Hops along DBpedia Graph*	38%	–	–
# Edges to Correct Answer Node (+)	5.07 ± 2.17	12.33 ± 5.59	1.87 ± 1.12
# Edges to Candidate Entity Node (-)	2.35 ± 0.99	4.41 ± 2.02	1.21 ± 0.35
# Evidence Sentences per Edge (+)	12.3 ± 11.1	8.83 ± 6.17	15.53 ± 17.52
# Evidence Sentences per Edge (-)	4.67 ± 3.14	4.48 ± 1.88	3.96 ± 3.33

Baselines

- Machine Reading (DrQA)
- BERT Ranker
- BERT Memory Network

Graph Visualization

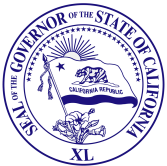
Q: Name **this person** ridiculed for the film Bedtime for Bonzo by incumbent Pat Brown during an election which he won to become governor of California.



Bedtime for Bonzo



Pat Brown



Governor of California

He co-starred in films like Bedtime for Bonzo. [0.96]

The election was between Pat Brown and Ronald Reagan. [0.08]

He served as the 33rd Governor of California. [0.06]

Jerry Brown is the son of Pat Brown. [0.05]

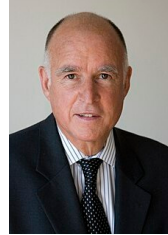
Jerry Brown served as the 34th and 39th governor of California. [0.04]

Ronald Reagan



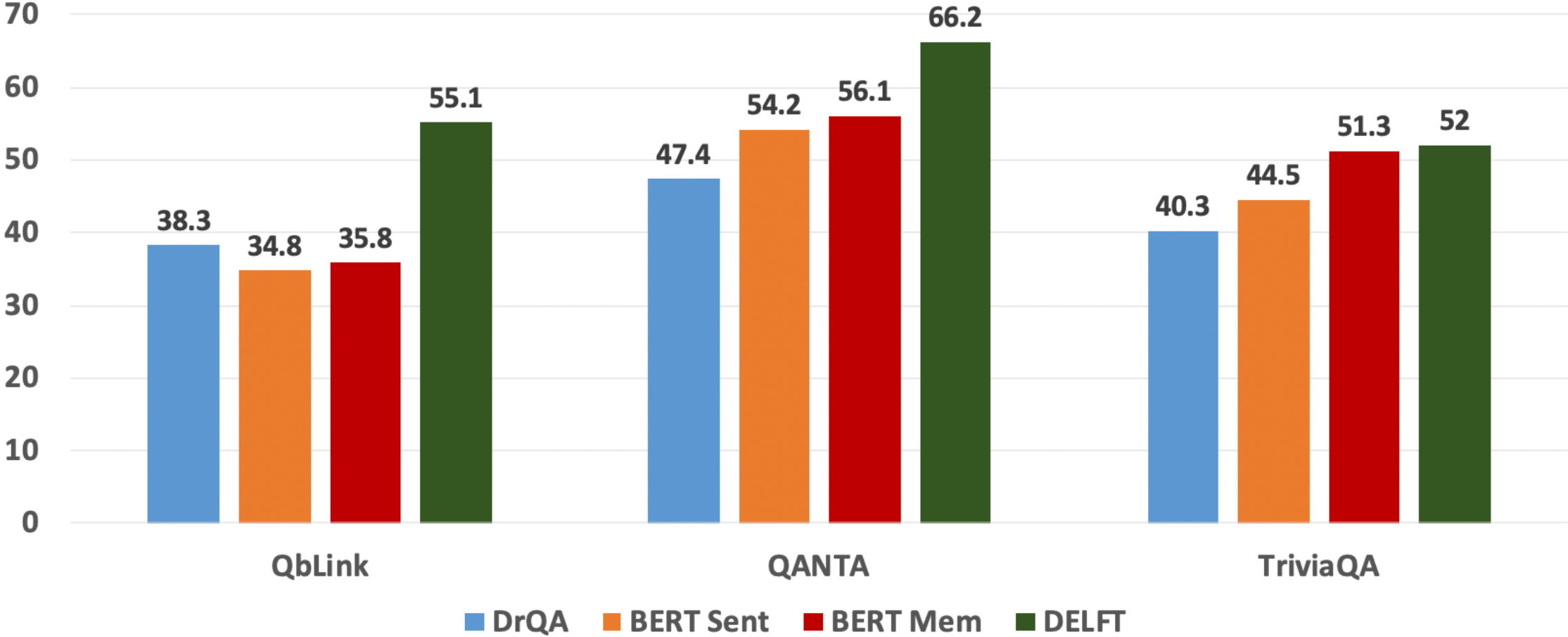
Score: 0.73 ✓

Jerry Brown



Score: 0.08 ✗

Overall Results

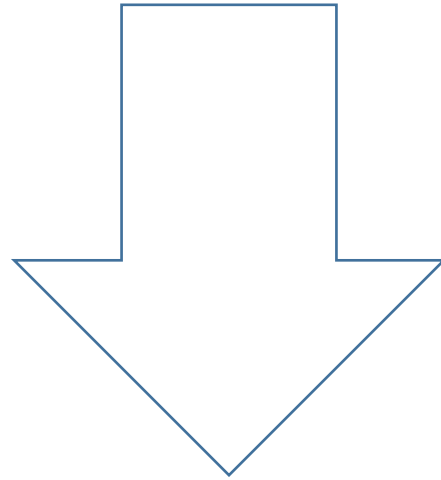


Outline

- DELFT: Multi-evidence QA with Free-text Knowledge Graph (WWW 2020)
- Transformer-XH: Multi-evidence Reasoning with eXtra Hop Attention (ICLR 2020)

DELFT works well, but ...

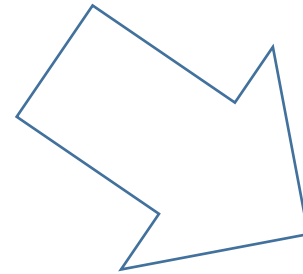
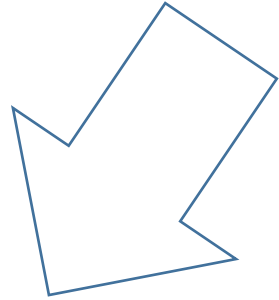
- Answer must be an entity
- Focus on single hop questions



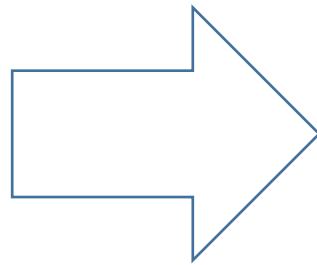
Transformer-XH: A more general model

Structured Text Sequence

Facebook was founded by [Mark Zuckerberg](#), along with fellow [Harvard College](#) students and roommates.



