Question Answering from Tables

Eli Sage-Martinson
CPSC 677: Advanced Natural Language Processing
12/17/2021

Outline

- Problem Definition
- Motivation
- Problem Variations
- Datasets
- Current Approaches
  - TAPAS
  - TABERT
- Paper 1: Open Domain Question Answering over Tables via Dense Retrieval
- Paper 2: MultimodalQA: Complex Question Answering over Tables
- Optional Papers
- TAPAS demo

Problem Definition

Q: What named native languages spoken in the warsaw governorate have more males then females?

Language	Number	percentage (%)	males	females
Polish	1 420 436	73.52	687 210	733 226
Yiddish	317 169	16.41	154 603	162 566
Russian	87 850	4.54	13 551	1 586
German	77 160	3.99	37 984	39 176
Ukrainian	15 930	0.82	15 623	307
Romanian	2 299	>0.01	2 293	6
Latvian	1 759	>0.01	1 738	21
Estonian	1 566	>0.01	1 555	11
Tatar	1 473	>0.01	1 437	36
Belarusian	1 343	>0.01	1 234	109
Other	4 824	0.24	3 289	1 535
Persons that didn't name their native language	54	>0.01	33	21
Total	1 931 867	100	977 948	953 919

A: 'Russian', 'Ukrainian', 'Romanian', 'Latvian', 'Estonian', 'Tatar', 'Belarusian'

Source: SQA dataset

Motivation

- Natural Language Interface to Database (NLIDB)
  - Replace complex and time-consuming structured queries (eg. SQL) with natural language queries.
  - Allows database queries to be performed more quickly and without technical knowledge.
- Multi-modal Question Answering
  - Sometimes the information we need to answer a question can only be found in a table.
  - Ex: Answering a question based on a Wikipedia article.

Problem Variations

- Structured Queries vs End-to-end
  - Structured Queries: model produces a structured program (ie. a SQL query) as output.
  - End-to-end: Model selects the output cell(s) (and a possible aggregation operation) with no intermediate program step.
- Strongly vs. weakly supervised:
  - Strongly Supervised: Model is trained on (question, program) pairs.
  - Weakly Supervised: Model is trained on (question, final answer) pairs.

Discussion Question: If you had to guess, what advantages/disadvantages might each of these approaches have compared to the other?

Datasets

- **WikiTableQuestions**: a crowdsourced set of complex questions, including aggregation operations, over Wikipedia tables.
- **SequentialQA**: a subset of highly compositional questions from WikiTableQuestions, broken down into series of ~3 queries.
- **WikiSQL**: Hand-generated natural language and SQL pairs.
- Stanford Question Answering Dataset (**SQUAD**): questions posed by crowdworkers over Wikipedia tables, answers are spans from the article.
- **SPIDER**: text-to-SQL database annotated by Yale students.

Current Approach: TAPAS

- TAPAS: Weakly Supervised Table Parsing via Pre-training
- TAPAS is a weakly supervised, end-to-end model from Google Research (2019) based on the BERT encoder.
- Demonstrates one approach to linearizing a 2D table for input to BERT.

TAPAS, cont.

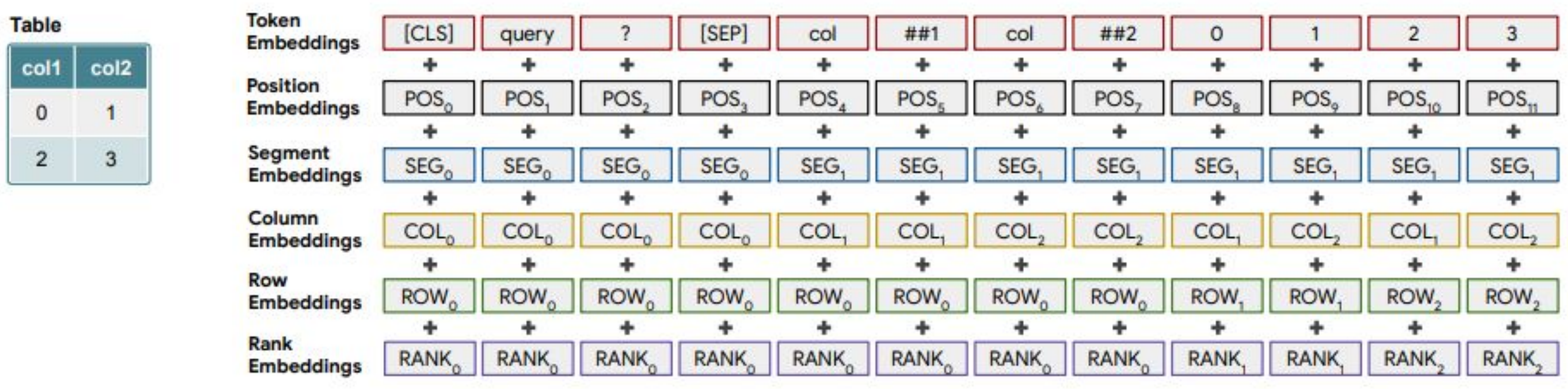
Example questions

#	Question	Answer	Example Type
1	*Which wrestler had the most number of reigns?*	Ric Flair	Cell selection
2	*Average time as champion for top 2 wrestlers?*	AVG(3749,3103)=3426	Scalar answer
3	*How many world champions are there with only one reign?*	COUNT(Dory Funk Jr., Gene Kiniski)=2	Ambiguous answer
4	*What is the number of reigns for* Harley Race?	7	Ambiguous answer
5	*Which of the following wrestlers were ranked in the bottom 3?*	{Dory Funk Jr., Dan Severn, Gene Kiniski}	Cell selection
	Out of these, who had more than one reign?	Dan Severn	Cell selection

Source: [TAPAS: Weakly Supervised Table Parsing via Pre-training]

TAPAS, cont.

