

Part VII

Open-Domain Question Answering using Text & Knowledge Bases

Large-scale Factual Knowledge Bases



Knowledge Graph



NELL



- Database that stores a large number of facts in an organized way
 - Freebase: 46m entities, 2.6b facts
 - WikiData: 87m items
 - Most knowledge bases are curated

Entity-centric Knowledge Graph

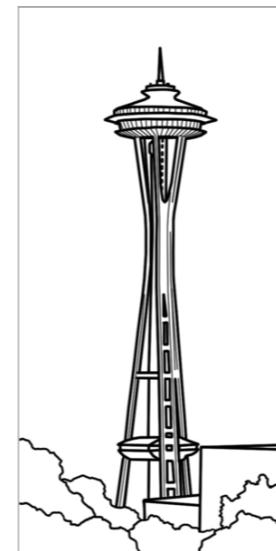
Super Bowl championships: 2013
Head coach: Pete Carroll
Founded: 1976
Division: NFC West



Seattle



Founded: Mar 30, 1971 · Pike Place Market
Customer service: +1 800-782-7282
CEO: Kevin Johnson
Founders: Jerry Baldwin · Zev Siegl · Gordon Bowker



Address: 400 Broad St, Seattle, 98109
Phone: (800) 937-9582
Opened: Apr 21, 1962
Height: 605 feet (184.41 m)
Floors: 6

Location

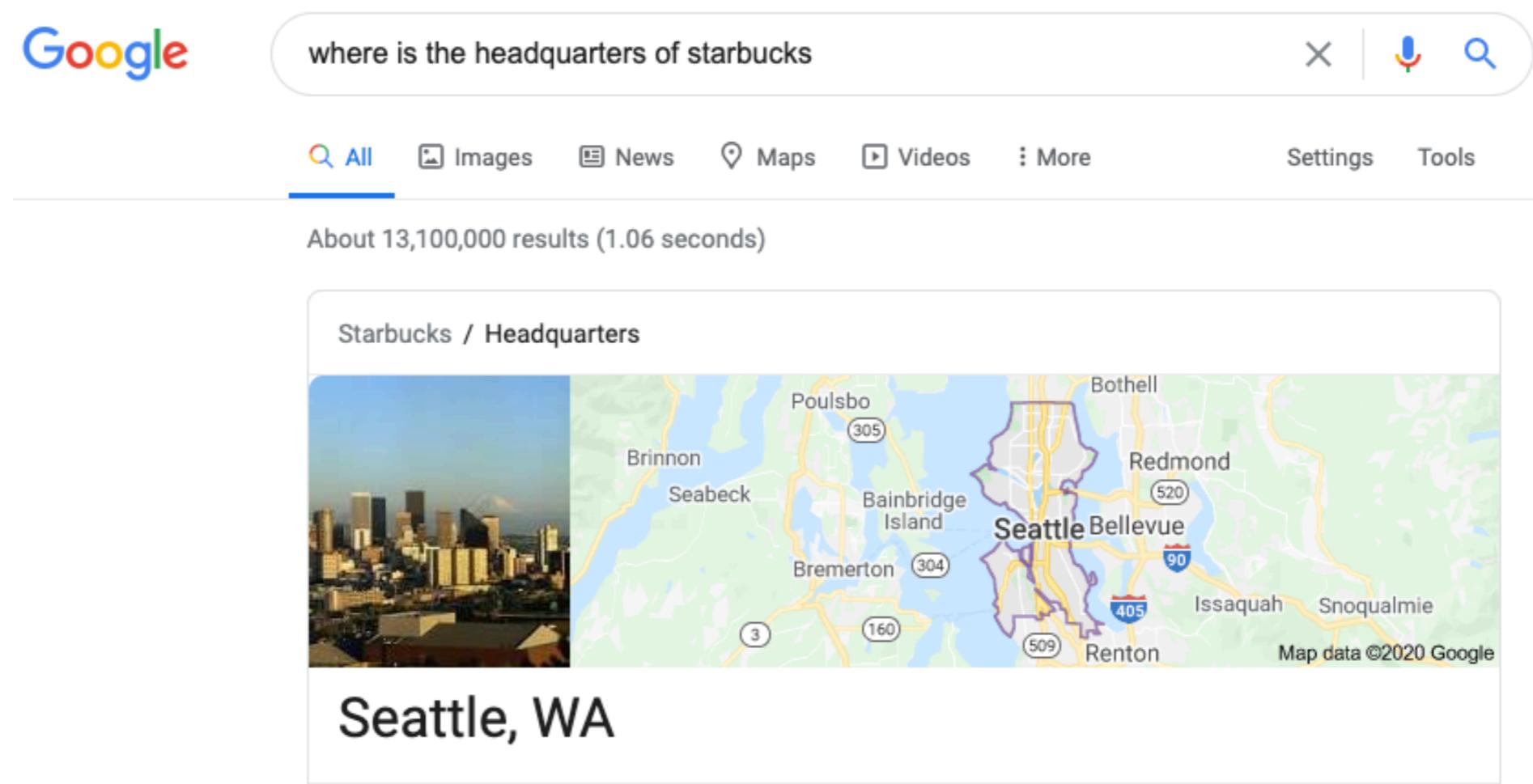
Home Field

Headquarters

Population: 744,955 (2018)
Area: 142.07 sq miles (369.20 km²)
Mayor: Jenny Durkan