Zuckerberg built a website called "Facemash" in 2003 while attending [Harvard University](#). The site



Harvard University is a private [Ivy League research university](#) in [Cambridge, Massachusetts](#), with about 6,800 undergraduate

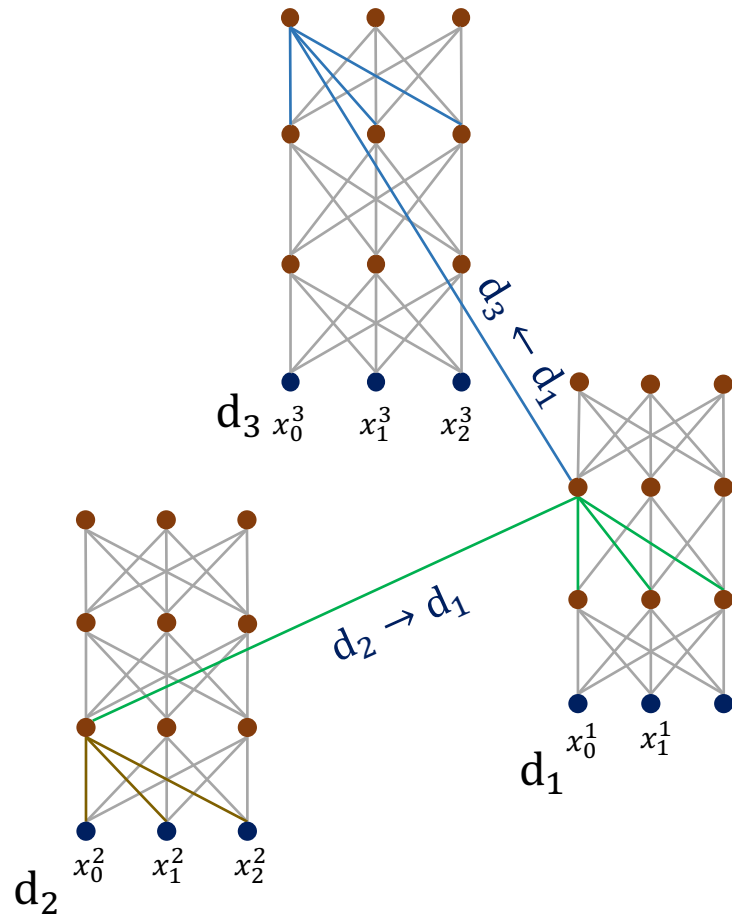
Transformer-XH

- For structured text:
 - Transformer with eXtra Hop attentions
 - Global representations of **multiple connected** text pieces
- Strong performance on different multi-evidence reasoning tasks
 - Multi-hop QA (Hotpot QA)
 - Multi-evidence Fact Verification (FEVER)

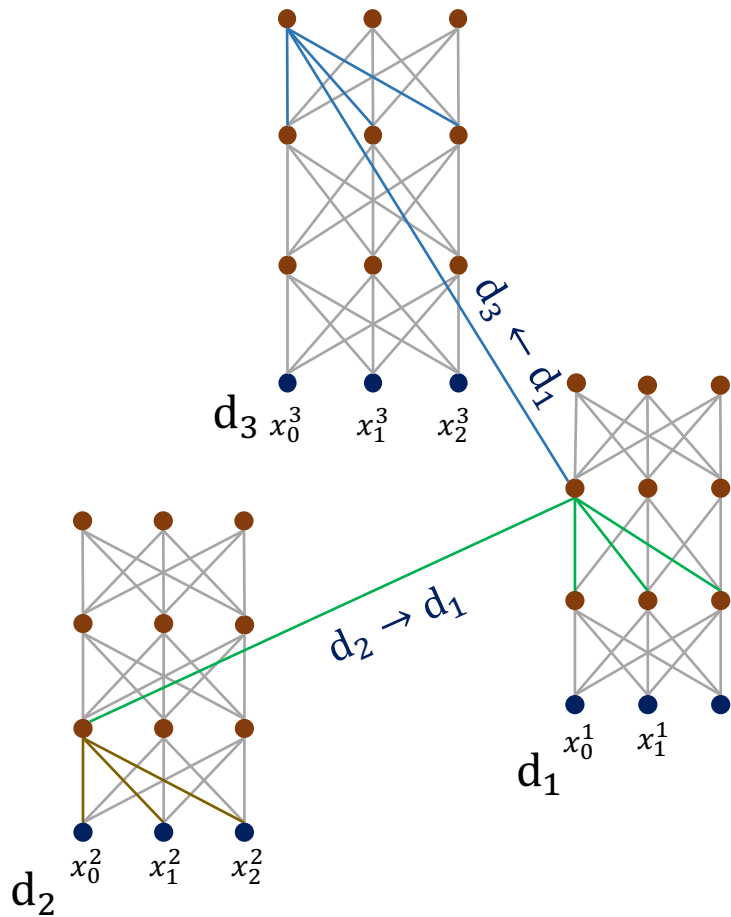
Transformer-XH: In Sequence Attention

In sequence (τ) attention in layer l , token i :

$$h_{\tau,i}^l = \sum_j \text{Softmax}_j \left(\frac{q_{\tau,i}^T \cdot k_{\tau,j}}{\sqrt{d_k}} \right) \cdot v_{\tau,j}$$



Transformer-XH: Extra Hop Attention



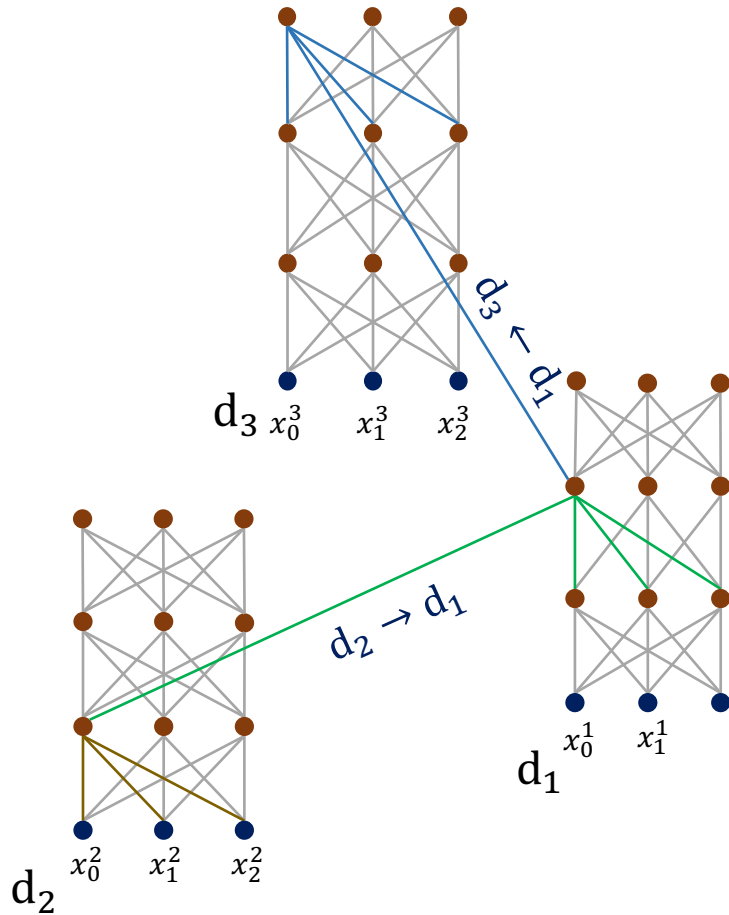
In sequence (τ) attention in layer l , token i :

$$h_{\tau,i}^l = \sum_j \text{Softmax}_j \left(\frac{q_{\tau,i}^T \cdot k_{\tau,j}}{\sqrt{d_k}} \right) \cdot v_{\tau,j}$$

Extra hop attention between sequences ($\eta \rightarrow \tau$):

$$\hat{h}_{\tau,0}^l = \sum_{\eta; e_{\tau\eta}=1} \text{Softmax}_{\eta} \left(\frac{\hat{q}_{\tau,0}^T \cdot \hat{k}_{\eta,0}}{\sqrt{d_k}} \right) \cdot \hat{v}_{\eta,0}$$

Transformer-XH: Layer Representation



In sequence (τ) attention in layer l , token i :

$$h_{\tau,i}^l = \sum_j \text{Softmax}_j \left(\frac{q_{\tau,i}^T \cdot k_{\tau,j}}{\sqrt{d_k}} \right) \cdot v_{\tau,j}$$