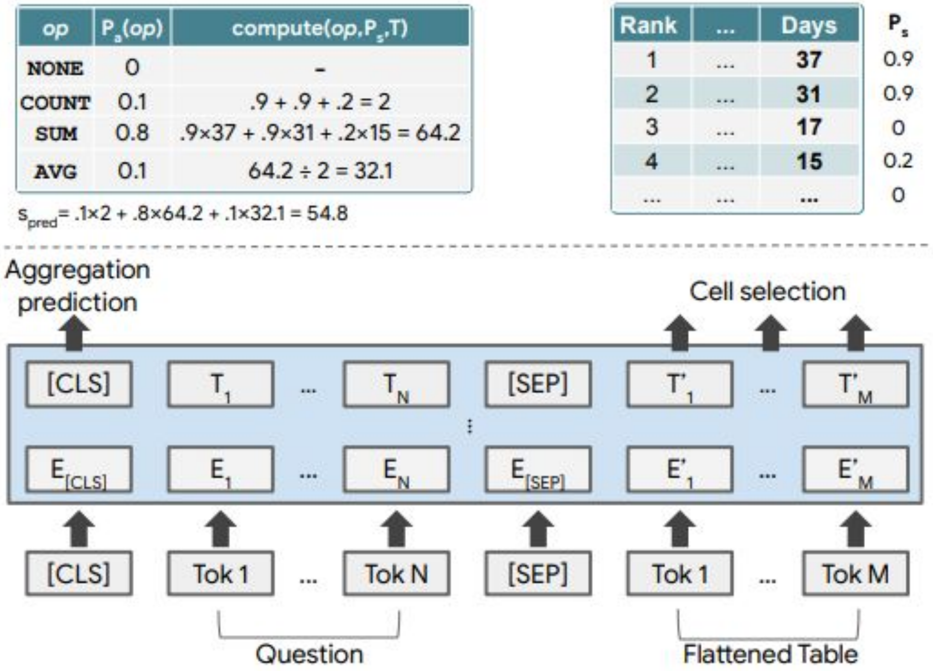
- Input is <question> + [SEP] + <flattened table>
- Additional positional embeddings are added the token-embeddings to account for table structure.



Source: [TAPAS: Weakly Supervised Table Parsing via Pre-training]

TAPAS, cont.

- An additional two classification layers are added to select table cells and an aggregation operation.



Source: [TAPAS: Weakly Supervised Table Parsing via Pre-training]

TAPAS, cont.

- TAPAS achieves competitive performance on **WikiSQL** and outperforms previous models on **SQA.**

Model	Dev	Test
Liang et al. (2018)	71.8	72.4
Agarwal et al. (2019)	74.9	74.8
Wang et al. (2019)	79.4	79.3
Min et al. (2019)	84.4	**83.9**
TAPAS	**85.1**	83.6
TAPAS (fully-supervised)	88.0	86.4

Table 3: WIKISQL denotation accuracy[4].

Model	ALL	SEQ	Q1	Q2	Q3
Pasupat and Liang (2015)	33.2	7.7	51.4	22.2	22.3
Neelakantan et al. (2017)	40.2	11.8	60.0	35.9	25.5
Iyyer et al. (2017)	44.7	12.8	70.4	41.1	23.6
Sun et al. (2018)	45.6	13.2	70.3	42.6	24.8
Müller et al. (2019)	55.1	28.1	67.2	52.7	46.8
TAPAS	**67.2**	**40.4**	**78.2**	**66.0**	**59.7**

Table 5: SQA test results. ALL is the average question accuracy, SEQ the sequence accuracy, and QX, the accuracy of the X'th question in a sequence.

Source: TAPAS: Weakly Supervised Table Parsing via Pre-training

Discussion Question

- Are there any types of questions that the cell selection + optional aggregator scheme cannot answer?

Current Approaches: TABERT

- TaBERT: Pretraining for Joint Understanding of Textual and Tabular Data
- Another question answering from tables model based on the BERT encoder.
- TABERT takes another approach to the fundamental problem of how to linearize a 2D table for input to BERT.

