

TriviaQA: A Large Scale Distantly Supervised Challenge Dataset for Reading Comprehension

Mandar Joshi, Eunsol Choi, Daniel S. Weld, Luke Zettlemoyer

Presenter: Zhuolun Xiang

Background

- Question Answering (QA) Formulation
 - Answer a question q given evidences D
 - Dataset of tuples $\{(q_i, a_i, D_i) | i = 1, \dots, n\}$
 - a_i is a substring of D_i
- Example

Question: American Callan Pinckney's eponymously named system became a best-selling (1980s-2000s) book/video franchise in what genre?

Answer: Fitness

Excerpt: Callan Pinckney was an American fitness professional. She achieved unprecedented success with her Callanetics exercises. Her 9 books all became international best-sellers and the video series that followed went on to sell over 6 million copies. Pinckney's first video release "Callanetics: 10 Years Younger In 10 Hours" outsold every other **fitness** video in the US.

Overview

- TriviaQA
 - Over 650K question-answer-evidence triples
 - First dataset where questions authored by trivia enthusiasts
 - Evidence documents from Web search and Wiki pages
 - A high percentage of the questions are challenging
- Dataset samples

QuestionId	Question	Answer	Web	Wikipedia
qw_3199	Miami Beach in Florida borders which ocean?	Atlantic	Evidence	Evidence
bt_1255	What was the occupation of Lovely Rita according to the song by the Beatles	Traffic Warden	Evidence	Evidence
qg_77	Who was Poopdeck Pappys most famous son?	Popeye	Evidence	Evidence
wh_1026	The Nazi regime was Germany's Third Reich; which was the first Reich?	HOLY ROMAN EMPIRE	Evidence	Evidence

Question:

Miami Beach in Florida borders which ocean?

Answer: Atlantic

The left column contains Web search results. The right column contains question entity pages from Wikipedia.

Web search results

- [Dania Beach Station | U.S. Customs and Border Protection](#)
<https://www.cbp.gov/.../miami-sector-florida/dania-beach-station>
Miami Sector Florida; Dania Beach Station; ... The Dania Beach Border Patrol Station area of responsibility is the ... bordered by the Atlantic Ocean on the east and ...
- [Hotel Ocean, Miami Beach, FL - Booking.com](#)
www.booking.com/hotel/us/ocean-miami-beach.html
Facing the Atlantic Ocean on South Beach's famous Ocean Drive, this Miami hotel is less than a 2-minute walk from the Art Deco ... Fort Lauderdale Hollywood Airport ...
- [What states border the Atlantic Ocean? | Reference.com](#)
<https://www.reference.com/geography/states-border-atlantic-ocean-1...>
... Delaware, Florida, Georgia, Louisiana, Maine ... The states in the U.S. that border the Atlantic Ocean include ... Daytona Beach, West Palm Beach, Miami and ...
- [Miami Beach Vacation Rentals & Condo Rentals - Airbnb](#)
<https://www.airbnb.com/s/Miami-Beach--FL>
It was my birthday and I chose to spend it in South Beach Miami. I was looking for a beach ... the beach, Ocean ... Lincoln Road, Miami Beach Golf ...
- [Hotel Ocean – Hotel Ocean](#)
hotelocean.com
Escape to the Luxury of Our Miami Beach Oceanfront Hotel. The Hotel Ocean is situated ideally in ... The Hotel Ocean is in just the right spot to be convenient ...

Question entity pages

- [Miami Beach, Florida - Wikipedia](#)
https://en.wikipedia.org/wiki/Miami_Beach,_Florida
- [Florida - Wikipedia](#)
<https://en.wikipedia.org/wiki/Florida>

Dataset Collection

- Gather question-answer pairs from 14 trivia websites
 - Remove short questions
- Collect evidence from Web search and Wikipedia
 - Web search
 - Pose questions on Bing
 - Exclude trivia websites
 - Crawl top 10 results
 - Wikipedia
 - Use TAGME to find Wikipedia entities in the question
 - Add these pages as evidence

Computer Trivia

Created in 2009, what was the first decentralized cryptocurrency?

Show Answer

Difficulty Rating:



Total votes: 336



Comments

Next Question ›

Dataset Collection

- Gather question-answer pairs
 - Remove short questions
- Collect evidence from Web
 - Web search
 - Pose questions on Bing
 - Exclude trivia websites
 - Crawl top 10 results
 - Wikipedia
 - Use TAGME to find Wikipedia entities in the question
 - Add these pages as evidence

The screenshot shows a Bing search results page for the query "Created in 2009, what was the first decentralized cryptocurrency?". The search bar at the top contains the query. Below the search bar, there are tabs for "全部" (All), "图片" (Images), "视频" (Videos), "地图" (Maps), "资讯" (News), "商店" (Shopping), and "我的收藏项" (My Collections). The search results show 5,110,000 results. The first result is a snippet from Wikipedia titled "Cryptocurrency - Wikipedia" with a link to [en.wikipedia.org](https://en.wikipedia.org/wiki/Cryptocurrency). Below this, there are several other search results, including "The History Of Cryptocurrency — Steemit" and "Bitcoin | Bitcoin, created in 2009, was the first ...". On the right side of the page, there is a "Cryptocurrency" knowledge panel with a Bitcoin logo and a brief definition: "A cryptocurrency (or crypto currency) is a medium of exchange using cryptography to secure the transactions and to control the creation of new units. Cryptocurrencies are a subset of alternative currencies, or specifically of digital currencies. Bitcoin became the first decentralized cryptocurrency in 2009. Since then, numerous cryptocurrencies have been created." Below the knowledge panel, there is a section titled "其他用户还搜索过" (Other users also searched for) with links to "Bitcoin", "Digital currency", "Double-spending", "Namecoin", and "Complementary currency".

