RINK: Reader-Inherited Evidence Reranker for Table-and-Text Open Domain Question Answering

Eunhwan Park, Sung-Min Lee, Donghyeon Jeon, Seonhoon Kim, Inho Kang, **Seung-Hoon Na***

nash@jbnu.ac.kr*

Jeonbuk National University*

NAVER Corporation

Introduction

Open-Domain Question Answering (ODQA)

- Task: Find an answer to a given question, usually contextual contents
- Requires the **Retrieval** step for searching relevant knowledge for the given question.

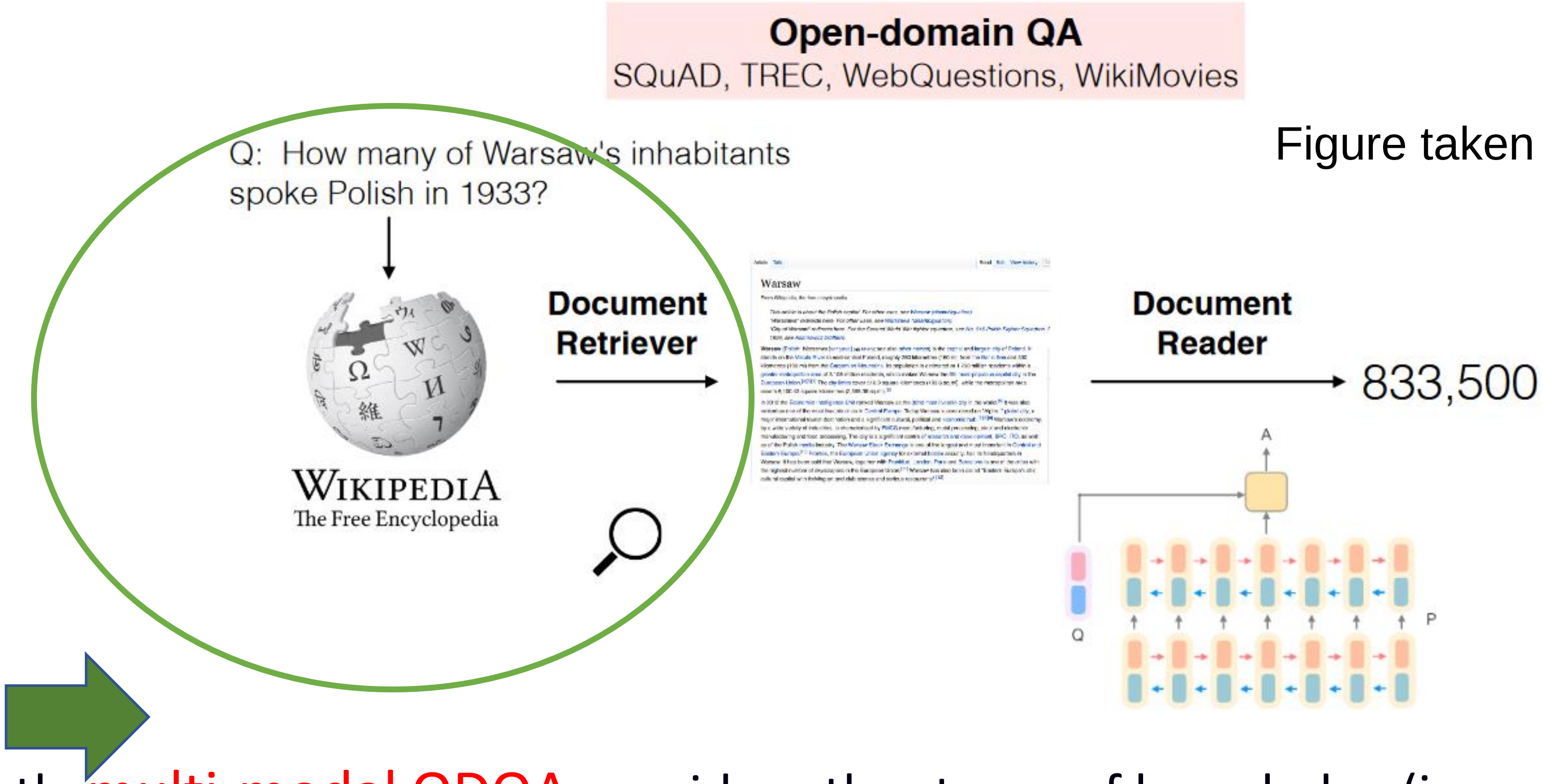


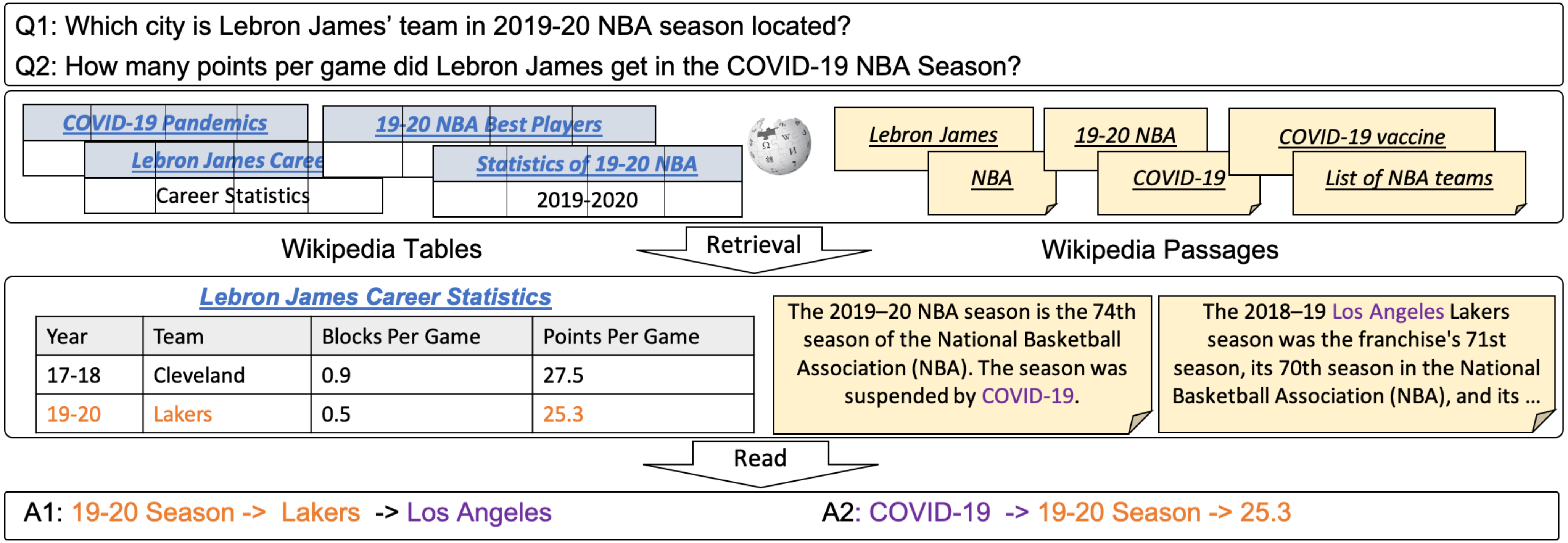
Figure taken from DrQA (Chen et al '17)