Extra hop attention between sequences ($\eta \rightarrow \tau$):

$$\hat{h}_{\tau,0}^l = \sum_{\eta; e_{\tau\eta}=1} \text{Softmax}_{\eta} \left(\frac{\hat{q}_{\tau,0}^T \cdot \hat{k}_{\eta,0}}{\sqrt{d_k}} \right) \cdot \hat{v}_{\eta,0}$$

Combine two attentions:

$$\begin{aligned} \tilde{h}_{\tau,0}^l &= \text{Linear}([h_{\tau,0}^l \circ \hat{h}_{\tau,0}^l]) \\ \tilde{h}_{\tau,l}^l &= h_{\tau,i}^l; \forall i \neq 0 \end{aligned}$$

Transformer-XH on Multi-Evidence Tasks

- Multi-hop QA
 - Multiple connected evidence is required
 - Evidence scatters in multiple documents
- Fact Verification
 - Claims often require multi-evidence to support

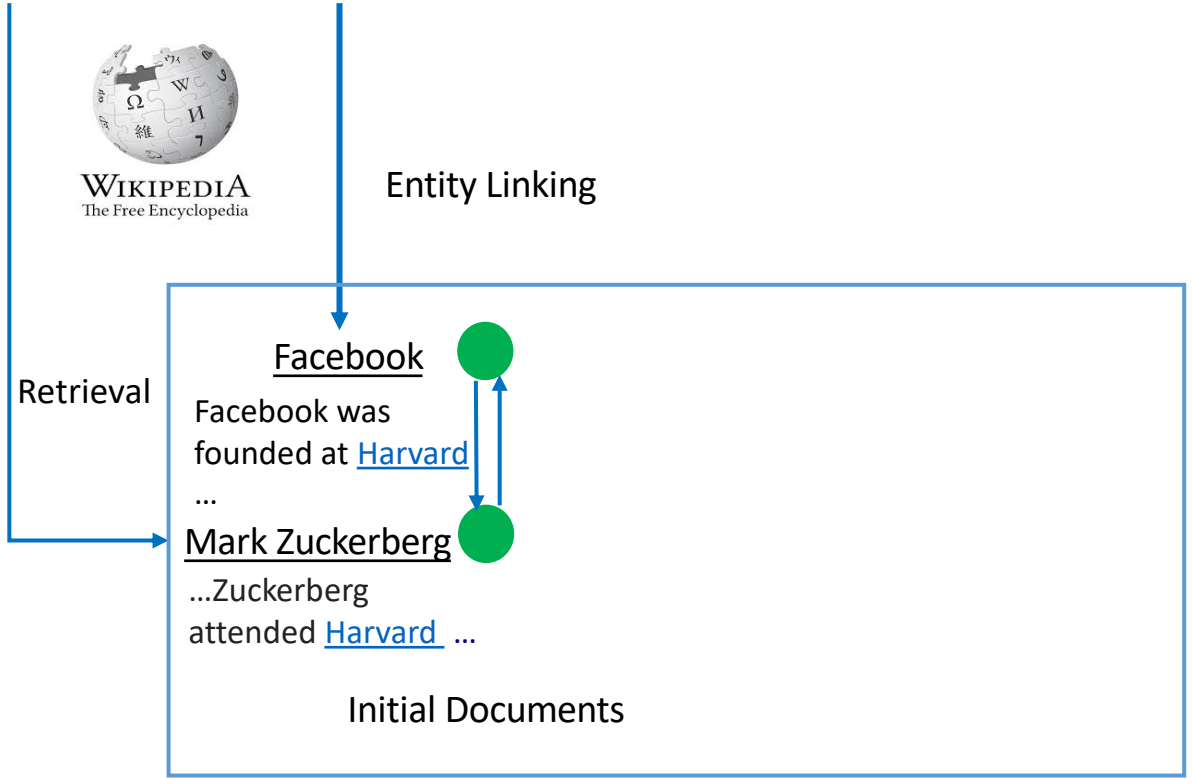
Transformer-XH on Multi-hop QA

- Graph Construction (Retrieval Step)
- Graph Modeling (Reasoning Step)

Multi-hop QA: Graph Construction

Input Question:

In which city was Facebook launched?

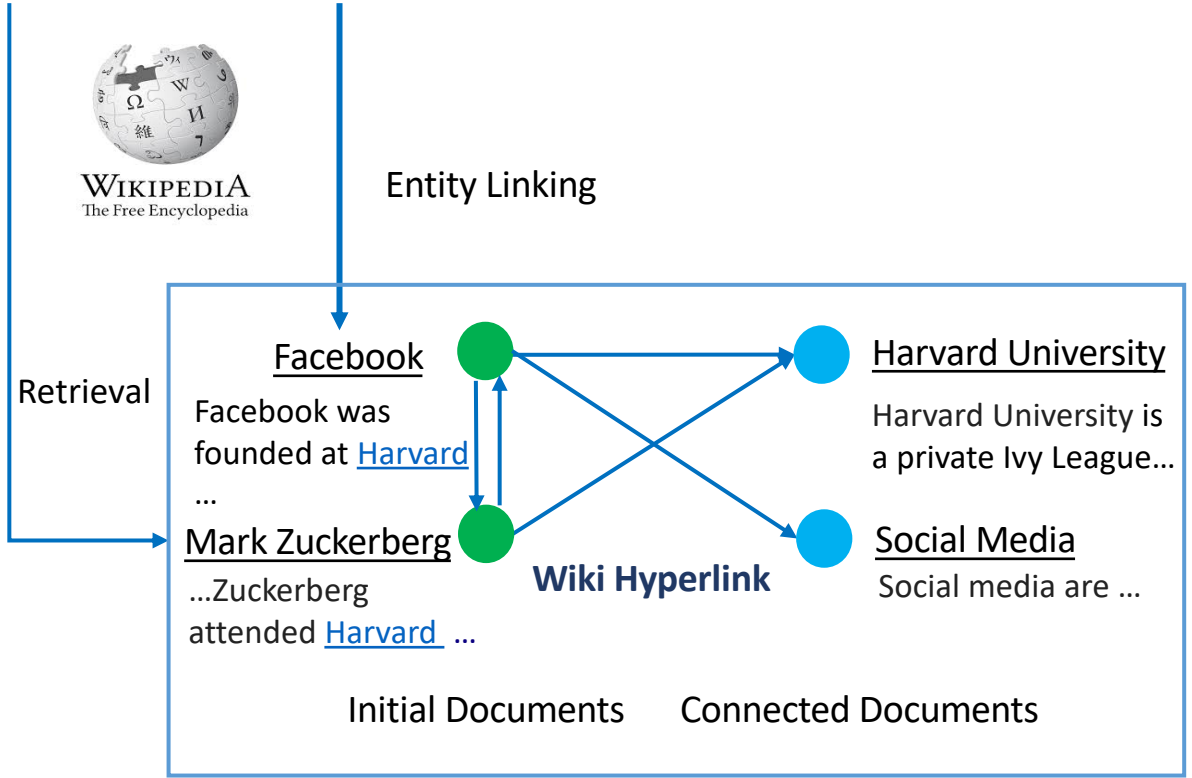


- Initial nodes

Multi-hop QA: Graph Construction

Input Question:

In which city was Facebook launched?