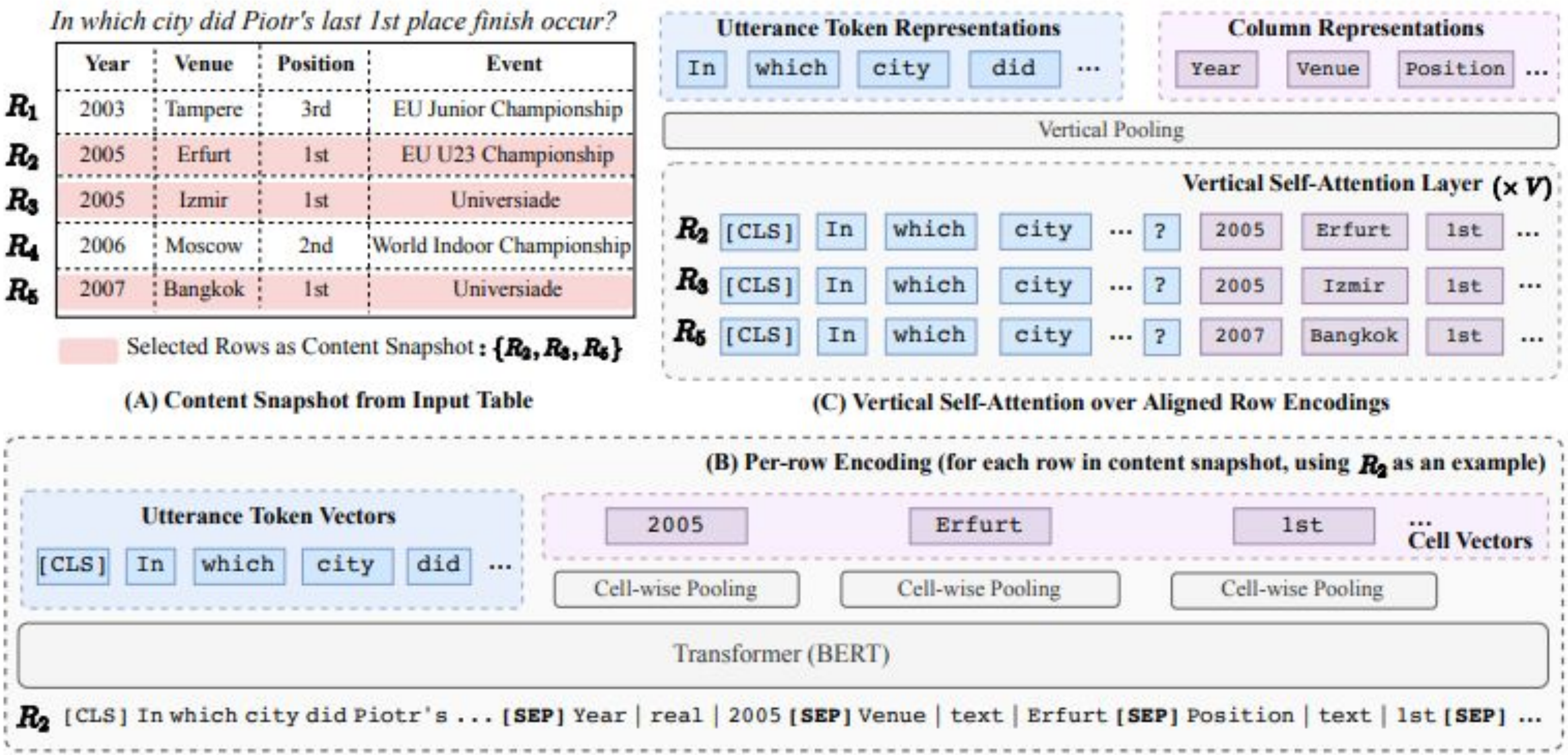
TABERT, cont.

- Input: an utterance *u* and a table *T*.
- Step 1: select a *snapshot* of the table data most relevant to *u*.
  - Select top *K* rows in the input table that have the highest *n*-gram overlap ratio with *u*.
  - For *K* = 1, create a synthetic row: for each column, select the cell that has the most n-gram overlap with *u*.
- Step 2: Row Linearization
  - Each cell in the snapshot is linearized as follows:

$$\underbrace{\texttt{Year}}_{\text{Column Name}} \;\; | \;\; \underbrace{\texttt{real}}_{\text{Column Type}} \;\; | \;\; \underbrace{\texttt{2005}}_{\text{Cell Value}}$$

  - Each cell in the row is concatenated and then the row is prefixed with *u*. This is given as input to the encoder.

Source: [TaBERT: Pretraining for Joint Understanding of Textual and Tabular Data]



Source: [TaBERT: Pretraining for Joint Understanding of Textual and Tabular Data]

TABERT, cont.

- Step 3: Vertical Self-Attention
  - Base transformer outputs vector representation of utterance and cell tokens for each row.
  - A vertical self-attention layer operates on vertically aligned elements of the output (utterance and cell vectors that correspond to the same question token or column).
  - This allows TABERT to capture cross-row dependencies on cell values.
- Step 4: Utterance and Column Representations
  - A representation **c_j** is computed for each **column**, and a representation **x_j** is computed for each **utterance token**.
  - These representation can be used for a variety of downstream tasks.

TABERT, cont.

- TABERT is pretrained on a corpus of web tables.
- Task 1: Masked Column Prediction: predict name / data type of masked columns.
- Task 2: Predict contents of masked cells in the content snapshot.

TABERT Applications

- Supervised Semantic Parsing: A general purpose semantic parser built on top of TABERT achieved competitive performance on the SPIDER dataset (text-to-SQL).
- Semi-Supervised Semantic Parsing: A general purpose semantic parser built on top of TABERT achieved new best performance on the WikiTableQuestions dataset (weakly supervised question answering).
- Even simple systems built with TABERT match or outperform more specialized systems built with plain BERT.
- It seems TABERT is a better representation layer for tabular tasks than BERT.

Previous Systems on WikiTableQuestions				
Model	Dev		Test	
Pasupat and Liang (2015)	37.0		37.1	
Neelakantan et al. (2016)	34.1		34.2	
Ensemble 15 Models	37.5		37.7	
Zhang et al. (2017)	40.6		43.7	
Dasigi et al. (2019)	43.1		44.3	
Agarwal et al. (2019)	43.2		44.1	
Ensemble 10 Models	–		46.9	
Wang et al. (2019b)	43.7		44.5	
Our System based on MAPO (Liang et al., 2018)				
	Dev	Best	Test	Best
Base Parser†	42.3 ±0.3	42.7	43.1 ±0.5	43.8
w/ BERT$_{Base}$ (K = 1)	49.6 ±0.5	50.4	49.4 ±0.5	49.2
– content snapshot	49.1 ±0.6	50.0	48.8 ±0.9	50.2
w/ TaBERT$_{Base}$ (K = 1)	51.2 ±0.5	51.6	50.4 ±0.5	51.2
– content snapshot	49.9 ±0.4	50.3	49.4 ±0.4	50.0
w/ TaBERT$_{Base}$ (K = 3)	51.6 ±0.5	52.4	51.4 ±0.3	51.3
w/ BERT$_{Large}$ (K = 1)	50.3 ±0.4	50.8	49.6 ±0.5	50.1
w/ TaBERT$_{Large}$ (K = 1)	51.6 ±1.1	52.7	51.2 ±0.9	51.5
w/ TaBERT$_{Large}$ (K = 3)	**52.2** ±0.7	**53.0**	**51.8** ±0.6	**52.3**