Recently multi-modal ODQA consider other types of knowledge (images and tabular contents) are provided as additional source of information

Table-and-Text Open-Domain Question Answering

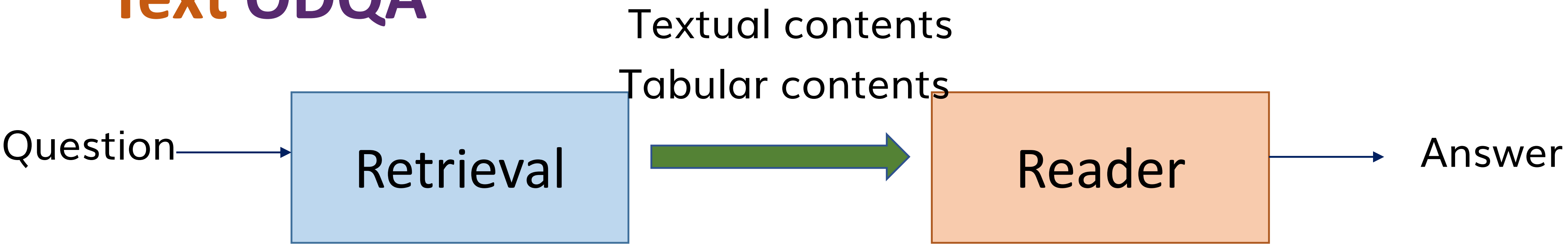
- **Table-and-Text ODQA**: Use heterogeneous data of tabular and textual contents as real-world knowledge for ODQA.
- **OTT-QA [Chen et al., 21]:** a new dataset for Open Table-and-Text Question Answering on the HybridQA dataset (Chen et al., 2020)

OTT-QA *Figure from `Open Question Answering over Tables and Text' [Chen et al 21]*



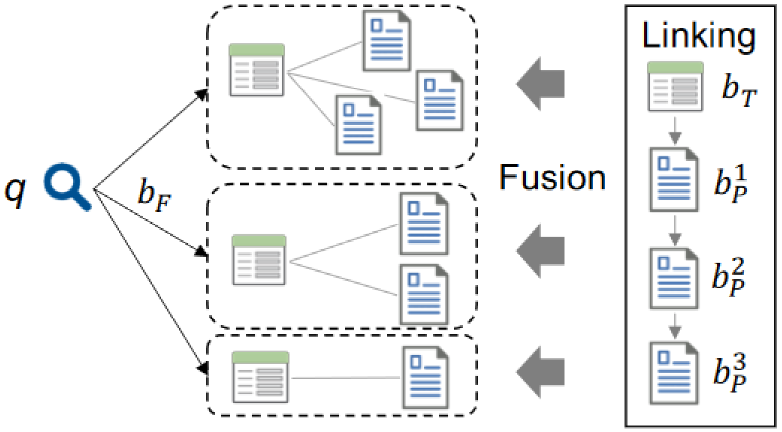
Multi-hop questions that requires aggregation across textual and tabular contents to get an answer.

Retrieval-Reader: Standard Framework for Table-and-Text ODQA



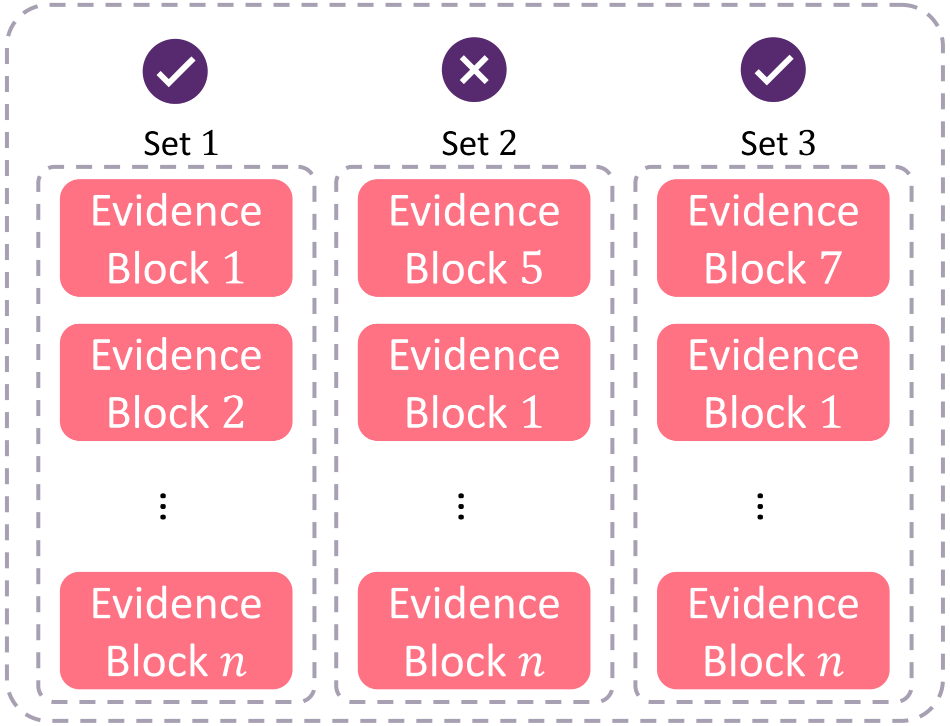
- The **Retrieval** module: finds a set of relevant “heterogenous” evidences
- The **Reader** module: Generates an answer using a decoder based on the retrieved evidences as an additional input.
- **Fusion block** [Chen et al ‘21] to handle **Multimodality**
  - **Early fusion** is applied to use a **fusion block** as the basic retrieval un by **combining table cells with their related passages**
  - The use of fusion block leads to **substantial improvement** over the single block approach
- The retrieval with fusion blocks is still challenging
  - the collection size of fusion blocks becomes readily huge because of its combinatorial nature

[Chen et al ‘21]



Retriever-Reranker-Reader: As an Initial Work for table-and-text ODQA

- Reader-INherited evidence reranKer (RINK)
  - **Set-level reranker:** Reranker module is inherited from the reader module with the same architecture, by taking a **set** as an input
  - **Set-level binary classification:** Performs a prompting-based **binary classification** to determine whether a given set contains a relevant block