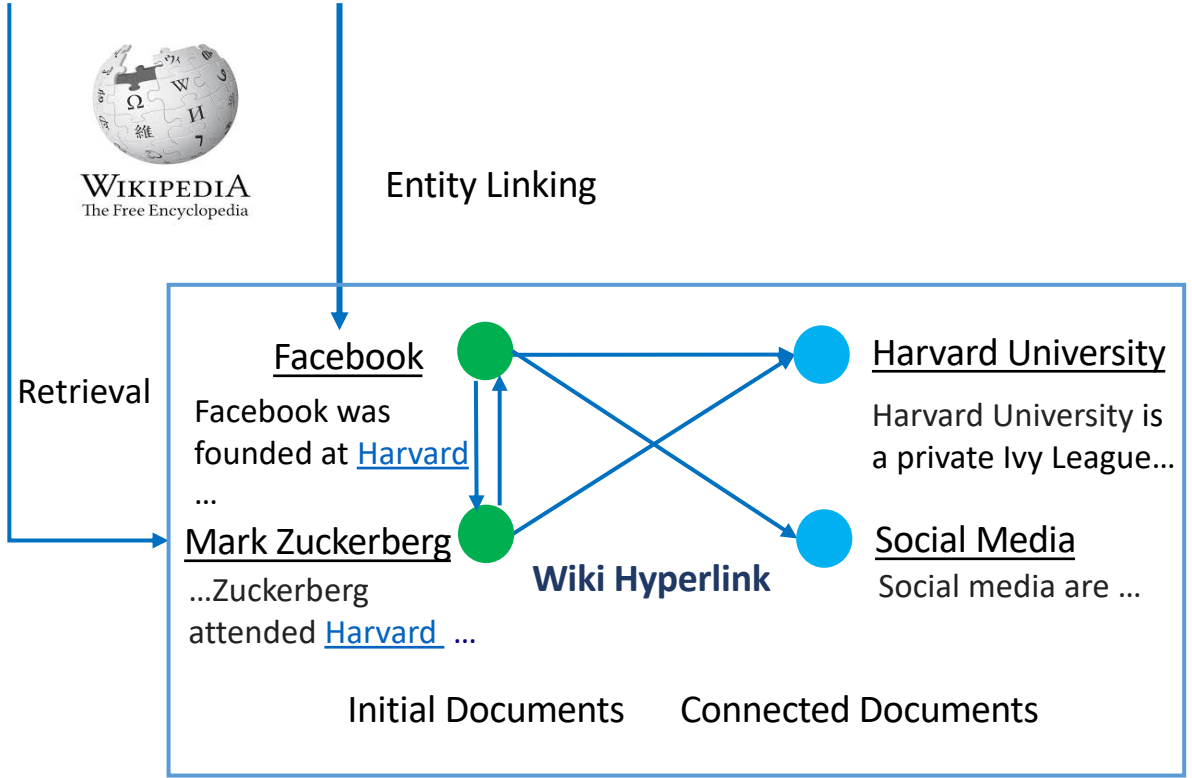


- Initial Nodes
- Node Expansion

Multi-hop QA: Graph Construction

Input Question:

In which city was Facebook launched?

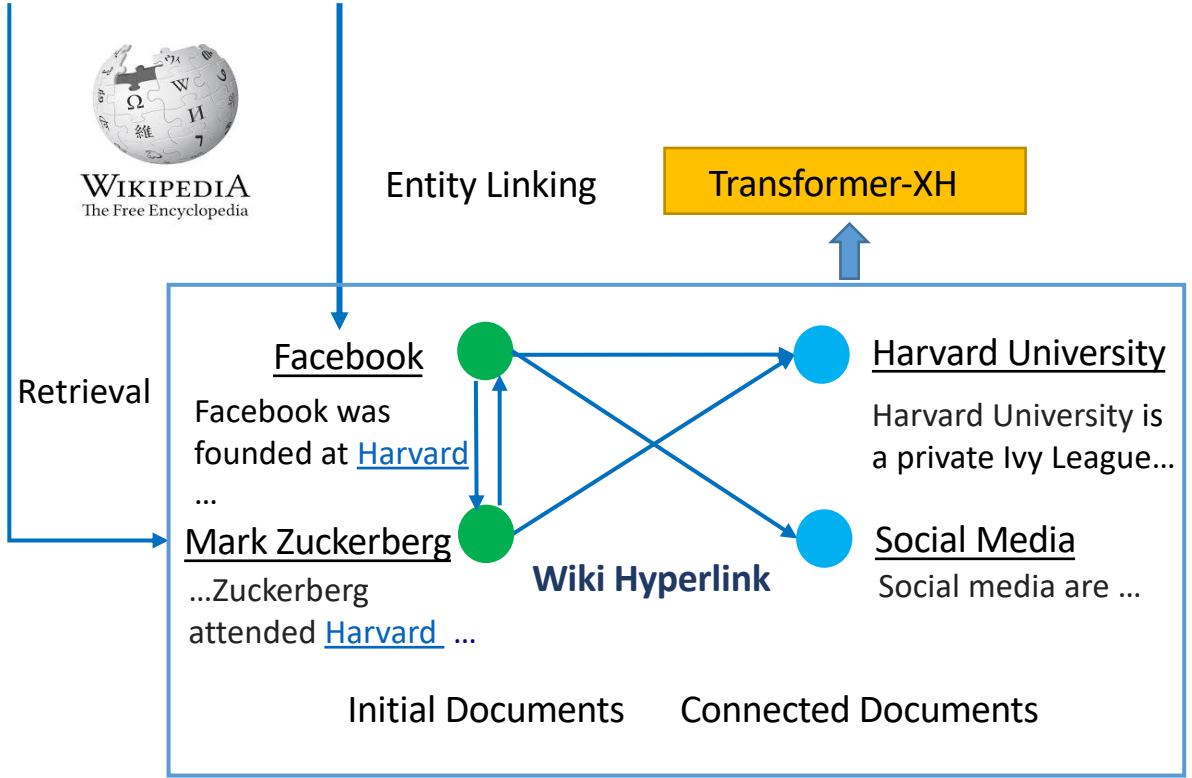


- Initial Nodes
- Node Expansion
- Node Filter

Multi-hop QA: Graph Modeling

Input Question:

In which city was Facebook launched?