Top-ranked Systems on Spider Leaderboard		
Model	Dev. Acc.	
Global–GNN (Bogin et al., 2019a)	52.7	
EditSQL + BERT (Zhang et al., 2019a)	57.6	
RatSQL (Wang et al., 2019a)	60.9	
IRNet + BERT (Guo et al., 2019)	60.3	
+ Memory + Coarse-to-Fine	61.9	
IRNet V2 + BERT	63.9	
RyanSQL + BERT (Choi et al., 2020)	**66.6**	
Our System based on TranX (Yin and Neubig, 2018)		
	Mean	Best
w/ BERT$_{Base}$ (K = 1)	61.8 ±0.8	62.4
– content snapshot	59.6 ±0.7	60.3
w/ TaBERT$_{Base}$ (K = 1)	63.3 ±0.6	64.2
– content snapshot	60.4 ±1.3	61.8
w/ TaBERT$_{Base}$ (K = 3)	63.3 ±0.7	64.1
w/ BERT$_{Large}$ (K = 1)	61.3 ±1.2	62.9
w/ TaBERT$_{Large}$ (K = 1)	64.0 ±0.4	64.4
w/ TaBERT$_{Large}$ (K = 3)	**64.5** ±0.6	**65.2**

Source: TaBERT: Pretraining for Joint Understanding of Textual and Tabular Data

Discussion Question

- Beyond performance enhancements, what are other advantages TABERT might have over systems like TAPAS?

[Open Domain Question Answering over Tables via Dense Retrieval]

Jonathan Herzig, Thomas Müller, Syrine Krichene, Julian Martin Eisenschlos
NAACL 2021

Open-Domain Question Answering

- Instead of assuming the table is given with the question, the model must select the best table from a corpus.
- More closely approximates tasks such as web search.
- Naively applying a QA model over every table in the corpus is prohibitively expensive.
- Most systems rely on a two-step approach:
  - A retriever model to select the appropriate table.
  - A QA model which outputs the answer.

Dataset

- There were previously no open-domain datasets for QA from tables.
- Use dataset **Natural Questions** as a base.
  - NQ questions are drawn from real Google searches.
  - NQ answers are spans of Wikipedia articles selected by annotators.
- The authors identify 12k examples where the answer resides in a table to produce **NQ-Tables,** which is comprised of (question, table, answer) triples.
- The set of all tables in **NQ-Tables** triples makes up the open domain corpus.

Dense Table Retrieval

- Dense Table Retrieval (DTR) retrieves 10 candidate tables from the corpus given a question *q*.
- Basic idea: compute a dense vector for each table that can be compared with a vector for the question.
- Makes use of TAPAS's ability to encode both questions and tables.

$$h_q = \mathbf{W_q}\mathrm{TAPAS_q}(q)\,\texttt{[CLS]}$$

$$h_T = \mathbf{W_T}\mathrm{TAPAS_T}(\texttt{title}(T), T)\,\texttt{[CLS]}$$

$$\mathrm{S_{ret}}(q, T) = h_q^T h_T,$$

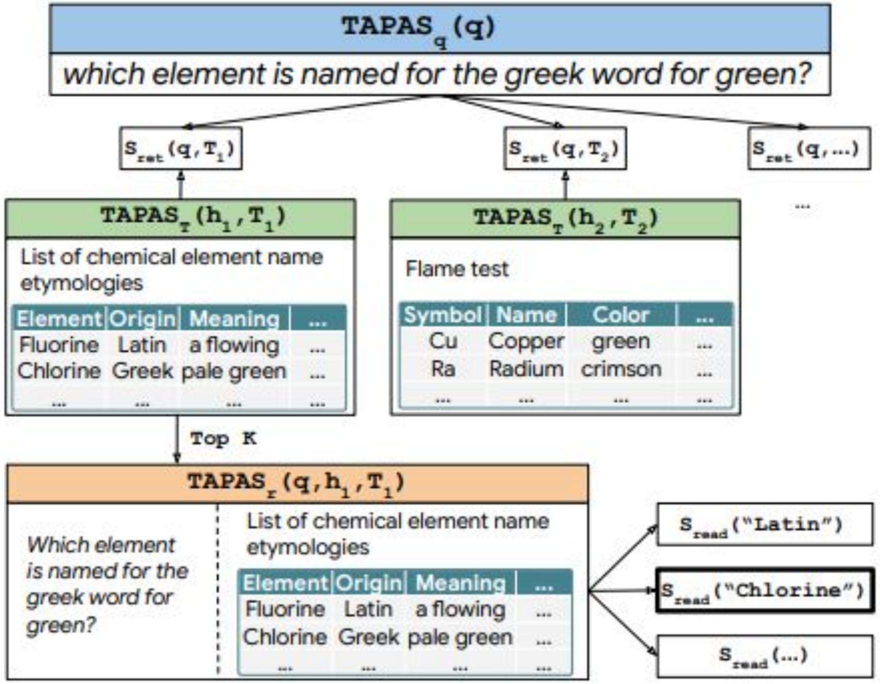
Source: Open Domain Question Answering over Tables via Dense Retrieval

Dense Table Retrieval, cont.

- Retriever is pre-trained on a variant of the Inverse Cloze Task (ICT): given a sentence *s*, predict the paragraph that contained *s* (*s* is masked).
- In this case, retriever tries to predict which Wikipedia text-span originally included a table *t* (*t* is masked).
- The retriever is then trained on (question, table) pairs from **NQ-Tables.**

Question Answering from Tables

- Once the candidate tables have been selected, TAPAS is applied to each of them.
- The answer with the highest score is selected and returned.