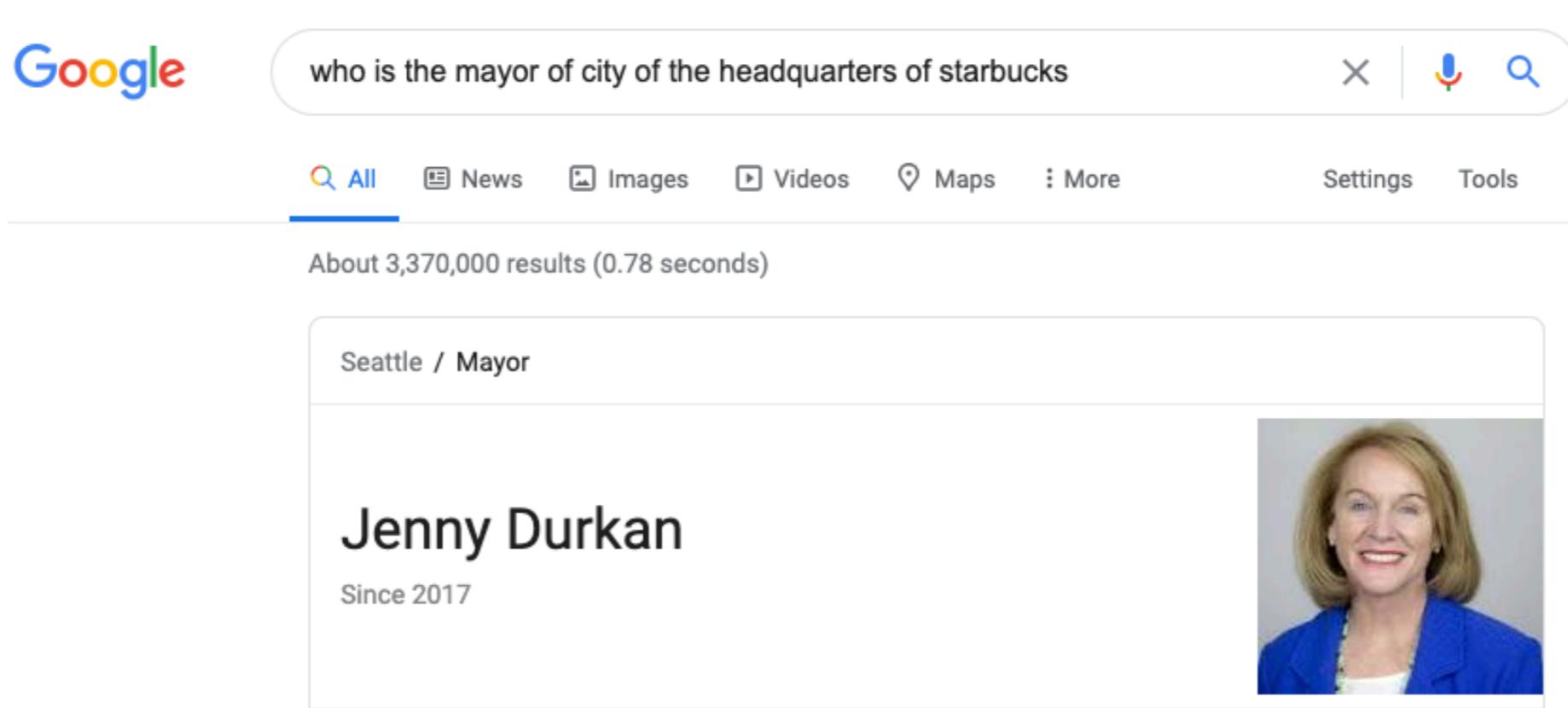
Open-Domain Question Answering using Knowledge Base

- Curated KB ensures the correctness of the information (answer)
- Common or "simple" questions can be answered easily
 - If semantic parsing (question → query) is done correctly



Open-Domain Question Answering using Knowledge Base

- Curated KB ensures the correctness of the information (answer)
- Common or "simple" questions can be answered easily
 - If semantic parsing (question → query) is done correctly
 - Graph structure enables multi-hop question answering



Open-Domain Question Answering using Knowledge Base

- Curated KB ensures the correctness of the information (answer)
- Common or "simple" questions can be answered easily
 - If semantic parsing (question → query) is done correctly
 - Graph structure enables multi-hop question answering



Open-Domain Question Answering using Knowledge Base

Limited coverage of questions that can be answered

- Common relationships & entities
- KB may not be timely updated

A screenshot of a Google search results page. The search query "where is the protest in seattle" is entered in the search bar. Below the search bar, there are navigation links for All, Images, News, Maps, Shopping, More, Settings, and Tools. A message indicates "About 90,900,000 results (0.84 seconds)". The main result is a card titled "1999 Seattle WTO protests / Location". It features a thumbnail image of the Seattle skyline and a map of the Seattle metropolitan area. The map shows major cities like Poulsbo, Brinnon, Seabeck, Bainbridge Island, Bremerton, Bothell, Redmond, Bellevue, Issaquah, and Snoqualmie, along with state routes 305, 304, 160, 509, 520, 90, and 405. The Seattle city boundary is highlighted in purple. The text "Seattle, WA" is displayed below the map.

Hybrid Approach: QA using Both Text and KB

Question: Who did Cam Newton sign with?

Knowledge graph

Cam Newton *plays* football
Cam Newton *career_starts* 2011
Cam Newton *date_of_birth* 19890511
...

Wikipedia Text

Cameron Newton (born May 11, 1989)
players for the Carolina Panthers of the
National Football League (NFL)
...

Answer: Carolina Panthers

Hybrid Approach: QA using Both Text and KB

- General Approach
 - Entity linking that connect mentions in question/text to KB
 - Graph representations for answer classification
- Papers to discuss
 - GRAFT-Net [Sun & Dhingra et al., 2018]
 - PullNet [Sun et al., 2019]
 - Knowledge-aware Reader [Xiong et al., 2019]
 - Knowledge-Guided Text Retrieval [Min et al., 2020]

WebQuestionsSP (WebQSP) [Yih et al., 2016]

- Answers to the questions are Freebase entities
- An enhanced version of WebQuestions
 - Remove ill-formed, ambiguous questions
 - Include full semantic parses (SPARQL) with entity annotations

Question: Who voiced Meg on Family Guy?

Parse:

```
PREFIX ns: <http://rdf.freebase.com/ns/>
SELECT ?x
WHERE {
  ns:m.035szd ns:tv.tv_character.appeared_in_tv_program ?y0 .
  ?y0 ns:tv.regular_tv_appearance.actor ?x ;
    ns:tv.regular_tv_appearance.series ns:m.019nnl ;
    ns:tv.regular_tv_appearance.special_performance_type
      ns:m.02nsjvf .
}
```

Answer:



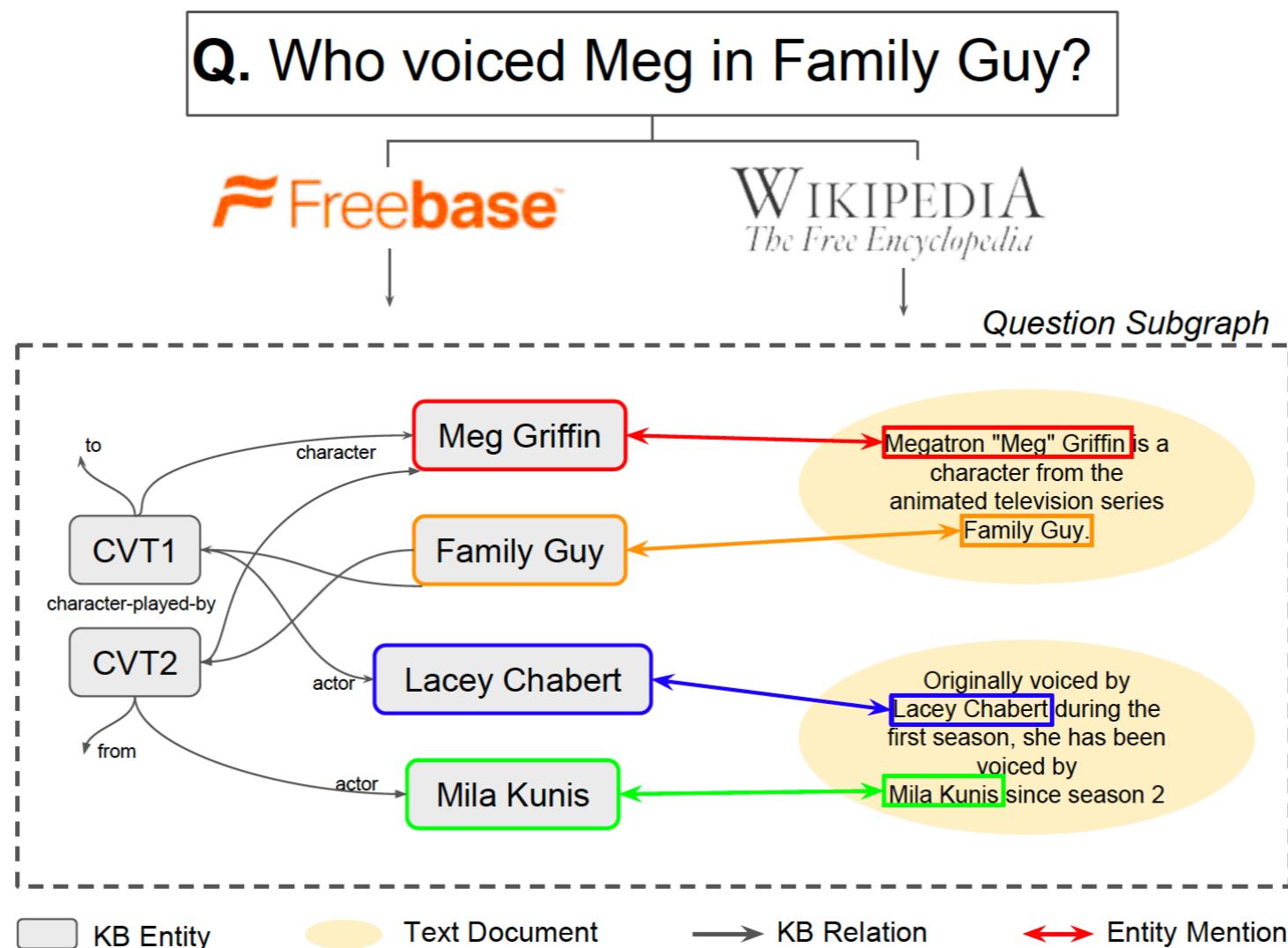
Lacey Chabert



Mila Kunis

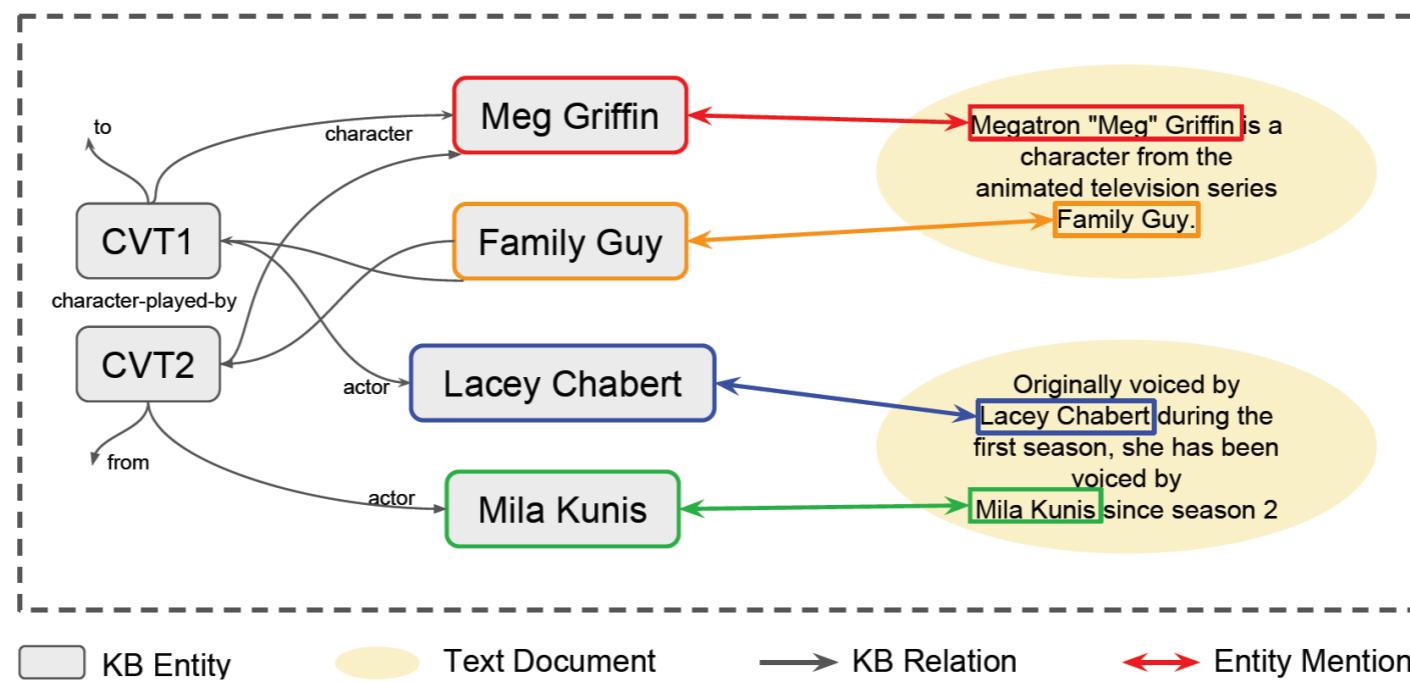
GRAFT-Net [Sun & Dhingra et al., 2018]

- Construct a heterogeneous **question graph** using both KB and text
- Use a variant of graph CNN to select an entity (node) as answer



Question Subgraph Retrieval

- KB Retrieval
 - Seed entities: entities in the question
 - Use Personalized PageRank (PPR) to find top entities around seed entities, weighted by similarity of question and relation (edge type)
- Text Retrieval
 - Top sentences from Top 5 relevant Wikipedia articles