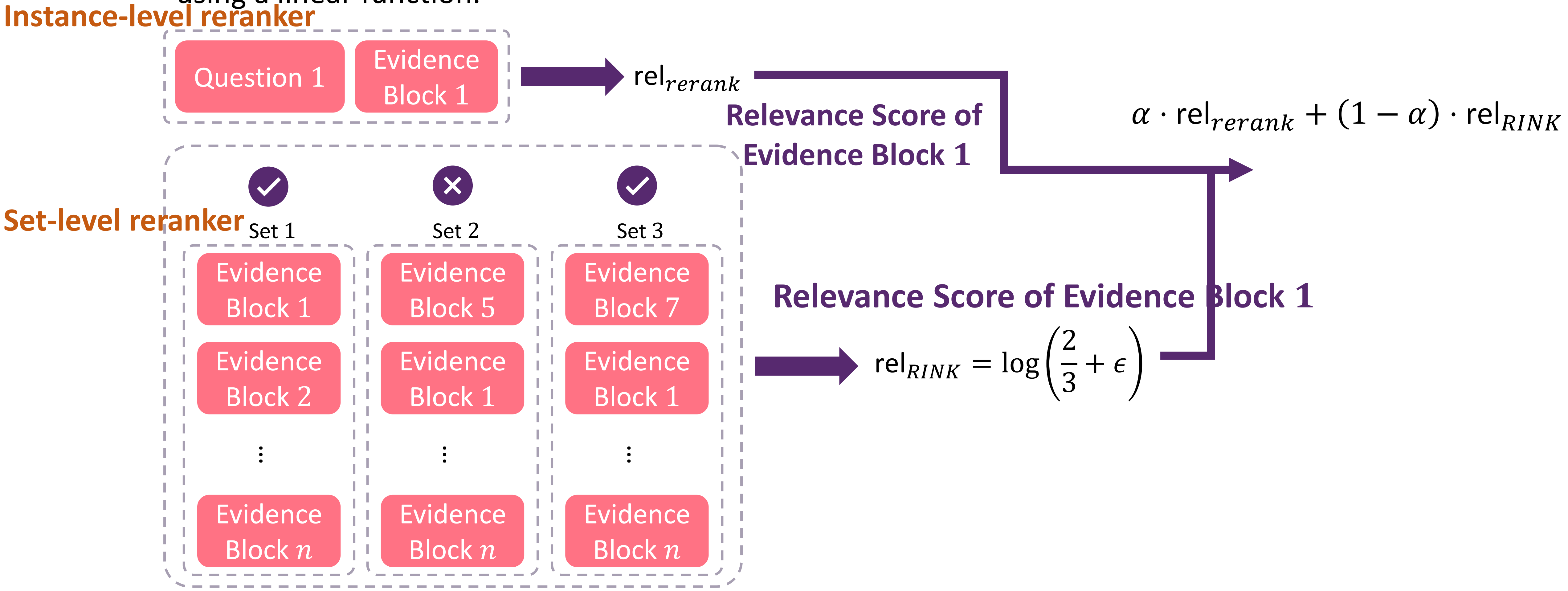
Evidence block 1 appears in three block sets.
But, only two sets are positive

Relevance Score of Evidence Block 1

$$\mathrm{rel}_{RINK} = \log\left(\frac{2}{3} + \epsilon\right)$$

Retriever-Reranker-Reader: for table-and-text ODQA

- Reader-INherited evidence reranKer (RINK)
  - **Combined reranker**: Combine set-level reranker and instance-level reranker
    - Utilize instance-level reranker and combine with rel_{RINK} to capture advantages of both rerankers using a linear function.



RINK: Reader-Inherited Evidence Reranker

Background: Notations & Fusion Block

- Table-and-text ODQA: aims to find an answer to a question q and a set of fusion blocks B
- b: a fusion block, presented by a table segment block b_T and its associated list of passages $b_P^1, \cdots, b_P^L$
- $BERT$: refers to the encoder of a pretrained language model (e.g., BERT or RoBERTa)
- $T5$: refers to the encoder-decoder language model (e.g., T5 and BART.)

Table segment block b_T Associated passages $b_P^1, \cdots, b_P^L$

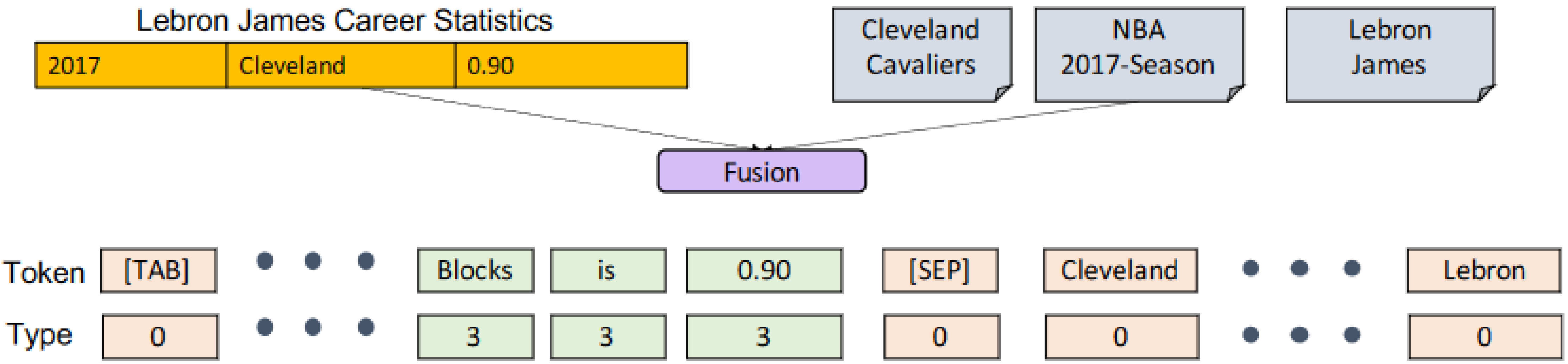
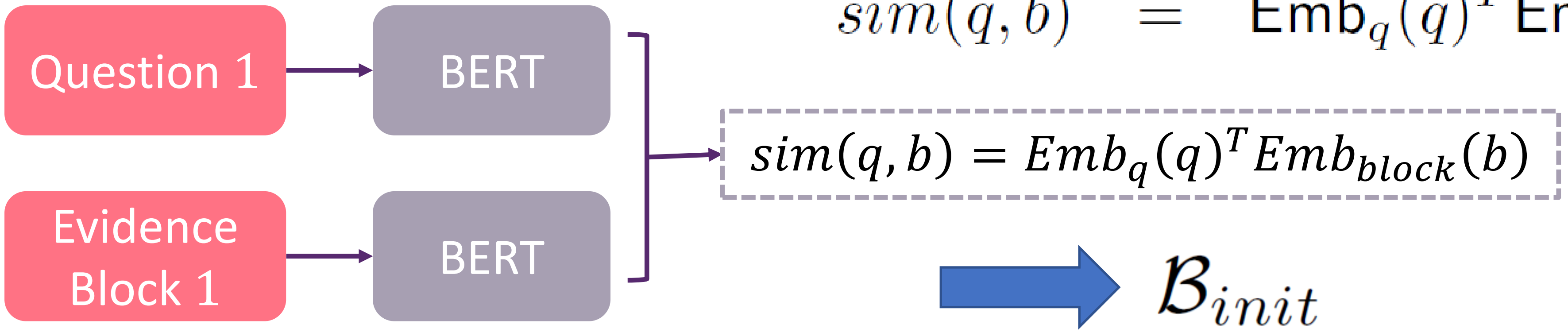