- Transformer-XH

Multi-hop QA: Graph Modeling

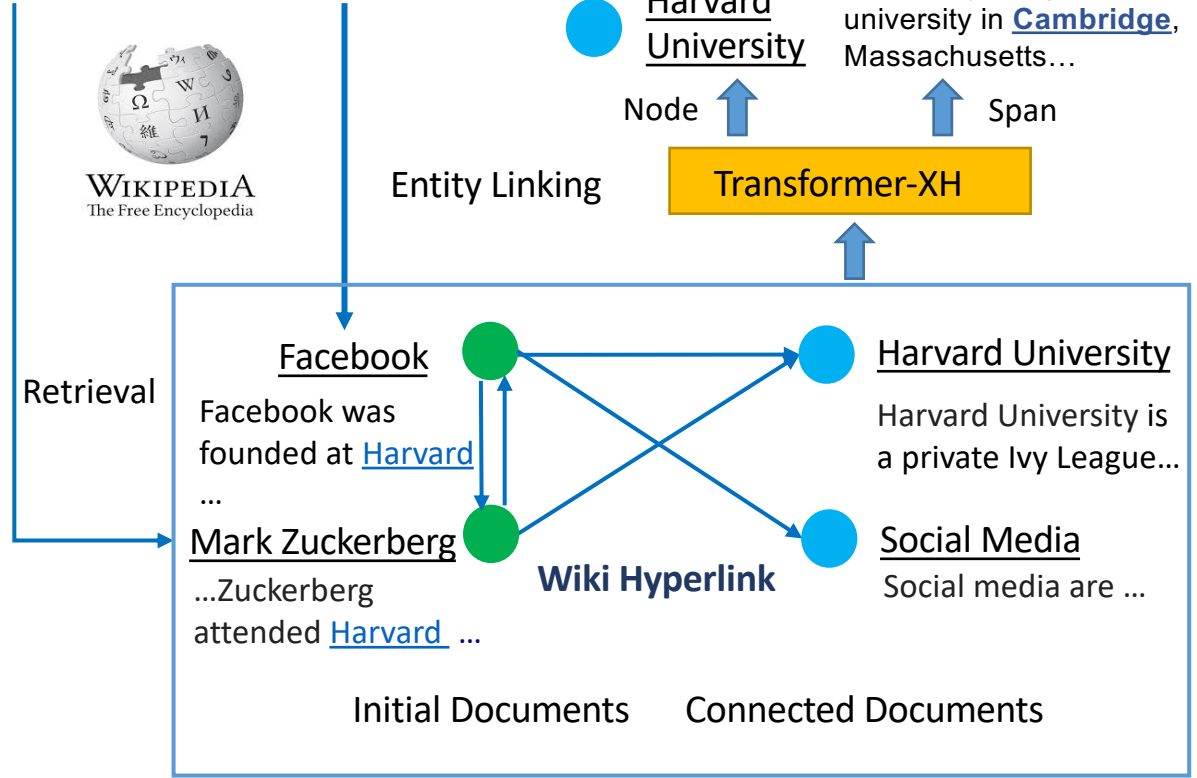
Input Question:

In which city was Facebook launched?



Answer:

Harvard University is a private Ivy League research university in Cambridge, Massachusetts...



- Transformer-XH
- Task specific layers (Multi-task)

Baselines

- Pipelined Approach with Single-hop BERT and Graph Neural Network

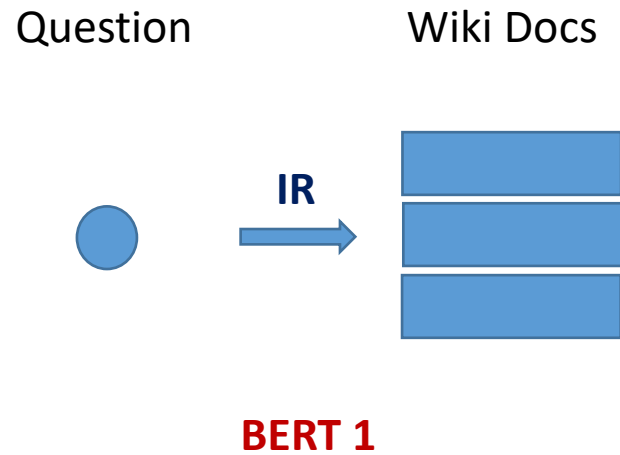
Question



- Cognitive QA (ACL 19). Published STOA.

Baselines

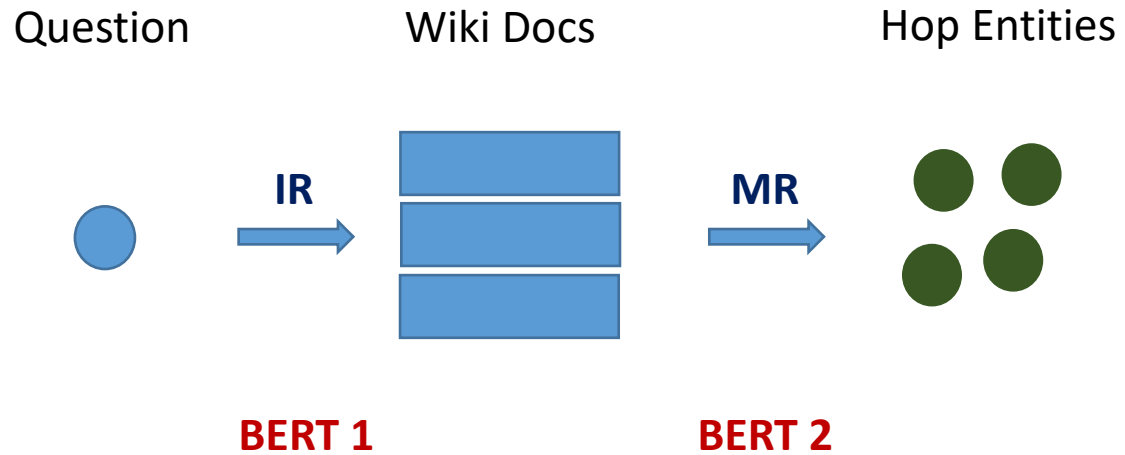
- Pipelined Approach with Single-hop BERT and Graph Neural Network



- Cognitive QA (ACL 19). Published STOA.

Baselines

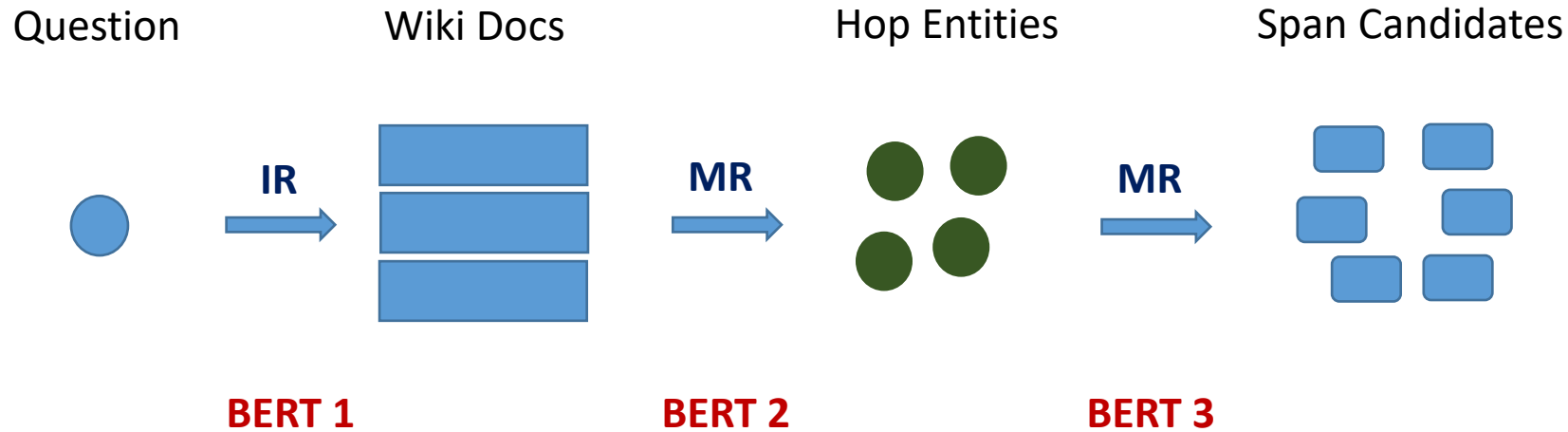
- Pipelined Approach with Single-hop BERT and Graph Neural Network



- Cognitive QA (ACL 19). Published STOA.