Graph Propagation

- Basic recipe

1. Initialize node representations $h_v^{(0)}$
2. For $l = 1, \dots, L$, update node representations

$$h_v^{(l)} = \phi \left(h_v^{(l-1)}, \sum_{v' \in N_r(v)} h_{v'}^{(l-1)} \right)$$

- Different update rules for KB and text entities in GRAFT-Net
- Answer selection: binary classification on nodes (entities)

$$\Pr(v \in \{a\}_q | G_q, q) = \sigma(w^T h_v^{(L)} + b)$$

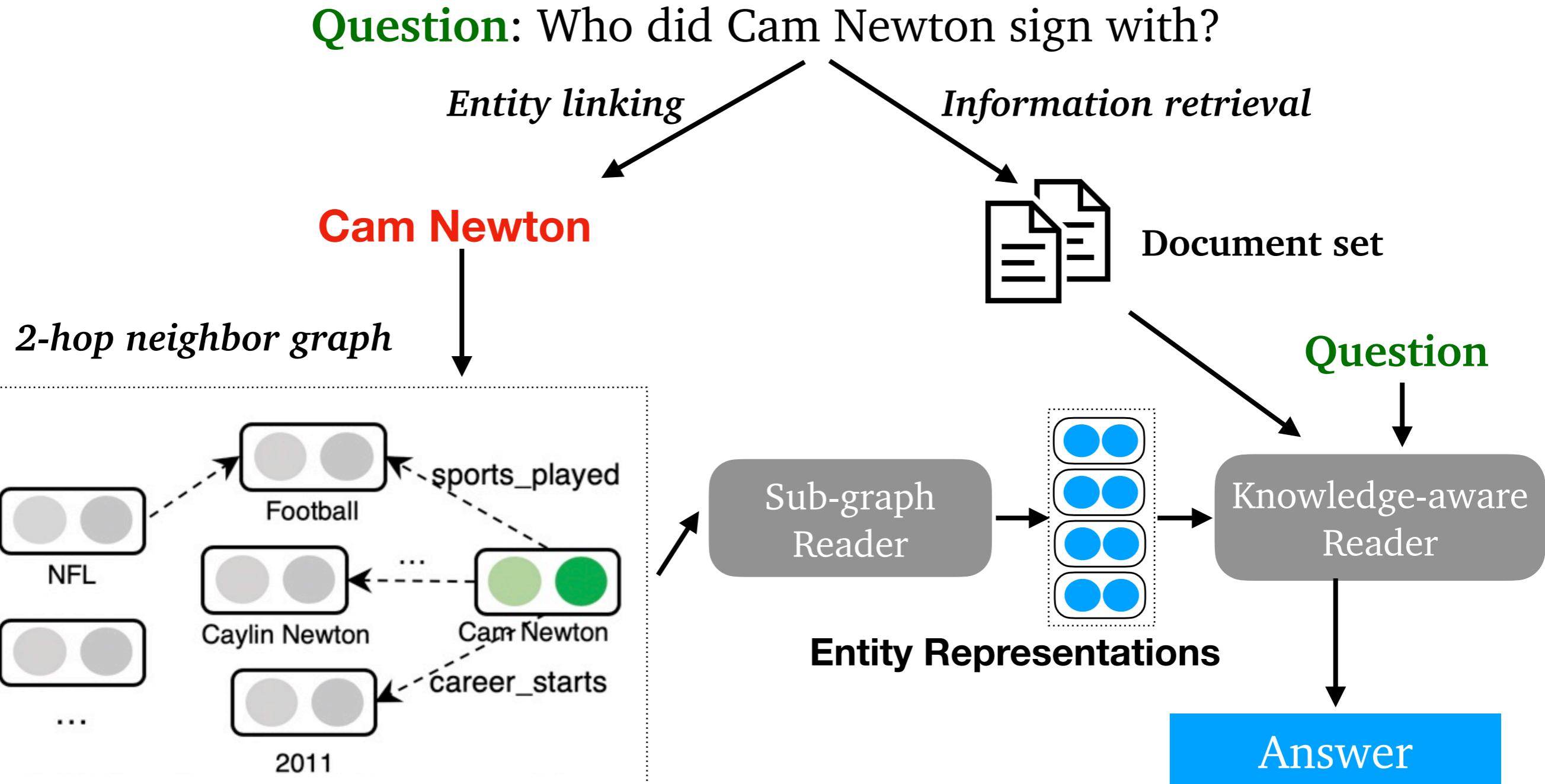
PullNet [Sun et al., 2019]

- Built on GRAFT-Net
- Incrementally construct the question graph using a graph CNN
- Algorithm
 1. Initial graph G^0 with question entities $V^0 = \{e_q\}$
 2. For $t = 1, \dots, T$ do
 - 1) Select entity nodes from V^{t-1} : $\{v_e\}$
 - 2) Find documents that contain entity e , and all other entities in the same documents
 - 3) Find all neighboring entities of entity e in the KB
 - 4) Add all entities and edges to G^t
 3. Select an entity node of V^T as answer

PullNet [Sun et al., 2019]

- The **Select** operation is determined by the same binary classifier as in GRAFT-Net.
- Training data: Approximate an ideal question graph using weak supervision (questions and their gold answers)
 - Find all shortest paths in the KB between question entities and answer entities
 - For each entity in the paths, include its neighboring entities up to a pre-determined length

Knowledge-aware Reader [Xiong et al., 2019]



Sub-graph Reader

- Start with the sub-graph embeddings from GRAFT-Net
- Reweigh the entity embeddings using neighboring entities and relations:

$$\vec{e}'_i = \sum_{e_j \in N(e_i)} \alpha_{ij} W \cdot [\vec{r}_{ij}; \vec{e}_j]$$