Figure taken from [Chen et al ‘21]

Baseline: (Initial) Retriever using bi-encoder

- Retriever
  - Based on **bi-encoder**, by using $BERT$ to **separately** encode a question q and a fusion block b
  - The similarity is computed using the inner-product between the encoded embeddings.
  - Employ FAISS (Johnson, Douze, and J´egou 2021) to obtain a set of top-N evidence blocks

$$\begin{aligned}
\mathrm{Emb}_q(q) &= \mathbf{W}_q \mathrm{BERT}_{[\mathrm{CLS}]}(q) \\
\mathrm{Emb}_{block}(b) &= \mathbf{W}_{block} \mathrm{BERT}_{[\mathrm{CLS}]}(b) \\
sim(q,b) &= \mathrm{Emb}_q(q)^T \mathrm{Emb}_{block}(q)
\end{aligned}$$



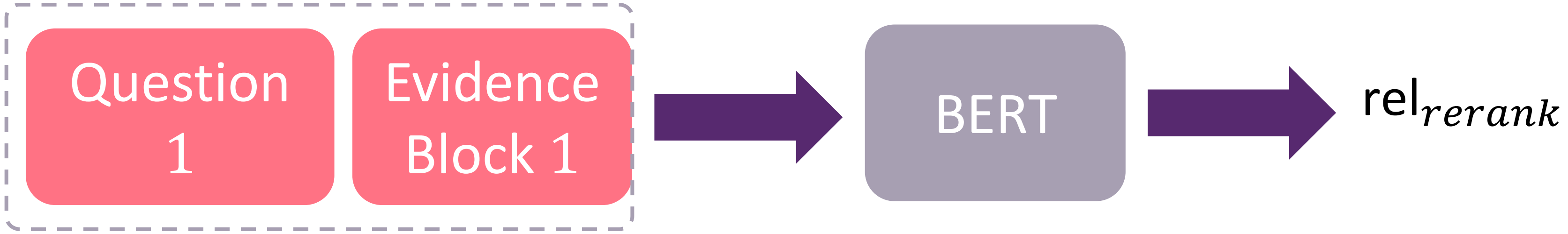
$\rightarrow \mathcal{B}_{init}$

a retrieved set of top-N fusion blocks

← We use these blocks for instance-level reranker too

Baseline: Reranker using cross-encoder

- Baseline Instance-level Reranker
  - Based on the **cross-encoder**, using BERT on the concatenated input of a question q and a fusion block b (q [SEP] b).
  - Perform the binary classification of whether the given block is relevant
  - The reranker takes B_{init}, the top N initial retrieval results, and produces the top M reranked results, B_{top}



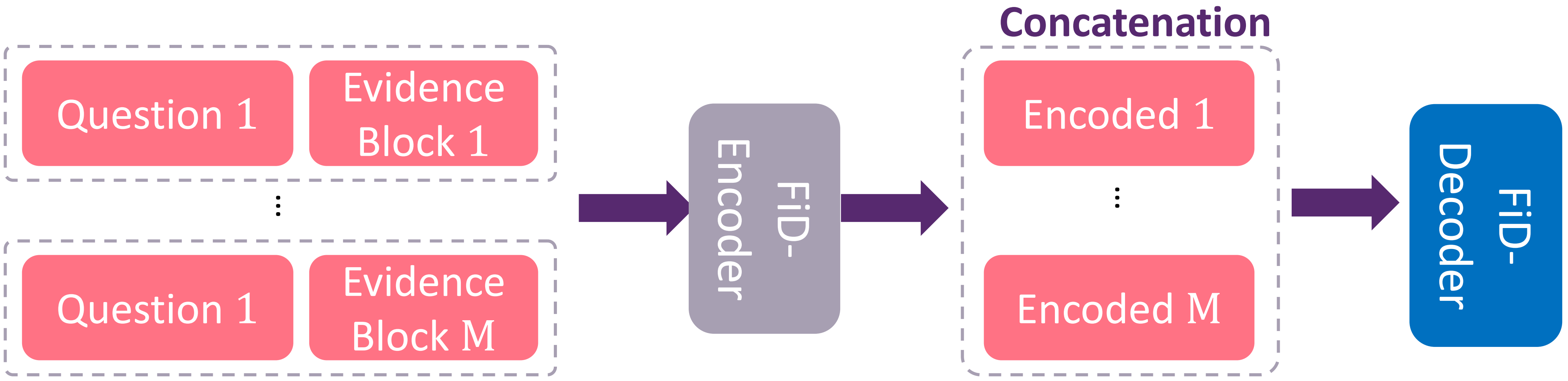
$$
\begin{aligned}
\mathrm{Emb}_{cross}(q, b) &= \mathrm{BERT}_{\mathrm{[CLS]}}(q\ \mathrm{[SEP]}\ b) \\
rel_{rerank}(q, b) &= \log \sigma\left(\mathrm{Linear}\left(\mathrm{Emb}_{cross}(q, b)\right)\right)
\end{aligned}
$$

$\mathcal{B}_{top}$

a set of top M reranked fusion blocks

Baseline: Reader using Fusion-in-Decoder (FiD)

- Reader
  - Based on the **(Fusion-in-Decoder (FiD))**: For each fusion block, its concatenated input with question c fed into T5's encoder to produce its contextualized representation
  - Concatenate all the contextualized block representations to fed into T5's decoder.