Source: Open Domain Question Answering over Tables via Dense Retrieval

Discussion Questions

- How well does the open-domain table task described in this paper reflect real-life QA tasks (for example, answering user questions in web searches)?
- **NQ-Tables** uses questions that are answered by highlighted spans of Wikipedia articles. What might be advantages and disadvantages of this approach?

MultimodalQA: Complex Question Answering over Tables

Alon Talmor, Alon Talmor, Ori Yoran, Amnon Catav, Dan Lahav, Yizhong Wang, Akari Asai, Gabriel Ilharco, Hannaneh Hajishirzi, Jonathan Berant

ICLR 2021

Multimodal Question Answering

- Combine a number of information sources (text, tables, images) to answer a question.
- Answer typically requires performing multihop reasoning: using the answer to one or more sub-questions as input to the next question.

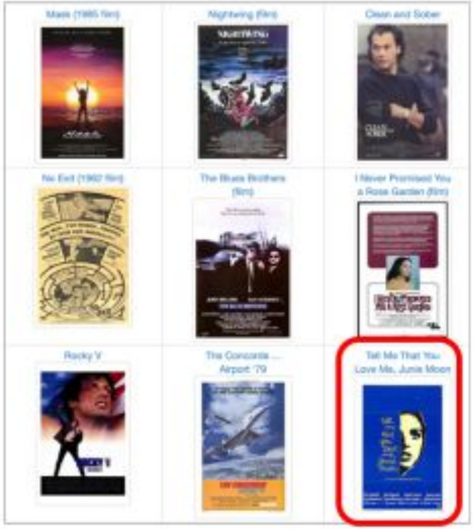
Steal This Movie!
The film follows Hoffman's (D'Onofrio) relationship with his second wife Anita (Garofalo) and their "awakening" and subsequent conversion to an activist life. The title of the film is a play on Hoffman's 1970 counter-culture guidebook titled "Steal This Book".

La liceale
La liceale (internationally released as The Teasers, "Under-graduate Girls", "Sophomore Swingers" and "Teasers") is a 1975 commedia sexy all'italiana directed by Michele Massimo Tarantini.
...
Guida. It was followed by "La liceale nella classe dei ripetenti".

Sage Stallone
Stallone made his acting debut alongside his father in Rocky V (1990), the fifth installment of the Rocky franchise, playing Robert Balboa Jr., the onscreen son of his father's title character. He did not, however,
...
After that, he acted in lesser profile films.

Pierino contro tutti
Pierino contro tutti (also known as "Desirable Teacher") is a 1981 comedy film directed by Marino Girolami. The main character of the film is Pierino, an
...
l as a short lived subgenre of joke-films in which the plot basically consists of a series of jokes placed side by side.

Year	Title	Role
1957	A Dangerous Age	David
1959	The Hanging Tree	Rune
1962	No Exit	Camarero
1970	Tell Me That You Love Me, Junie Moon	Jesse
1972	The Outside Man	Desk Clerk
...	...	...
1985	Mask	Mr. Simms
1988	Clean and Sober	Kramer
1990	Rocky V	Doctor
1991	Guilty by Suspicion	Darryl Zanuck



Q: Which B. Piazza title came earlier: the movie S. Stallone's son starred in or the movie with half of a lady's face on the poster?
A: Tell Me That You Love Me, Junie Moon

Source: [MultiModalQA: Complex Question Answering over Text, Tables and Images]

Dataset

- The paper introduces **MMQA**, a multimodal QA dataset.
- Tables are harvested from Wikipedia, then linked to text and images in existing datasets.
- Single modality questions are devised for each of the tables, texts, and images.
- The linked structure of the table/text/image data is used to automatically link the single-modality questions to produce multi-modal pseudo-language questions.
- Paid annotators paraphrase the pseudo-language questions into more natural text.

Type	Q&A	%
TextQ	*What was the territorial capital of the territory opposing Ohio in the Toledo War? Detroit*	31.0
TableQ	*Does the German state Baden-Wurttemberg or Thuringia have more residents? Baden-Württemberg*	18.3
ImageQ	*What weapon is the statue in Nottingham holding? bow*	8.9
Compose(TextQ,TableQ)	*At what age did the Cleveland Cavaliers player with 6190 rebounds enter the NBA? 19*	7.8
ImageListQ	*What is the common name of the bush warbler in Thailand that has an orange stripe above its eye? Chestnut-crowned bush warbler*	6.1
Compose(TableQ,ImageListQ)	*The film that starred Chris Ellison where a man was holding a newspaper on the poster, was released what year? 1988*	5.4
Compose(ImageQ,TableQ)	*On the poster for the TV show in which Tom Mison played Dorian Crane, what kind of structure can be seen behind the two men? castle*	4.5
Compare(Compose(TableQ,ImageQ),TableQ)	*Which manufacturer has fewer wins at the First Data 500: Buick or the brand with a cross for a logo? Buick*	3.5
Compose(TableQ,TextQ)	*On what date did the original artist who sang Sweet Child of Mine have a concert at US Bank Stadium? July 30, 2017*	3.2
Intersect(TableQ,TextQ)	*Who was the artist for Damon Fox in 2006 who also sings "You got the moves like Jagger"? Christina Aguilera*	2.6
Compose(TextQ,ImageListQ)	*On the poster for the movie based on the book "Act like a Lady, Think Like a Man," how many people are there in total? nine*	2.4
Intersect(ImageListQ,TableQ)	*What covers of the Chandler Canterbury films from 2009 has more than one person? Powder Blue Balls Out, Gary the Tennis Coach, After.Life*	2.3
Compare(TableQ,Compose(TableQ,TextQ))	*Did Chelsea or club that sings You'll Never Walk Alone rank higher in Deloitte Football Money League 2007? Chelsea*	2.1
Compose(ImageQ,TextQ)	*Did Gary Oldman take part in the movie whose poster features two men holding handguns, and which had Mark L. Smith as a writer? no*	1.0
Compare(Compose(TableQ,ImageQ), Compose(TableQ,TextQ))	*Was the film that features a giant eye on its poster or the first Wolverine movie the earlier film that Scott Silver worked on? Requiem for a Dream*	0.8
Intersect(ImageListQ,TextQ)	*What common law state with an eagle on the flag has an institution in the North region of Division II of the NCCAA? Iowa*	0.2

Source: [MultiModalQA: Complex Question Answering over Text, Tables and Images]

QA Models

- The authors train single-modality models for each modality (images, tables, and text).
  - Text QA Module: Based on RoBERTa-Large, selects a span from the paragraph to answer the question.
  - Table QA Module: Based on TAPAS.
  - Image QA Module: Based on VILBERT-MT, predicts answer from (question, image) pair.