Baselines

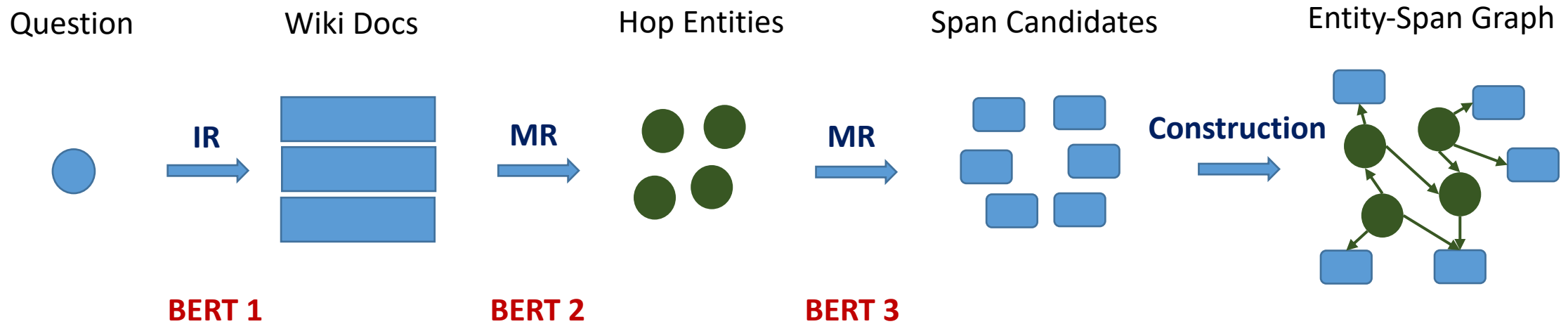
- Pipelined Approach with Single-hop BERT and Graph Neural Network



- Cognitive QA (ACL 19). Published STOA.

Baselines

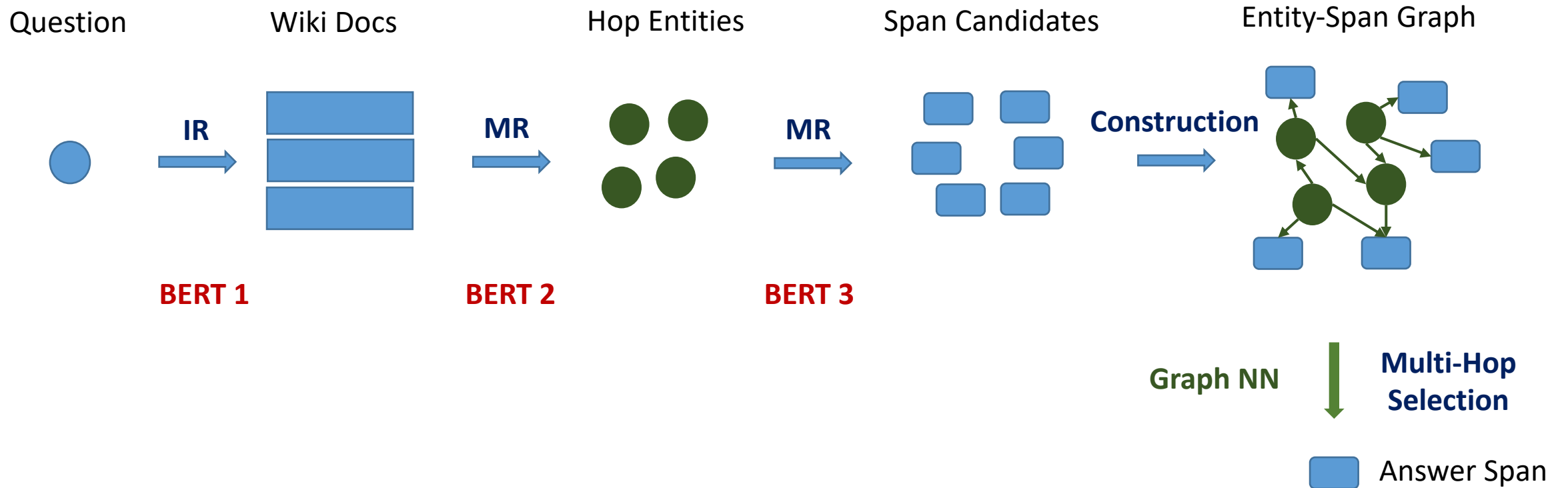
- Pipelined Approach with Single-hop BERT and Graph Neural Network



- Cognitive QA (ACL 19). Published STOA.

Baselines

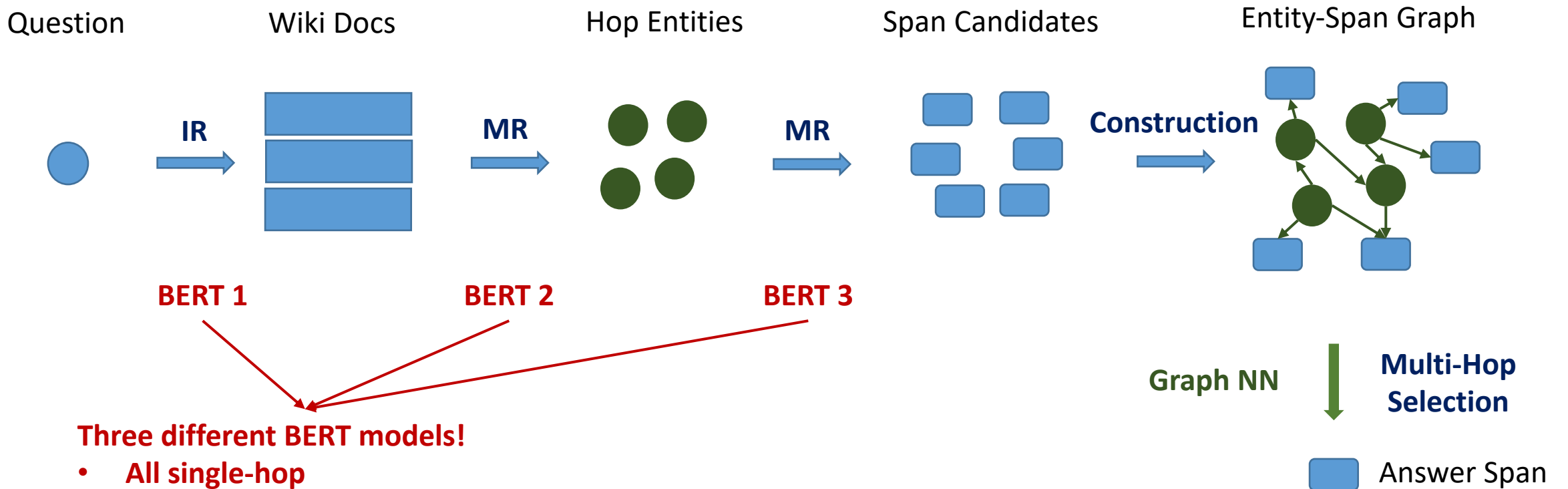
- Pipelined Approach with Single-hop BERT and Graph Neural Network



- Cognitive QA (ACL 19). Published STOA.