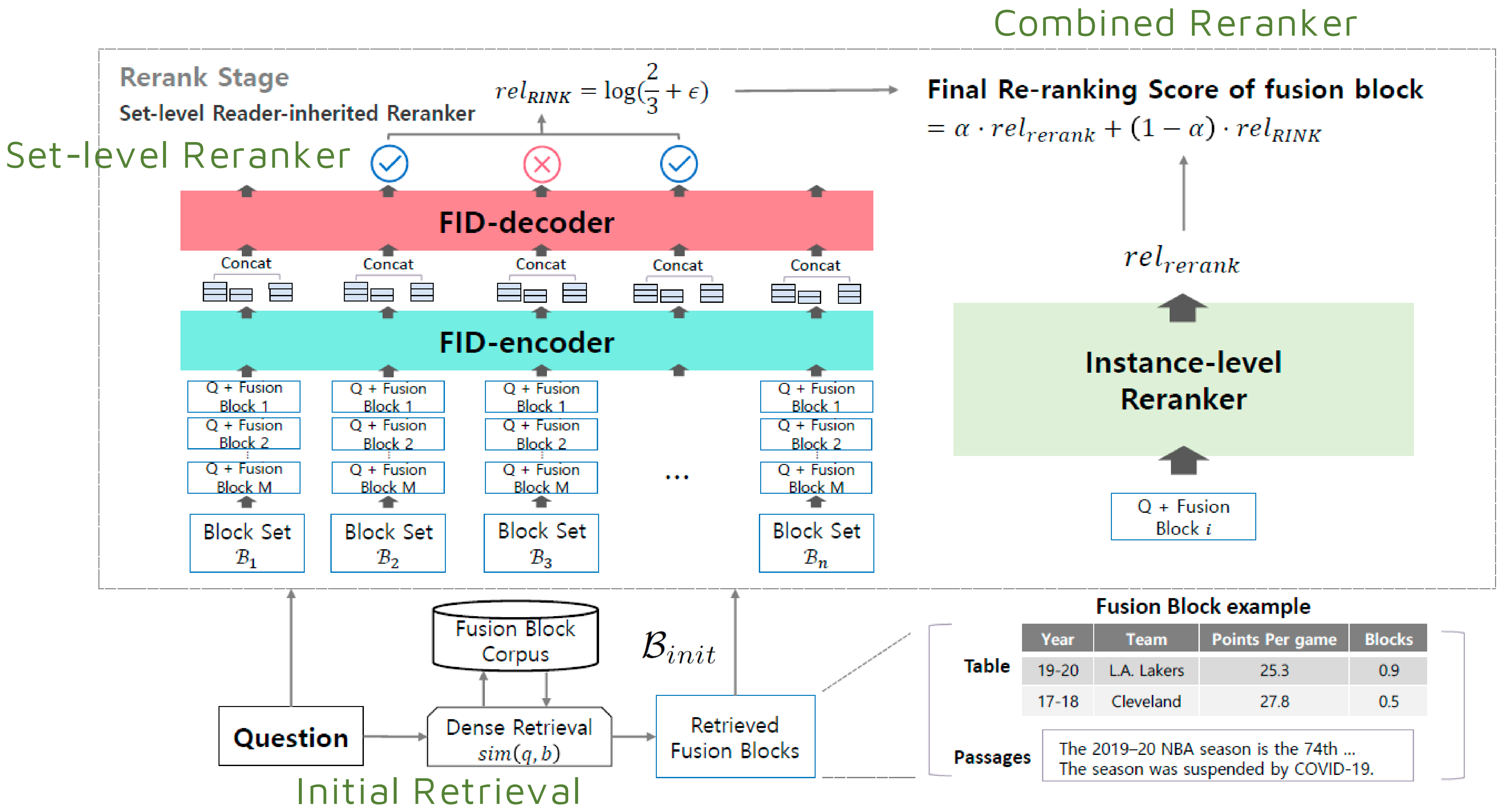


$$\mathbf{C}_i = \text{T5-enc}(q\ [\text{SEP}]\ b_i)$$

$$decode(q, \mathcal{B}_{top}) = \text{T5-dec}([\mathbf{C}_1; \cdots; \mathbf{C}_M], [\text{PAD}])$$

Padding token used for the start token of the decoder.

RINK: Neural architecture of Our Reranker



RINK: Set-level Relevance Classification

- Reader-inherited Reranker
  - Directly employs the reader module based on FiD, which takes a set of blocks as an input.
  - Reranker is obtained by finetuning the FiD reader module based on a prompting method to determine whether a given set of blocks is relevant.
- Set-level Relevance Classification
  - Relevance label: A set of blocks is relevant if at least one element block is relevant.
  - Input: a set of fusion blocks– the concatenated sequence of M fusion blocks $\{b_1, b_2, \cdots, b_M\}$ each of which is associated with a question q
  - Output: the classification result - a probability vector over V to compute whether a set of M fusion blocks is relevant

Prompt tokens

For i-th block

$$\mathbf{C}'_i = \text{T5-enc}(\text{“query:”}\ q\ \text{“block:”}\ b_i\ \text{“relevant:”})$$

$$\mathbf{p}(q, \mathcal{B}) = \text{T5-dec}_{\text{token}}(\underbrace{[\mathbf{C}'_1; \cdots; \mathbf{C}'_M]}_{\text{Concatenated}}, [\text{PAD}])) \in \mathbb{R}^{|\mathcal{V}|}$$

$\text{T5-dec}_{\text{token}}(x, y)$: the autoregressive language model of T5 that produces a probability vector over V for the next token

RINK: Set-level Relevance Classification

- Prompting method for set-level relevance classification

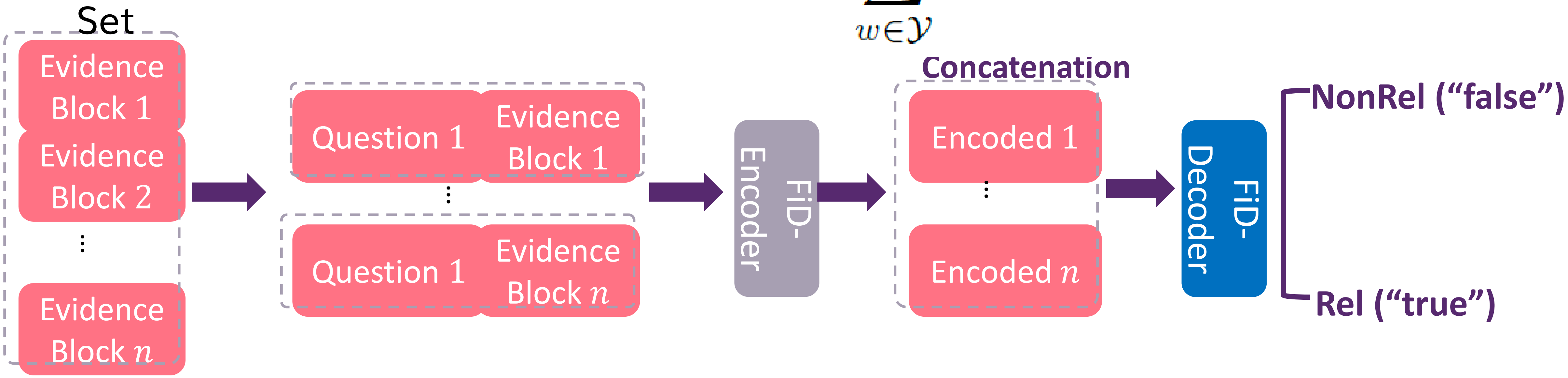
$$\mathbf{C}'_i = \text{T5-enc}(\text{“query:”}\ q\ \text{“block:”}\ b_i\ \text{“relevant:”})$$

$$\mathbf{p}(q, \mathcal{B}) = \text{T5-dec}_{\text{token}}([\mathbf{C}'_1; \cdots; \mathbf{C}'_M], [\text{PAD}])) \in \mathbb{R}^{|\mathcal{V}|}$$

The verbalizer $v: \mathcal{Y} \rightarrow \mathcal{V}$ converts a label into individual words:
v(Nonrel)= “false” and v(Rel)= “true”

- The probability of the relevance label: obtained by normalization over two token probs.