QA Models

- Multi-hop answering is performed by *ImplicitDecomp*:
  - First use a classifier to predict the question type (ie. Compose(TableQ, TextQ).
  - At each hop, the model is given (question, question type, hop number, context, answers from previous hops).
  - Note that the question is not explicitly decomposed into parts—decomposition is performed *implicitly.*
- A baseline comparison is provided by a model called *AutoRouting*: after predicting the question type, hand the entire question to the single-modality model for the last hop.
  - Ex: Compose(TextQ, ListImageQ) is given directly to the image QA module.

	Single Modality		Mutli Modality		All	
	EM	F_1	EM	F_1	EM	F_1
Question-only[2]	14.2	17.0	16.9	19.5	15.3	18.0
Context-only	8.0	10.2	6.6	8.5	7.4	9.5
AutoRouting	48.9	57.1	32.0	38.2	42.1	49.5
ImplicitDecomp	**51.1**	**58.8**	**46.5**	**51.7**	**49.3**	**55.9**
Human	87.9	92.5	84.8	90.1	86.2	91.2

Table 4: Test set results

Source: [MultiModalQA: Complex Question Answering over Text, Tables and Images]

Discussion Questions

- In your opinion, does *ImplicitDecomp* achieve true multi-modal reasoning? Is the framework used by that model (predicting question types) feasible in a real-world context?
- What other sources might we use to construct a multi-modal reasoning dataset?

Additional Paper

[CLTR: An End-to-End, Transformer-Based System for Cell Level Table Retrieval and Table Question Answering]

Feifei Pan , Mustafa Canim, Michael Glass, Alfio Gliozzo, Peter Fox
ACL 2021

Transformer-based End-to-End Table QA system

- The paper deals with open domain question answering, like previous Dense Table Retrieval paper.
- First, inexpensive BM35 algorithm is used to generate a pool of relevant tables from the corpus.
- Then row-column intersection (RCI) model is used to generate scores for every column and row in the table.
  - The score for each table is taken as the maximum cell-level score:
  - P_t = MAX(P_col) + MAX(P_row)
- Once ranking is complete, the top *k* tables are returned to users, with the maximum-likelihood cell highlighted.

Dataset Generation

- E2E_GNQ: created from GNQTables, which is a open-domain table retrieval dataset which marks each table as relevant (1) or not (0) to a given question.
- Each table in GNQTables is additionally annotated with the correct table cell(s), the correct row ID, and the correct column ID.
- Evaluate on Precision @ k, Normalized Discounted Gain @ k, and Mean Average Precision (MAP).

Results

- CLTR outperforms baselines on the ESE_GNQ dataset.
- Qualitative: many of CLTR's errors are caused by Wikipedia Infoboxes, which are "noisy" tables without a clear row / column structure.

	P@5	P@10	N@5	N@10	N@20	MAP
BM25	0.0413	0.0242	0.1650	0.1764	0.1852	0.1601
MTR_{point}	0.1460	0.0767	0.6227	0.6349	0.6359	0.5920
MTR_{pair}	0.1826*	0.0990*	0.6945*	0.7198*	0.7220*	0.6328*
CLTR	**0.2203**	**0.1660**	**0.7235**	**0.7402**	**0.7458**	**0.7176**

(b) E2E_GNQ

Source: CLTR: An End-to-End, Transformer-Based System for Cell Level Table Retrieval and Table Question Answering

Discussion Question

- Which approach to open-domain question answering is more useful: returning a set of relevant tables with likely cells highlighted (CLTR), or simply returning the proposed answer to the question (DTR)?