Baselines

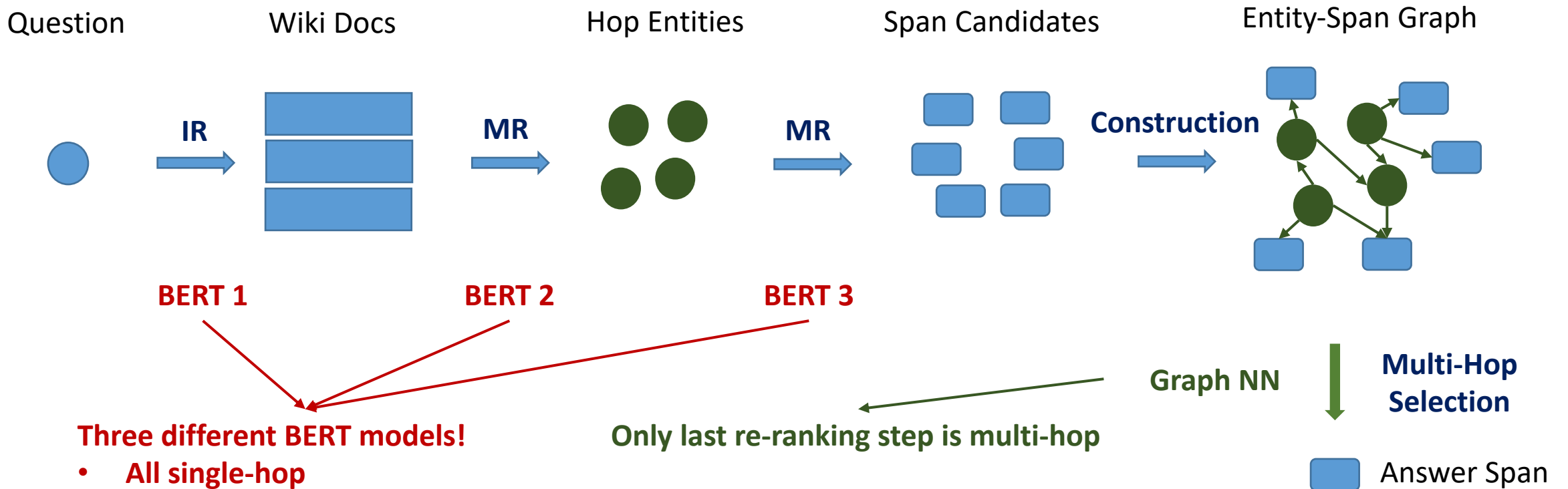
- Pipelined Approach with Single-hop BERT and Graph Neural Network



- Cognitive QA (ACL 19). Published STOA.

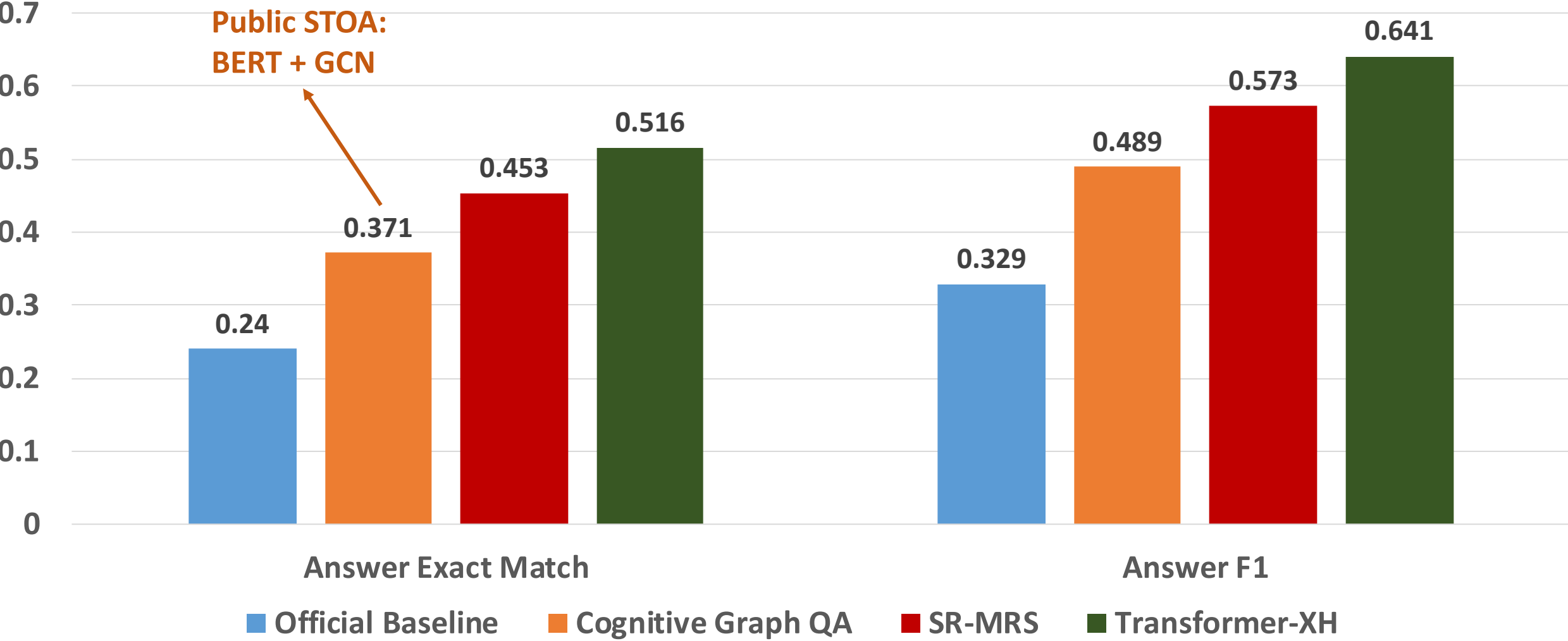
Baselines

- Pipelined Approach with Single-hop BERT and Graph Neural Network

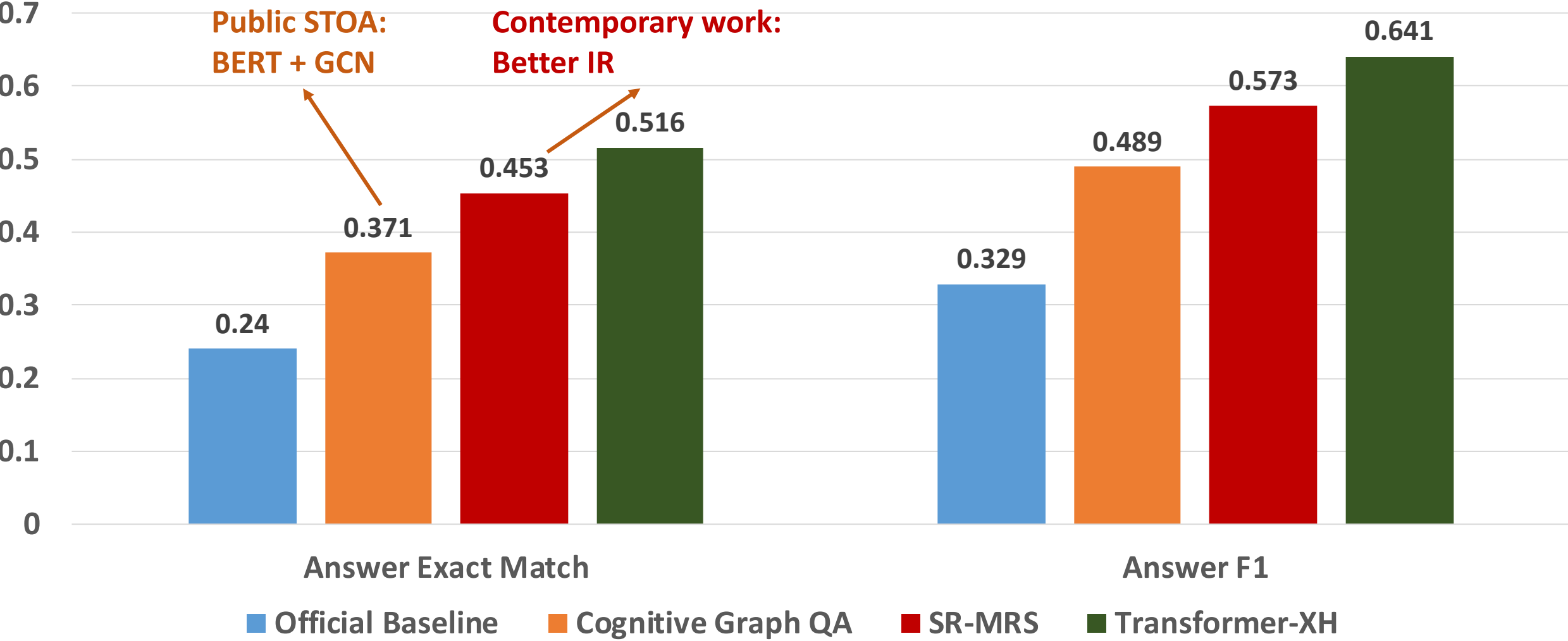


- Cognitive QA (ACL 19). Published STOA.