Dataset Collection

- Gather question-answer pairs from 14 trivia websites
 - Remove short questions
- Collect evidence from Web search and Wikipedia
 - Web search
 - Pose questions on Bing
 - Exclude trivia websites
 - Crawl top 10 results
 - Wikipedia
 - Use TAGME to find Wikipedia entities in the question
 - Add these pages as evidence

The screenshot displays the TAGME (Tagging Agent for Generic Mention Extraction) web interface. At the top, a text input field contains the question: "Created in 2009, what was the first decentralized cryptocurrency?". To the right of the input field is a vertical slider control labeled "Many links" at the top and "Few links" at the bottom, with a blue slider knob positioned in the middle. Below the slider are two buttons: "Reset" and "TAGME!". A green circular icon with a white 'G' is located at the bottom right of the input field. Below the input field, there are two tabs: "Tagged text" (which is selected) and "Topics". The "Tagged text" tab shows the same question with the phrase "first decentralized cryptocurrency" highlighted in blue, indicating it has been identified as a named entity.

Dataset Collection

- Gather question-answer pairs from 14 trivia websites
 - Remove short questions
- Collect evidence from Web search and Wikipedia
 - Web search
 - Pose questions on Bing
 - Exclude trivia websites
 - Crawl top 10 results
 - Wikipedia
 - Use TAGME to find Wikipedia entities in the question
 - Add these pages as evidence

Total number of QA pairs	95,956
Number of unique answers	40,478
Number of evidence documents	662,659
Avg. question length (word)	14
Avg. document length (word)	2,895

Dataset Analysis

- Question-answer pairs
 - Avg length = 14
 - Manually analyze 200 sampled questions

Property	Example annotation	Statistics
Avg. entities / question	Which politician won the Nobel Peace Prize in 2009?	1.77 per question
Fine grained answer type	What fragrant essential oil is obtained from Damask Rose?	73.5% of questions
Coarse grained answer type	Who won the Nobel Peace Prize in 2009?	15.5% of questions
Time frame	What was photographed for the first time in October 1959	34% of questions
Comparisons	What is the appropriate name of the largest type of frog?	9% of questions

Property of questions

Dataset Analysis

- Question-answer pairs
 - Avg length = 14
 - Manually analyze 200 sampled questions

Type	Percentage
Numerical	4.17
Free text	2.98
Wikipedia title	92.85
Person	32
Location	23
Organization	5
Misc.	40

Property of answers

Dataset Analysis

- Question-answer pairs
 - Avg length = 14
 - Manually analyze 200 sampled questions
- Evidences
 - 75.4%/79.7% of Web/Wiki evidences contain answers
 - Human test achieves 75.3/79.6 accuracy on Web/Wiki domains
- Answer 40% of questions needs information from multiple sentences

Experiments: Baseline Methods

- Random entity baseline (Wiki domain only)
 - Entities in Wiki pages form candidate answer set
 - Randomly pick one that not occur in question
- Entity classifier
 - Ranking problem over candidate answers
 - Function learnt using LambdaMART (Wu et al., 10)
- Neural model
 - Use BiDAF model (Seo et al., 17)

Experiments

- Metrics
 - Exact match(EM) and F1 score
 - For numerical and freeform answer: single given answer as ground truth
 - For Wiki entity: use Wiki aliases as well
- Setup
 - Random partition into train(80%)/development(10%)/test(10%)

		Train	Dev	Test
Wikipedia	Questions	61,888	7,993	7,701
	Documents	110,648	14,229	13,661
Web	Questions	76,496	9,951	9,509
	Documents	528,979	68,621	65,059
Wikipedia verified	Questions	-	297	584
	Documents	-	305	592
Web verified	Questions	-	322	733
	Documents	-	325	769

Experiments

- Results

Method	Domain	Distant Supervision						Verified					
		Dev			Test			Dev			Test		
		EM	F1	Oracle	EM	F1	Oracle	EM	F1	Oracle	EM	F1	Oracle
Random	Wiki	12.67	22.45	15.80	12.56	22.14	15.80	13.80	23.35	16.83	15.06	24.36	18.32
Classifier		23.42	27.67	71.32	22.45	26.51	71.60	23.57	27.93	78.78	26.03	30.66	76.54
BiDAF		40.26	45.74	82.54	40.33	45.91	82.82	46.46	52.75	90.23	44.17	50.52	86.8
Classifier	web	24.63	29.08	66.71	24.00	28.38	66.29	27.07	31.64	74.15	29.51	34.00	75.68
BiDAF		41.08	47.40	82.92	40.73	47.05	82.95	50.15	54.89	90.56	48.63	55.07	89.98

- Human baseline: 79.7% on Wiki, 75.4% on web

Conclusion

- TriviaQA
 - 650K question-answer-evidence triples
 - Questions authored by trivia enthusiasts
 - Evidence documents from Web search and Wiki pages
 - Experiments show TriviaQA is a challenging testbed

Thanks!