Additional Paper

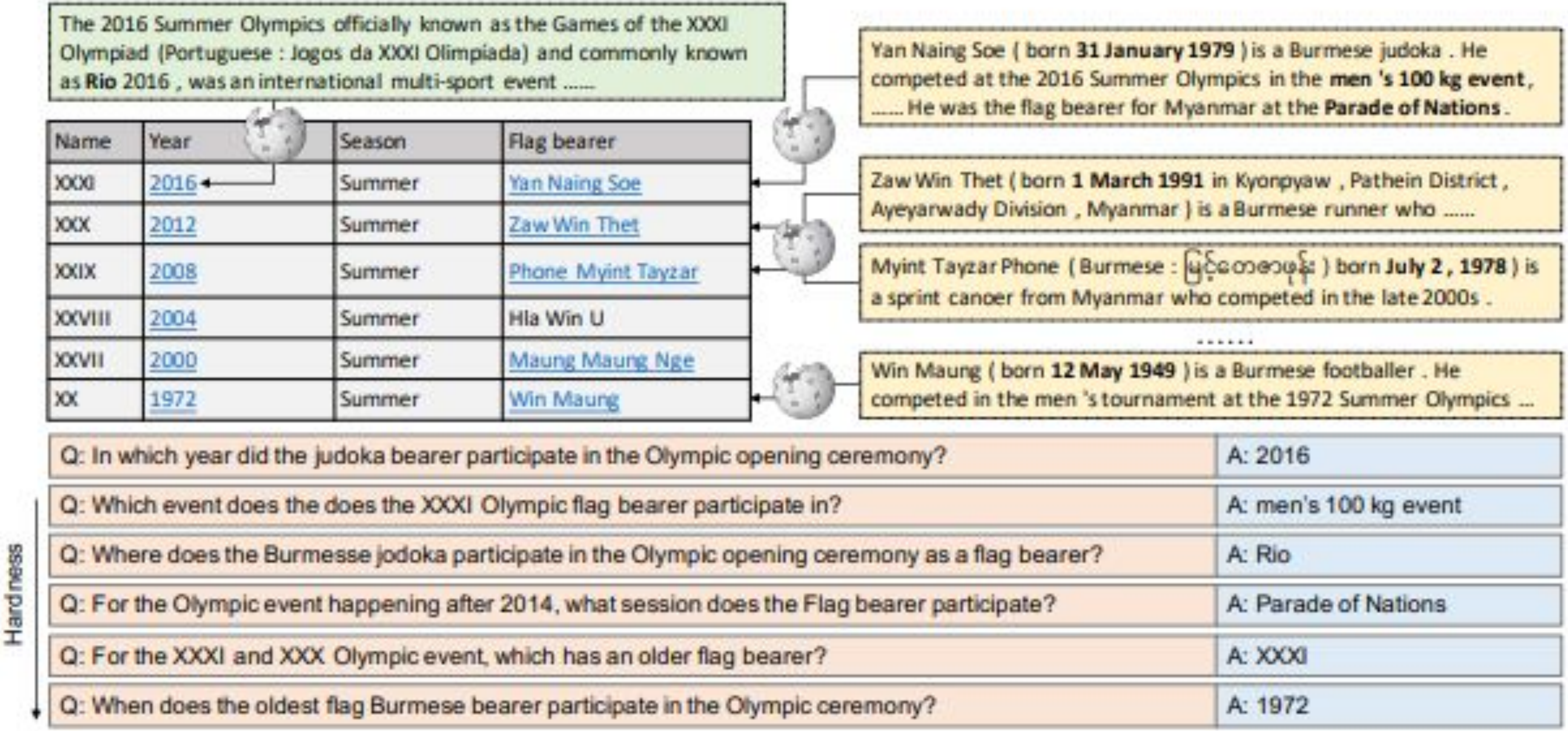
HybridQA: A Dataset of Multi-Hop Question Answering over Tabular and Textual Data

Wenhu Chen, Hanwen Zha, Zhiyu Chen, Wenhan Xiong, Hong Wang, William Wang

ACL 2020

Hybrid Dataset

- This paper constructs HybridQA, a question answering dataset over heterogeneous (multimodal) data.
- Human annotators are presented with Wikipedia tables and associated hyperlinked pages.
- Annotators asked to produce questions requiring aggregating information from both sources.
- Deal with multiple sources of bias:
    - Portions of the table are randomly highlighted to encourage annotators to select rows uniformly.
    - Use algorithm to detect and reject questions which have their text spans clustered too closely in the first few sentences.
    - Use human experts to detect “fake hybrid” questions which don’t actually require multi-hop reasoning.



Source: HybridQA: A Dataset of Multi-Hop Question Answering over Tabular and Textual Data

Discussion Question

- What are advantages and disadvantages of this dataset (HybridQA) compared to the multimodal dataset we saw in the *ImplicitDecomp* paper?
- What do you think might be other sources of bias encountered when having human annotators write questions?

TAPAS Demo

- Try predictions on SQA here: https://colab.research.google.com/github/google-research/tapas/blob/master/notebooks/sqa_predictions.ipynb
- A sample of test question screenshots are on the following pages.

Pos	No	Driver	Team	Laps	Time/Retired	Grid	Points
1	32	Patrick Carpentier	Team Player's	87	1:48:11.023	1	22
2	1	Bruno Junqueira	Newman/Haas Racing	87	+0.8 secs	2	17
3	3	Paul Tracy	Team Player's	87	+28.6 secs	3	14
4	9	Michel Jourdain, Jr.	Team Rahal	87	+40.8 secs	13	12
5	34	Mario Haberfeld	Mi-Jack Conquest Racing	87	+42.1 secs	6	10
6	20	Oriol Servia	Patrick Racing	87	+1:00.2	10	8
7	51	Adrian Fernandez	Fernandez Racing	87	+1:01.4	5	6
8	12	Jimmy Vasser	American Spirit Team Johansson	87	+1:01.8	8	5
9	7	Tiago Monteiro	Fittipaldi-Dingman Racing	86	+ 1 Lap	15	4
10	55	Mario Dominguez	Herdez Competition	86	+ 1 Lap	11	3
11	27	Bryan Herta	PK Racing	86	+ 1 Lap	12	2
12	31	Ryan Hunter-Reay	American Spirit Team Johansson	86	+ 1 Lap	17	1
13	19	Joel Camathias	Dale Coyne Racing	85	+ 2 Laps	18	0
14	33	Alex Tagliani	Rocketsports Racing	85	+ 2 Laps	14	0
15	4	Roberto Moreno	Herdez Competition	85	+ 2 Laps	9	0
16	11	Geoff Boss	Dale Coyne Racing	83	Mechanical	19	0
17	2	Sebastien Bourdais	Newman/Haas Racing	77	Mechanical	4	0
18	15	Darren Manning	Walker Racing	12	Mechanical	7	0
19	5	Rodolfo Lavin	Walker Racing	10	Mechanical	16	0

```
> what were the drivers names?
Ryan Hunter-Reay, Sebastien Bourdais, Adrian Fernandez, Jimmy Vasser, Bruno Junqueira, Joel Camathias,
> of these, which points did patrick carpentier and bruno junqueira score?
22, 17
> who scored higher?
Patrick Carpentier
```