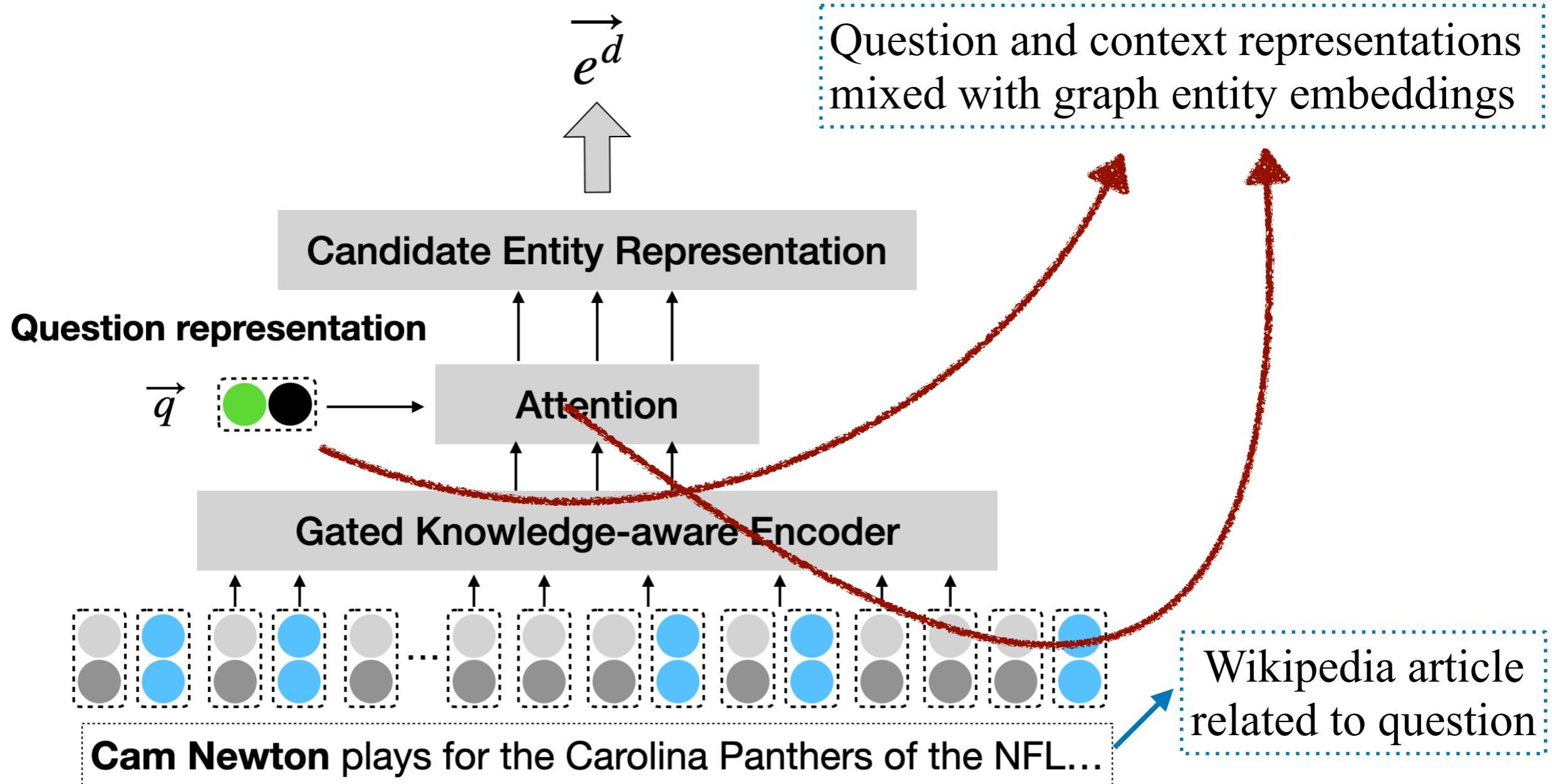
Similarity between question and r_{ij}

Learned weighting parameters

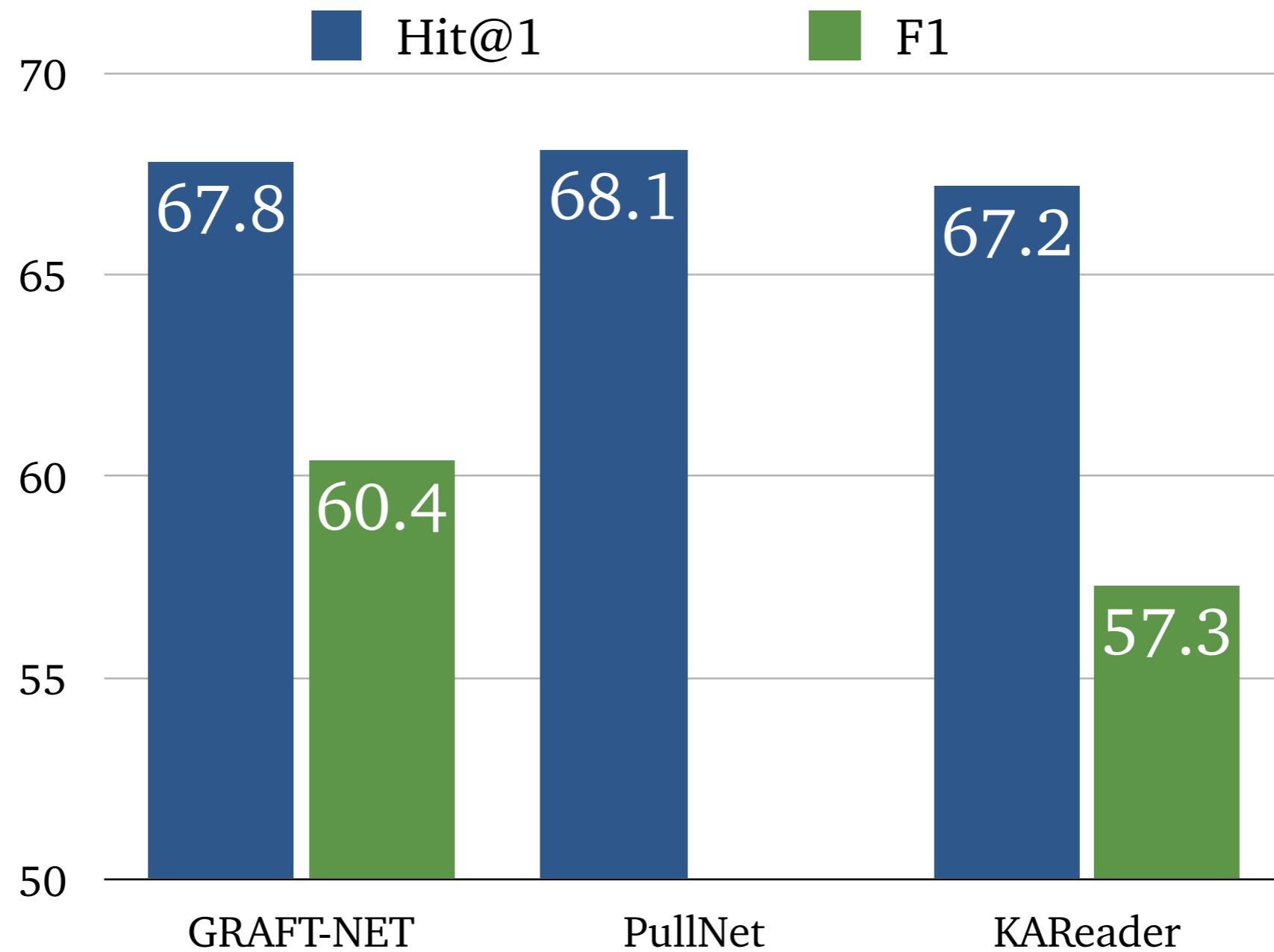
Relationship between e_i and e_j

Neighboring node

Knowledge-Aware Text Reader



Results on WebQSP

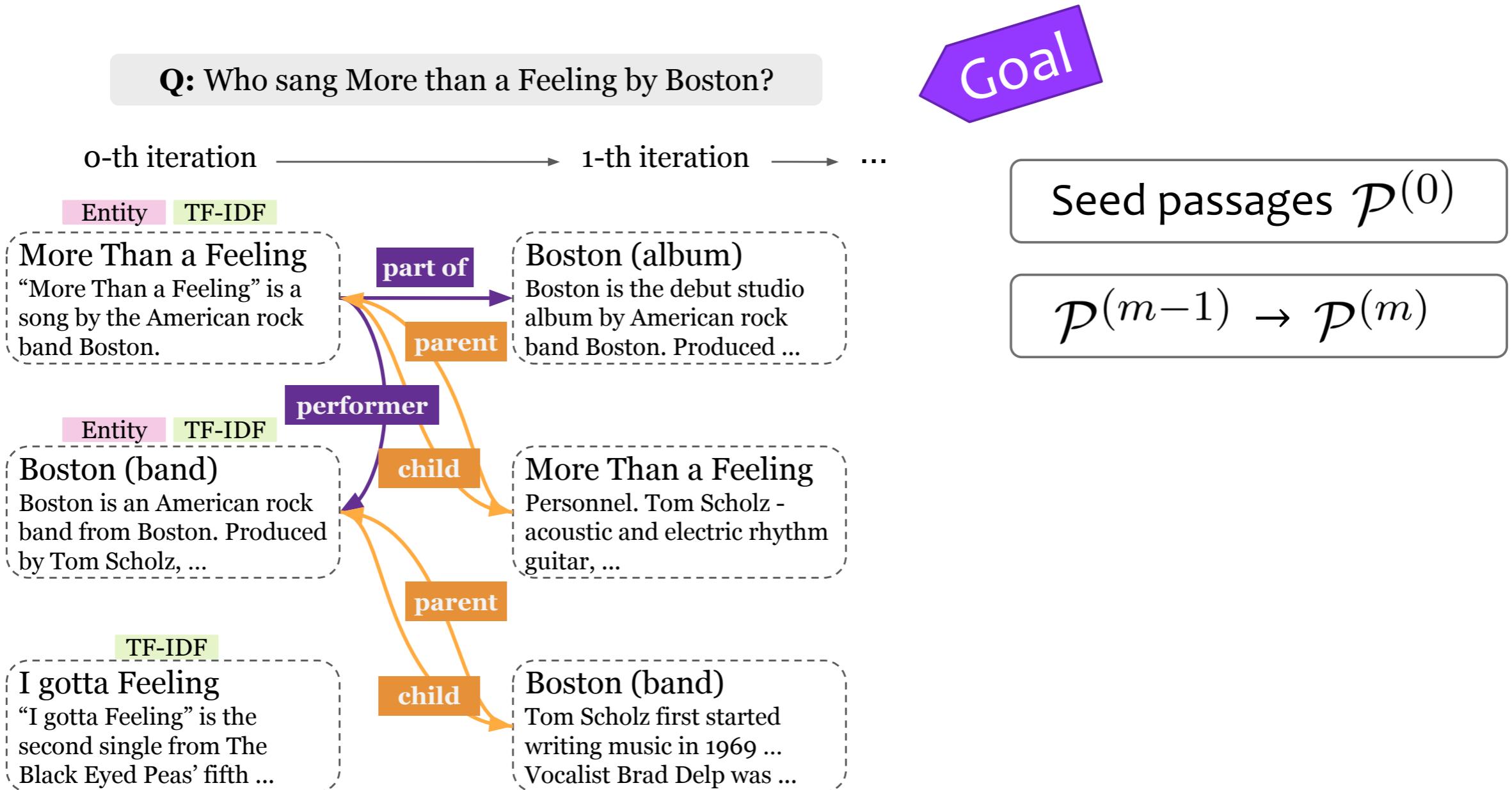


Knowledge-Guided Text Retrieval and Reading

[Min et al., 2020]



Graph Retriever



Graph Retriever

Seed passages $\mathcal{P}^{(0)}$

Q: Who sang More than a Feeling by Boston?

0-th iteration

Entity TF-IDF

More Than a Feeling
“More Than a Feeling” is a song by the American rock band Boston.

Entity TF-IDF

Boston (band)