$$\text{p}(\text{Rel}|q, \mathcal{B}) = \mathbf{p}(q, \mathcal{B})[v(\text{Rel})] / \sum_{w \in \mathcal{Y}} \mathbf{p}(q, \mathcal{B})[v(w)]$$



RINK: Multiple Set-level Evidences

- RINK uses multiple set-level evidences and aggregate them
- Construct a collection of block sets $\mathcal{S} = \{\mathcal{B}_1, \cdots, \mathcal{B}_n\}$ $B_i \subset \mathcal{B}_{init}$
  - Here, we randomly construct n set samples, with the constraint that the number of sets containing each block is the same
- The constraint is restated as: $|S(b)| = K$ for any block b.
  - $\mathcal{S}(b) \subseteq \mathcal{S}$: the collection of sets containing b $\mathcal{S}(b) = \{\mathcal{B} | b \in \mathcal{B} \text{ and } \mathcal{B} \in \mathcal{S}\}$
- Given $|B_{init}| = N$, $|B_i| = M$, and $|S(b)| = K$, the number of set samples:

$$\boxed{n = N \times K/M}$$

  - Here, for consistency with the reader' s setting, we maintain the size of the set $|B_i|$ as the same as the reader ➔ $|B_{\mathrm{i}}| = M$ for any sampled set $B_{\mathrm{i}} \in S$

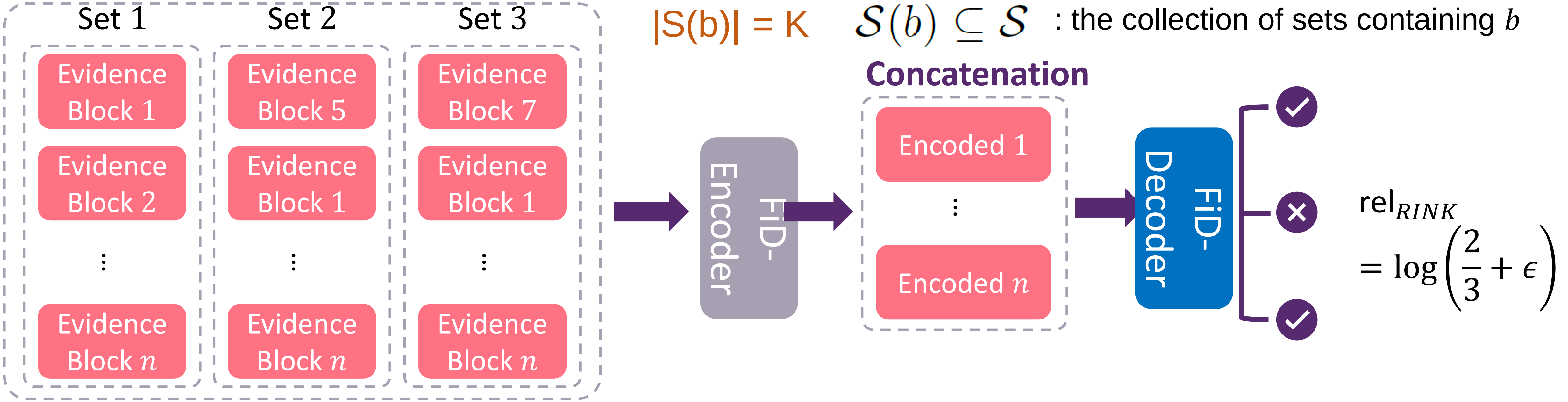
RINK: Aggregation over Set-level Relevance Results

- Apply the set-level classification for all n set samples to obtain set-level relevance scores for n sets in S
- Aggregate the set-level evidences to obtain the instance-level score of block b

The indicator function that is one if e is true and zero otherwise.

$$rel_{RINK}(q,b) = \log\left(\frac{\sum_{\mathcal{B}\in\mathcal{S}(b)} \mathcal{I}(\mathrm{p}(\mathrm{Rel}|q,\mathcal{B}) > 0.5)}{K} + \epsilon\right)$$

|S(b)| = K $\mathcal{S}(b) \subseteq \mathcal{S}$: the collection of sets containing b



RINK: Combining with the Baseline Instance-level Reranker

- We combine these two types of rerankers using a simple linear function

$$rel_{combined}(q,b) = \alpha \cdot rel_{rerank} + (1-\alpha) \cdot rel_{RINK}$$

  - α is an interpolation parameter, which is tuned on the development set

RINK: Pretraining the Encoder of Retriever

- Data augmentation: Inspired by (Iida et al, 2021), we construct a pre-training corpus using Wikipedia by employing ***cell corruption*** and ***cell reordering***.
- Pretraining Task 1: **Tabular-and-Textual Entailment**
  - A pair of a table and a passage is assumed to be an "entailment" class when there is a table cell that is hyperlinked to the passage
  - Other pairs are merely regarded as "contradiction" class.

Tabular-Textual Entailment Task

(a). Original Table

Year	Team	Points Per game	Blocks Per Game
18-19	Cleveland	27.5	0.9
19-20	L.A. Lakers	25.3	0.3

Entailment

Passage (L.A. Lakers):
The Los Angeles Lakers are an American professional basketball team based in Los Angeles. The Lakers … compete in the National Basketball league's Western Conference Pacific Division.

(b). Cell Corruption

Year	Team	Points Per game	Blocks Per Game
18-19	Cleveland	1984	0.9
19-20	New York	25.3	12.0

Contradiction

Passage (Michael Jordan)
Michael Jeffrey Jordan (born February 17, 1963), also known by his initials MJ, is an American businessman and former professional basketball player

(c). Cell Re-ordering

Year	Team	Points Per game	Blocks Per Game
18-19	27.5	Cleveland	0.9
19-20	L.A. Lakers	25.3	0.3

Entailment

Passage (Cleveland Cavaliers)
The Cleveland Cavaliers (often referred to as the Cavs) are an American professional basketball team based in Cleveland. The Cavaliers compete in the National Basketball Association (NBA) as a member of the league's Eastern Conference Central Division.

RINK: Pretraining the Encoder of Retriever

- Pretraining Task 2: **Cross-modal masked language modeling task**
  - Multi-modal extension of the masked token prediction
    - The masked token prediction: a "masked" concatenated sequence of tabular and textual contents is provided as an input to the retriever's encoder
    - The whole-cell masking for table contents and the BERT's whole-word masking for textual ones, as proposed by (Herzig et al. 2020)

Cross-Modal Masked Language Modeling Task

(a). Whole Cell Masking

Year	Team	Points Per game	Blocks Per Game
18-19	[MASK]	27.5	0.9
19-20	L.A. Lakers	[MASK]	0.3

(b). Whole Word Masking

Passage (L.A. Lakers):
The [MASK] are an [MASK] professional basketball team based in Los Angeles. The Lakers … compete in the National Basketball league's Western Conference Pacific Division.

Experiments

Main Results

- Outperforms CARP, the best baseline model, by increasing EM by 3.5 and 3.0 points on the development & blind test sets.