Pos	No	Driver	Team	Laps	Time/Retired	Grid	Points
1	32	Patrick Carpentier	Team Player's	87	1:48:11.023	1	22
2	1	Bruno Junqueira	Newman/Haas Racing	87	+0.8 secs	2	17
3	3	Paul Tracy	Team Player's	87	+28.6 secs	3	14
4	9	Michel Jourdain, Jr.	Team Rahal	87	+40.8 secs	13	12
5	34	Mario Haberfeld	Mi-Jack Conquest Racing	87	+42.1 secs	6	10
6	20	Oriol Servia	Patrick Racing	87	+1:00.2	10	8
7	51	Adrian Fernandez	Fernandez Racing	87	+1:01.4	5	6
8	12	Jimmy Vasser	American Spirit Team Johansson	87	+1:01.8	8	5
9	7	Tiago Monteiro	Fittipaldi-Dingman Racing	86	+ 1 Lap	15	4
10	55	Mario Dominguez	Herdez Competition	86	+ 1 Lap	11	3
11	27	Bryan Herta	PK Racing	86	+ 1 Lap	12	2
12	31	Ryan Hunter-Reay	American Spirit Team Johansson	86	+ 1 Lap	17	1
13	19	Joel Camathias	Dale Coyne Racing	85	+ 2 Laps	18	0
14	33	Alex Tagliani	Rocketsports Racing	85	+ 2 Laps	14	0
15	4	Roberto Moreno	Herdez Competition	85	+ 2 Laps	9	0
16	11	Geoff Boss	Dale Coyne Racing	83	Mechanical	19	0
17	2	Sebastien Bourdais	Newman/Haas Racing	77	Mechanical	4	0
18	15	Darren Manning	Walker Racing	12	Mechanical	7	0
19	5	Rodolfo Lavin	Walker Racing	10	Mechanical	16	0

```
> which driver completed the fewest laps?
Rodolfo Lavin
```

Pos	No	Driver	Team	Laps	Time/Retired	Grid	Points
1	32	Patrick Carpentier	Team Player's	87	1:48:11.023	1	22
2	1	Bruno Junqueira	Newman/Haas Racing	87	+0.8 secs	2	17
3	3	Paul Tracy	Team Player's	87	+28.6 secs	3	14
4	9	Michel Jourdain, Jr.	Team Rahal	87	+40.8 secs	13	12
5	34	Mario Haberfeld	Mi-Jack Conquest Racing	87	+42.1 secs	6	10
6	20	Oriol Servia	Patrick Racing	87	+1:00.2	10	8
7	51	Adrian Fernandez	Fernandez Racing	87	+1:01.4	5	6
8	12	Jimmy Vasser	American Spirit Team Johansson	87	+1:01.8	8	5
9	7	Tiago Monteiro	Fittipaldi-Dingman Racing	86	+ 1 Lap	15	4
10	55	Mario Dominguez	Herdez Competition	86	+ 1 Lap	11	3
11	27	Bryan Herta	PK Racing	86	+ 1 Lap	12	2
12	31	Ryan Hunter-Reay	American Spirit Team Johansson	86	+ 1 Lap	17	1
13	19	Joel Camathias	Dale Coyne Racing	85	+ 2 Laps	18	0
14	33	Alex Tagliani	Rocketsports Racing	85	+ 2 Laps	14	0
15	4	Roberto Moreno	Herdez Competition	85	+ 2 Laps	9	0
16	11	Geoff Boss	Dale Coyne Racing	83	Mechanical	19	0
17	2	Sebastien Bourdais	Newman/Haas Racing	77	Mechanical	4	0
18	15	Darren Manning	Walker Racing	12	Mechanical	7	0
19	5	Rodolfo Lavin	Walker Racing	10	Mechanical	16	0

```
> How many more laps did patrick carpentier complete than darren manning?
87
```

Pos	No	Driver	Team	Laps	Time/Retired	Grid	Points
1	32	Patrick Carpentier	Team Player's	87	1:48:11.023	1	22
2	1	Bruno Junqueira	Newman/Haas Racing	87	+0.8 secs	2	17
3	3	Paul Tracy	Team Player's	87	+28.6 secs	3	14
4	9	Michel Jourdain, Jr.	Team Rahal	87	+40.8 secs	13	12
5	34	Mario Haberfeld	Mi-Jack Conquest Racing	87	+42.1 secs	6	10
6	20	Oriol Servia	Patrick Racing	87	+1:00.2	10	8
7	51	Adrian Fernandez	Fernandez Racing	87	+1:01.4	5	6
8	12	Jimmy Vasser	American Spirit Team Johansson	87	+1:01.8	8	5
9	7	Tiago Monteiro	Fittipaldi-Dingman Racing	86	+ 1 Lap	15	4
10	55	Mario Dominguez	Herdez Competition	86	+ 1 Lap	11	3
11	27	Bryan Herta	PK Racing	86	+ 1 Lap	12	2
12	31	Ryan Hunter-Reay	American Spirit Team Johansson	86	+ 1 Lap	17	1
13	19	Joel Camathias	Dale Coyne Racing	85	+ 2 Laps	18	0
14	33	Alex Tagliani	Rocketsports Racing	85	+ 2 Laps	14	0
15	4	Roberto Moreno	Herdez Competition	85	+ 2 Laps	9	0
16	11	Geoff Boss	Dale Coyne Racing	83	Mechanical	19	0
17	2	Sebastien Bourdais	Newman/Haas Racing	77	Mechanical	4	0
18	15	Darren Manning	Walker Racing	12	Mechanical	7	0
19	5	Rodolfo Lavin	Walker Racing	10	Mechanical	16	0

```
> What was the average number of points scored by all drivers?
10, 0, 5, 22, 12, 0, 6, 2, 0, 3, 0, 0, 4, 14, 1, 8, 0, 0, 17
```