Boston is an American rock band from Boston. Produced by Tom Scholz, ...

TF-IDF

I gotta Feeling
“I gotta Feeling” is the second single from The Black Eyed Peas’ fifth ...

Q: Who sang More than a Feeling by Boston?

Entity linking

More than a Feeling

Boston (band)

Q: Who sang More than a Feeling by Boston?

TF-IDF article retrieval

More than a Feeling

Boston (band)

I gotta Feeling

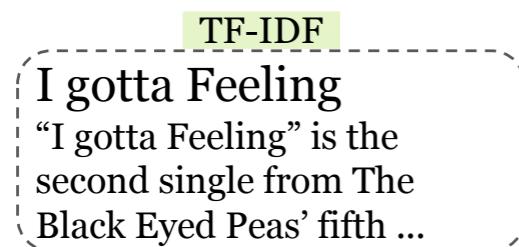
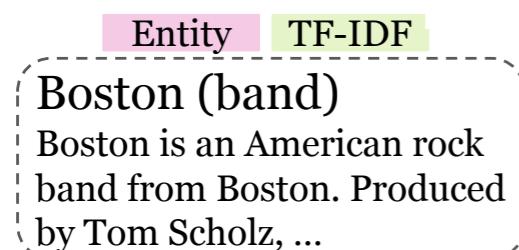
Boston, MA

Graph Retriever

$$\mathcal{P}^{(m-1)} \rightarrow \mathcal{P}^{(m)}$$

Q: Who sang More than a Feeling by Boston?

0-th iteration → 1-th iteration → ...