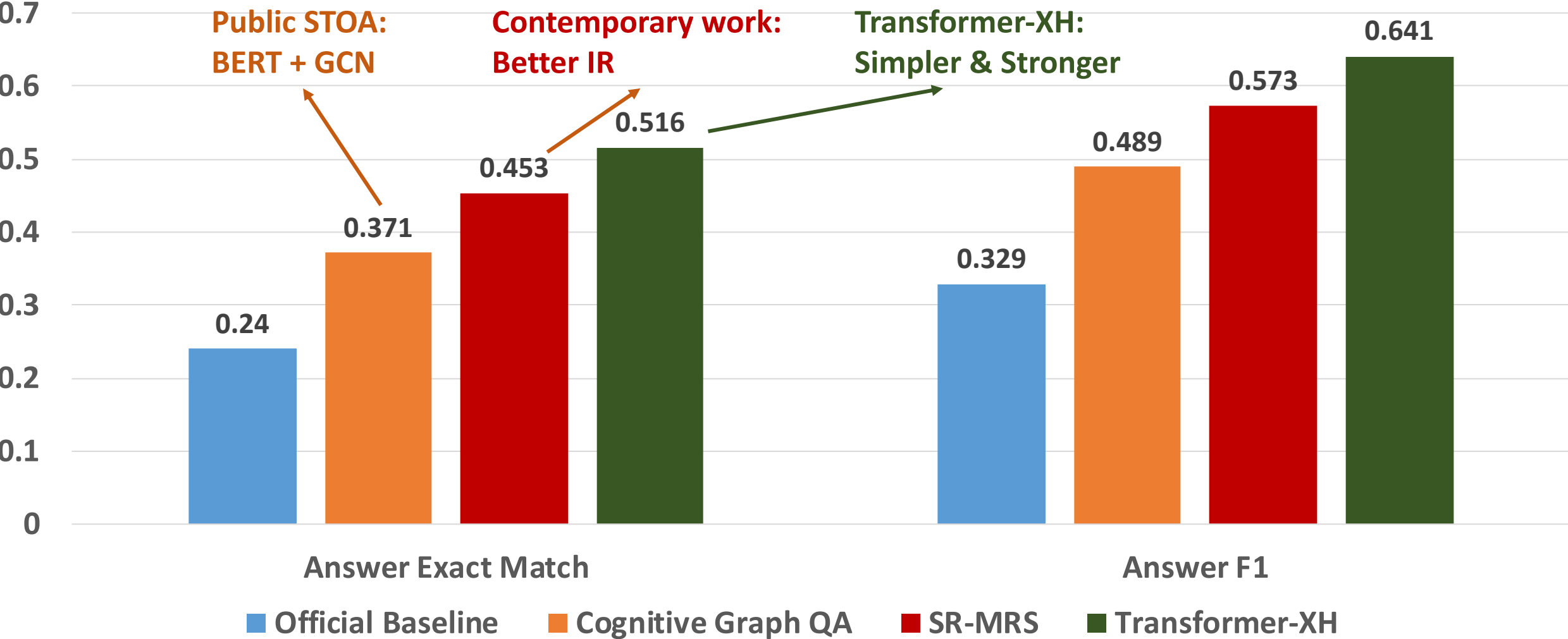
Results on Hotpot QA Full-Wiki Test Set



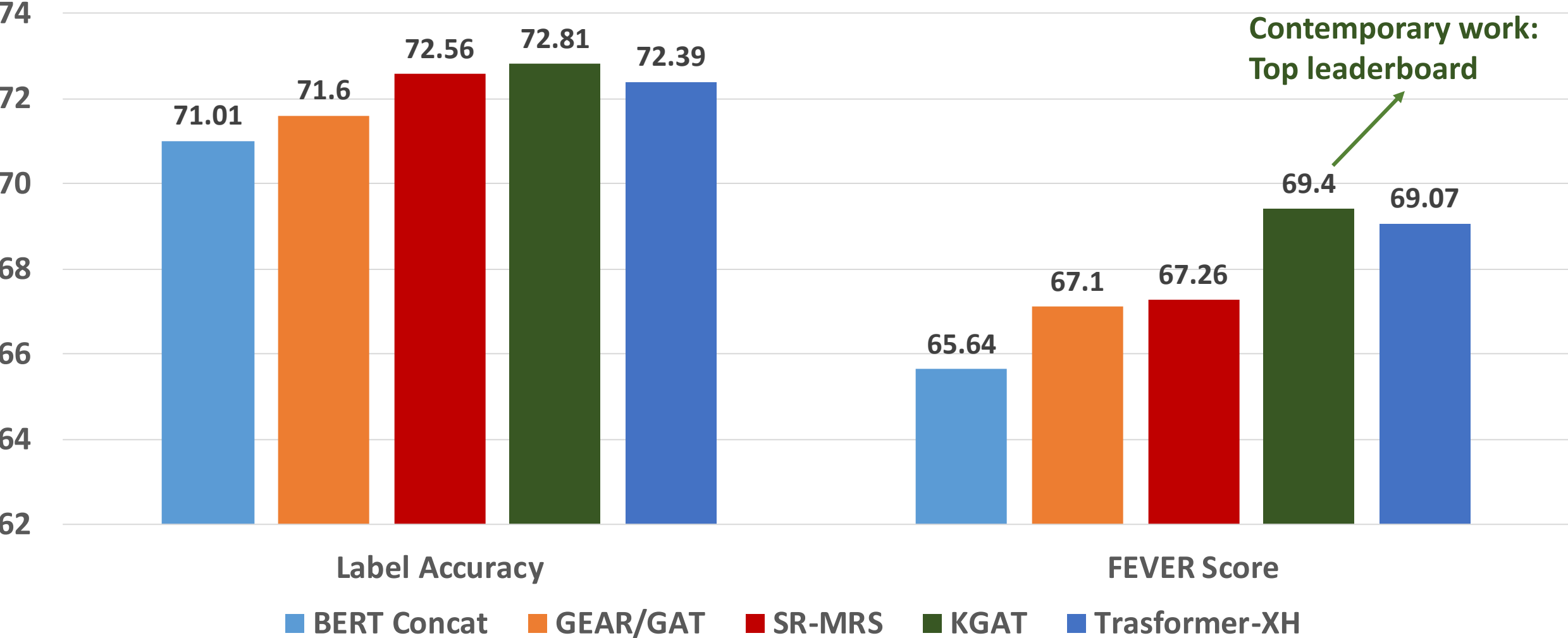
Results on Hotpot QA Full-Wiki Test Set



Results on Hotpot QA Full-Wiki Test Set



Results on FEVER 1.0



Summary

- Multi-Evidence QA: Beyond current MR style QA
- Free-text Knowledge Graph: Graph Construction -> Graph Modeling
- General solution to model structured text
- Strong performance & Generalizable to multiple tasks
- Code, Data, Processed graph are available
<http://users.umiacs.umd.edu/~chenz/>

Thanks!