Model	Dev		Test	
	EM	F1	EM	F1
HYBRIDER	10.3	13.0	9.7	12.8
IR + SBR	7.9	11.1	9.6	12.8
FR + SBR	13.8	17.2	13.4	16.9
IR + CBR	14.4	18.5	16.9	20.9
FR + CBR	28.1	32.5	27.2	31.5
DUREPA	15.8	–	–	–
CARP	33.2	38.6	32.5	38.5
Instance-level Reranker ($M = 15$)	35.2	41.2	34.5	40.4
RINK ($\alpha = 0, M = 15, K = 30$)	34.6	40.3	32.7	38.6
RINK ($\alpha = 0, M = 10, K = 30$)	35.2	40.8	33.3	39.3
RINK ($\alpha = 0.7, M = 15, K = 30$)	36.2	42.1	35.0	41.0
RINK ($\alpha = 0.7, M = 10, K = 30$)	**36.7**	**42.4**	**35.5**	**41.5**

HYBRIDER (Chen et al. 2020b): employs a sparse retriever (i.e., BM25 and TF-IDF) to retrieve relevant tables and passages, and uses a reasoning model based on ranking, hop, and reading comprehension (RC) models to extract an answer

{Iterative, Fusion}-Retriever / {Single, Cross}-Block Reader (Chen et al. 2021)

- Iterative retriever (**IR**): uses the iterative retrieval protocol
- Fusion retriever (**FR**): adopts the "early fusion" strategy.
- Single block reader (**SBR**): feeds the top k-retrieved blocks to the reader one by one, and selects the answer with the highest confidence score.
- Cross block reader (**CBR**). feeds all concatenated top-k blocks together into the reader.

DUREPA (Li et al. 2021): jointly reads tables and passages using the dual-reader architecture and generates either an answer or an executable SQL query to derive the answer.

CARP (Zhong et al. 2022): proposed the use of a hybrid chain defined as a sequence of nodes from a heterogeneous graph

Retrieval Performance

RINK shows the substantial improvements, with increases of more than 10% at R@1 and R@5.

RINK shows further improvements over the "Instance-level reranker," particularly showing an increase of approximately 2% at R@5, R@10, and R@15, when M = 15.

Model	R@1	R@5	R@10	R@15
Bi-Encoder	–	–	72.9	–
Tri-Encoder	–	–	73.8	–
CARP	49.0	–	74.0	–
Retriever	51.0	66.2	76.6	79.6
RINK ($M = 15, \alpha = 0$)	61.4	77.4	81.6	83.7
RINK ($M = 10, \alpha = 0$)	62.7	78.0	81.9	84.0
Instance-level Reranker ($M = 15$)	65.0	77.4	81.2	83.6
RINK ($M = 15, \alpha = 0.7$)	66.1	79.9	83.3	**85.5**
RINK ($M = 10, \alpha = 0.7$)	**66.9**	**80.6**	**83.5**	85.4

Ablation Study: Effect of Varying K for RINK

The larger the value of K, the more effective the reranking, because the instance-level aggregated score is estimated more accurately for larger values of K.

Model	K	R@1	R@5	R@10	R@15
RINK ($\alpha = 0$)	6	58.7	75.7	80.6	82.8
	9	59.8	76.5	81.1	83.6
	15	60.8	76.6	81.3	83.6
	24	**62.0**	77.0	81.0	83.0
	30	61.4	**77.4**	**81.6**	**83.7**

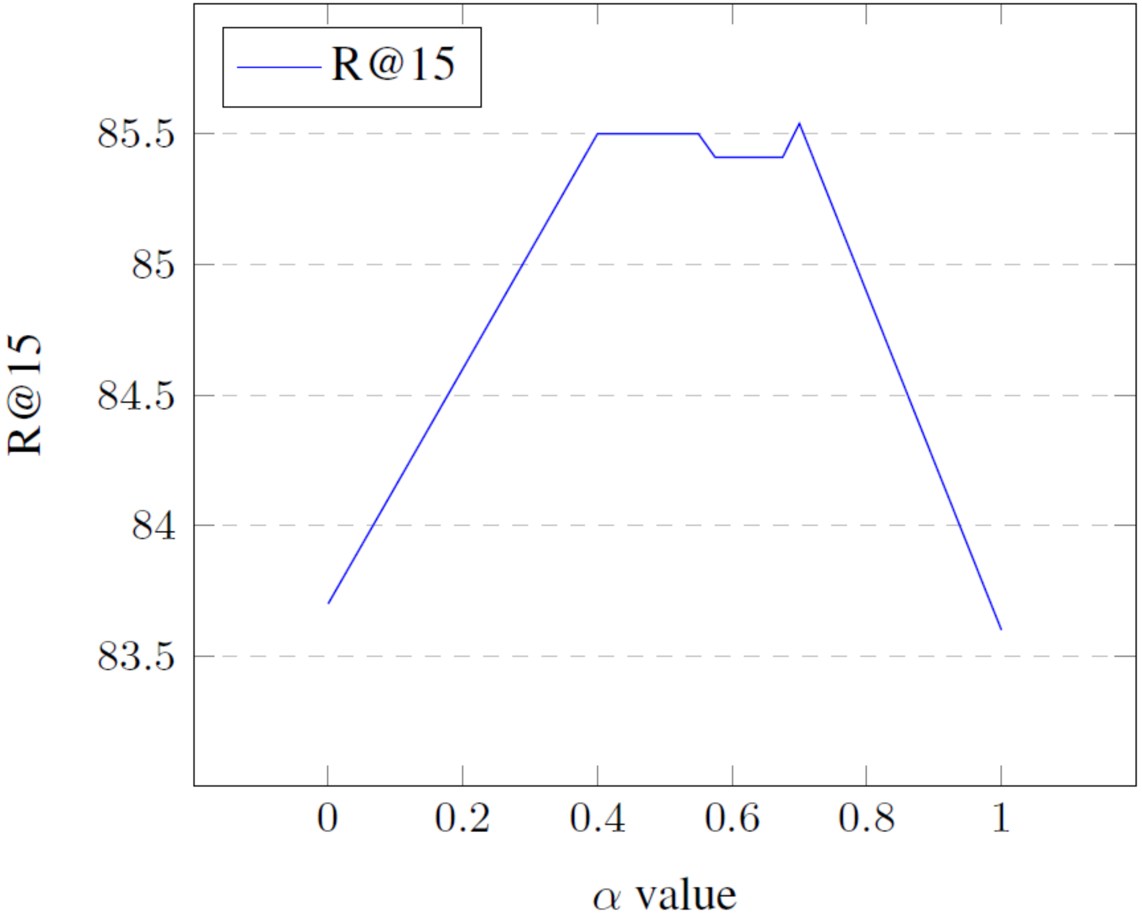
Retrieval performances of RINK (M = 15, α = 0) with varied K values

Model	K	EM	F1
RINK ($\alpha = 0$)	6	33.4	39.3
	9	33.7	39.4
	15	34.1	40.0
	24	34.4	40.2
	30	**34.6**	**40.3**

QA performances of RINK (M = 15, α = 0) with varied K values

Ablation Study: Effect of Varying α for RINK

The performances are relatively high in the range of [0.4, 0.7] for α, where α = 0.7 shows the best performance.