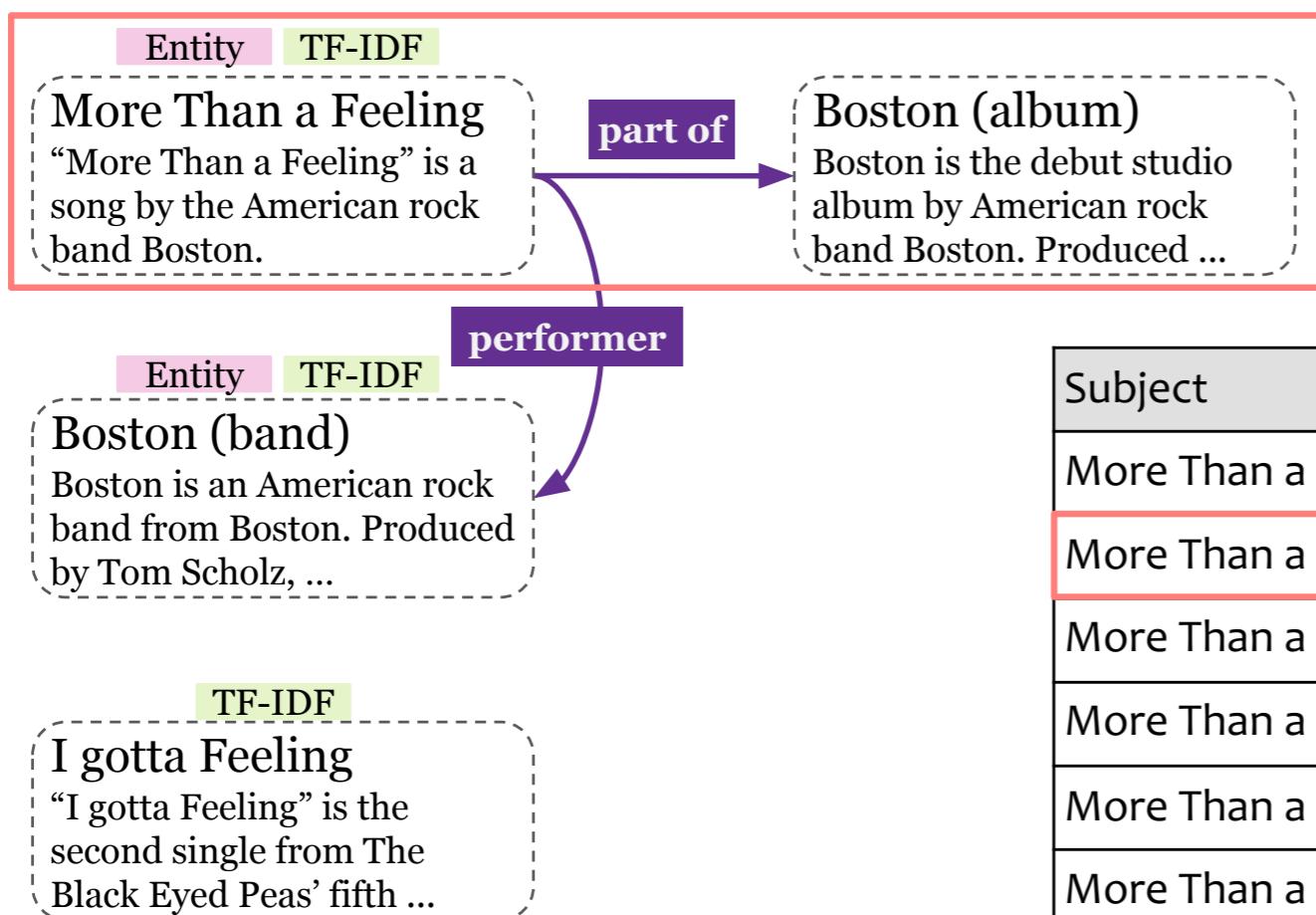
Subject	Relation	Object
More Than a Feeling	performer	Boston (band)
More Than a Feeling	part of	Boston (album)
More Than a Feeling	genre	Hard rock
More Than a Feeling	country of origin	USA
More Than a Feeling	record label	Epic
More Than a Feeling	followed by	Foreplay/Long Time

Graph Retriever

$$\mathcal{P}^{(m-1)} \rightarrow \mathcal{P}^{(m)}$$

Q: Who sang More than a Feeling by Boston?

0-th iteration → 1-th iteration → ...

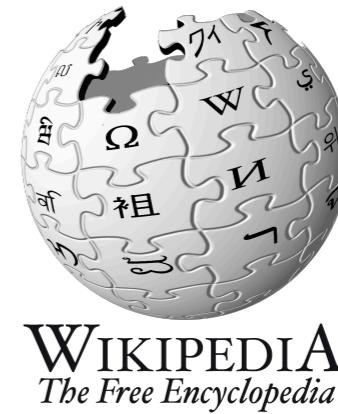
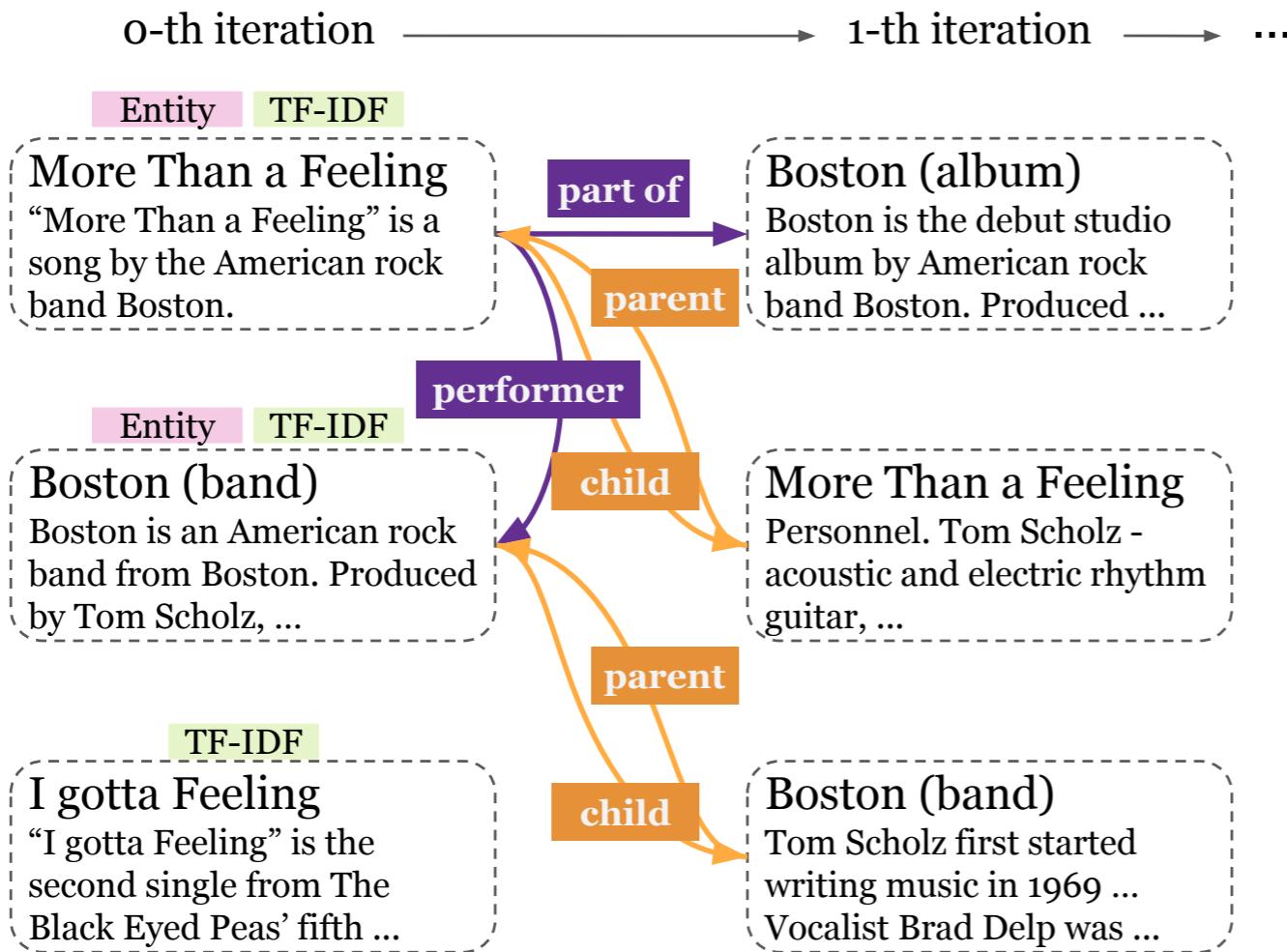


Subject	Relation	Object
More Than a Feeling	performer	Boston (band)
More Than a Feeling	part of	Boston (album)
More Than a Feeling	genre	Hard rock
More Than a Feeling	country of origin	USA
More Than a Feeling	record label	Epic
More Than a Feeling	followed by	Foreplay/Long Time

Graph Retriever

$$\mathcal{P}^{(m-1)} \rightarrow \mathcal{P}^{(m)}$$

Q: Who sang More than a Feeling by Boston?



More Than a Feeling

From Wikipedia, the free encyclopedia

“More Than a Feeling” is a song by the American rock band Boston. Written by Tom Scholz, it was released ...

Background and writing

The song took Tom Scholz five years to complete. It is one of six songs (five of which eventually appeared ...)

Content

The Book of Rick Lists suggests that the chorus ...

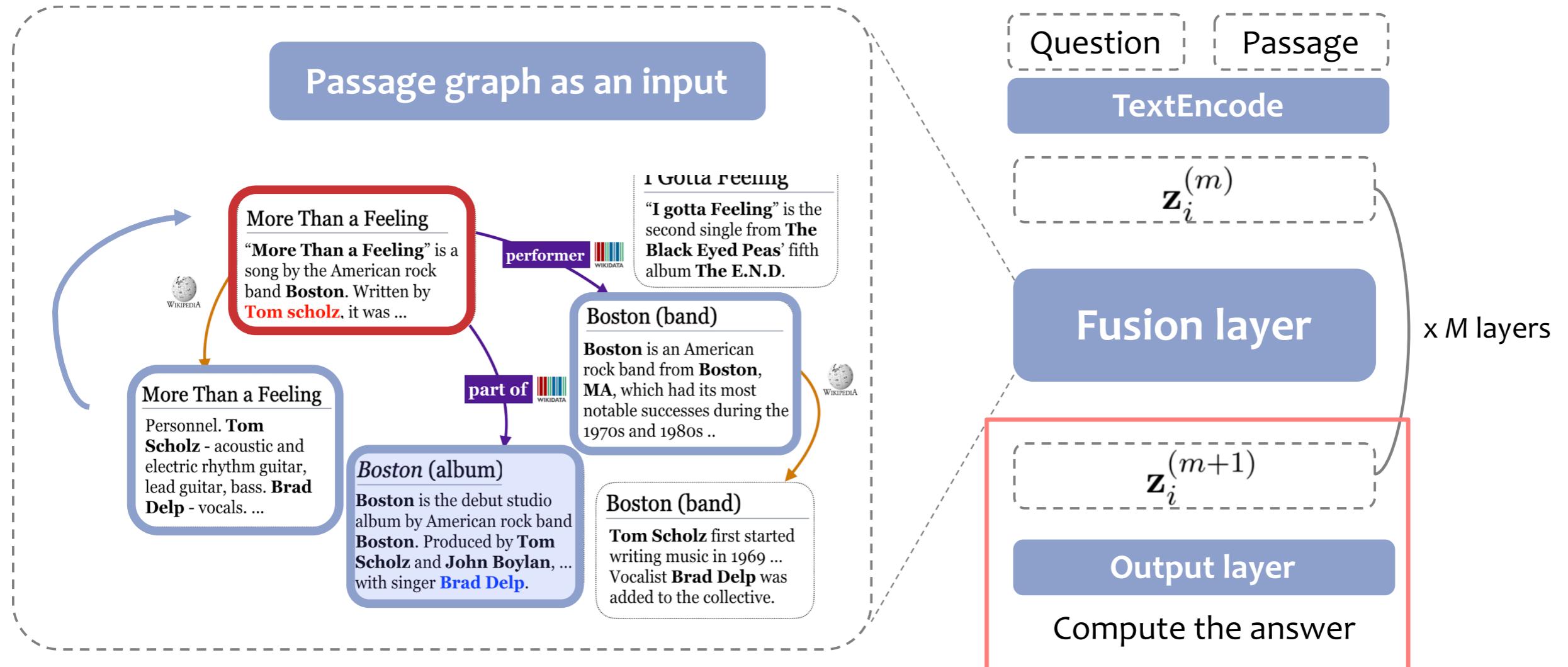
Reception

Guitar World states that when the radio plays ...

Single version

Epic released an edited version of the song for the ...

Graph Reader



Summary

- Basics of large-scale (entity-centric) knowledge bases
 - Pros and cons as the knowledge source for QA
- Open-domain QA using both KB and Text
 - GRAFT-Net [Sun & Dhingra et al., 2018]
 - PullNet [Sun et al., 2019]
 - Knowledge-aware Reader [Xiong et al., 2019]
 - Knowledge-Guided Text Retrieval [Min et al., 2020]
- Extended reading
 - Asai et al., 2020. Learning to Retrieve Reasoning Paths over Wikipedia Graph for Question Answering