Performance curve of the reranking by RINK with M = 15 in terms of R@15, varying α

Ablation Study: Effect of Pretraining via Data Augmentation

The effect of the pretraining on QA performance is more dominant than that on retrieval performance, leading to increases of about 1.5% over nonpretraining runs both at EM and F1, under both of FiD and RINK.

Model	R@1	R@5	R@10	R@15
Retriever (w/ pre-train)	**52.2**	**70.2**	76.6	**79.8**
Retriever (w/o pre-train)	51.0	66.2	76.6	79.6

Retrieval performance of baseline retriever on development set with and without pretraining.

Model	Pretraining	EM	F1
FiD	✗	28.5	34.4
	✓	**30.1**	**36.0**
RINK ($M = 10, \alpha = 0.7$)	✗	33.8	40.0
	✓	**35.5**	**41.5**

QA performances of FiD and RINK on blind test set with and without pretraining.

Case Studies

The correct cases where their reasoning types are text → table in B2 (a) and table → text in B1 ((b)).

(a)

Q: This 70 's Kishore Kumar song was in a film produced by Alankar Chitra and directed by Shanker Mukherjee ?
B1 (Kishore_Kumar_0): [TITLE] Kishore Kumar [SECTITLE] Awards [TABLE] Year is 1970. Song is Roop Tera Mastana. Film is Aradhana. Music director is Sachin Dev Burman. Lyricist is Anand Bakshi. ⋯
B2 (Kishore_Kumar_1): [TITLE] Kishore Kumar [SECTITLE] Awards [TABLE] Year is 1975. Song is Main Pyaasa Tum. Film is Faraar. Music Director is Kalyanji Anandji. Lyricist is Rajendra Krishan. [PASSAGE] ⋯ The film is produced by Alankar Chitra and directed by Shanker Mukherjee.
Gold Table Id: Kishore_Kumar_1
Generated Answer: Main Pyaasa Tum ✓
Reference Answer: Main Pyaasa Tum

(b)

Q: What position does 2009-10 season Vancouver Canucks player Rob Davison currently hold with the Toronto Marlies ?
B1 (2009-10_Vancouver_Canucks_season_14): [TITLE] 2009-10 Vancouver Canucks season [SECTITLE] Free agents lost [TABLE] Player is Rob Davison. New team is New Jersey Devils. Contract terms is . [PASSAGE] ⋯ He is currently serving as assistant coach of the Toronto Marlies ⋯
B2 (2009-10_Vancouver_Canucks_season_15): [TITLE] 2009-10 Vancouver Canucks season [SECTITLE] Draft picks [TABLE] Round is 2. # is 53. Player is Anton Rodin (RW). Nationality is Sweden. College/junior/club team (League) is Brynas IF Jr. (J20 SuperElit) ⋯
Gold Table Id: 2009-10_Vancouver_Canucks_season_15
Generated Answer: assistant coach ✓
Reference Answer: assistant coach

Case Studies

(c): the case with the retrieval error where the gold table segment is not appear in the top M retrieved blocks, while its reasoning path seems to be ended with B6 after trials of question matching

(c)

Q: What is Spain 's oldest sporting club solely devoted to football with a 2014-15 Fenerbahçe S.K . season result F-A of 0-2 ?
B1 (2014-15_Fenerbahçe_S.K._season_7): [TITLE] 2014-15 Fenerbahçe S.K. season [SECTITLE] Pre-season friendlies [TABLE] Date is 16 August 2014. Opponents is Olympiacos. Stadium is Şükrü Saracoğlu Stadium. Result F-A is 2-1. ⋯
B2 (2014-15_Fenerbahçe_S.K._season_7): [TITLE] 2014-15 Fenerbahçe S.K. season [SECTITLE] Pre-season friendlies [TABLE] Date is 8 August 2014. Opponents is Chelsea. Stadium is Şükrü Saracoğlu Stadium. Result F-A is 0-2. Attendance is 21,300. ⋯
B6 (2014-15_FC_Barcelona_season_8): [TITLE] 2014-15 FC Barcelona season [SECTITLE] Competitions -- Score overview [TABLE] Opposition is Real Madrid. Home score is 2-1. Away score is 1-3. Double is 3-4. ⋯ Real Madrid is one of three founding members of La Liga that have never been relegated from the top division since its inception in 1929 , along with Athletic Bilbao and Barcelona .
Gold Table Id: 2014-15_Fenerbahçe_S.K._season_7
Generated Answer: Barcelona
Reference Answer: Sevilla Fútbol Club

(d): the case with the numerical reasoning error where both relevant blocks B1 and B14 are successfully retrieved, while the answer extraction is failed to precisely perform the numerical reasoning that selects the cell with the highest gloss among B1 and B14.

(d)

Q: Is 20th Century Fox or Walt Disney Pictures the creative studio behind the most highest grossing film of Singapore ?
B1 (List_of_highest-grossing_films_in_Singapore_4): [TITLE] List of highest-grossing films in Singapore [SECTITLE] Top-grossing films in 2006 [TABLE] Rank is 1. Title is X-Men : The Last Stand. Studio is 20th Century Fox. Gross is $ 4.8m. ⋯
B14 (List_of_highest-grossing_films_in_Singapore_8): [TITLE] List of highest-grossing films in Singapore [SECTITLE] Top-grossing films of all time [TABLE] Rank is 2. Title is Avengers : Infinity War. Studio is Walt Disney Pictures. Lifetime Gross is S $ 16.22mil. Year is 2018.
Gold Table Id: List_of_highest-grossing_films_in_Singapore_16
Generated Answer: 20^{th} Century Fox
Reference Answer: Walt Disney Pictures

Summary & Conclusion

- We propose RINK, a novel set-level reranking method that reuses the reader's module without any modification of the reader's module.
- We present pre-training method for the initial retriever based on tabular-and-text entailment and cross-modal masked language modeling tasks with data augmentation.
- Experiments results on OTT-QA shows that RINK leads to state-of-the-art performance.
- Results confirm that the retrieval step is the key component for improving the QA performance.
- Results also suggest us to further investigate the Retriever-Reranker-Reader framework as a promising approach to table-and-text ODQA.

Future Work

- Extend the set-level RINK by using the cross-attention scores of (Izacard and Grave 2021a) as an additional relevance signal.

- Pursue a reranker-aware joint learning framework of Retriever-Reranker-Reader for table-and-text ODQA
  - By extending REALM (Guu et al. 2020) and RAG (Lewis et al. 2020c), we would like to establish an end-to-end learning framework of Retriever-Reranker-Retriever for table-and-text ODQA and explore data augmentation methods directly to train all the components in the framework in a joint manner

Thank you!

Appendix