



CONVERSATIONAL QA



Conversational Question Answering

- Questions and answers in free-form text
- Different forms, different challenges
 - Chit Chat
 - Multi-turn QA
 - Clarifications
- Different from MRC:
 - Isolated vs contextual
 - Question lengths: shorter for conversational QA datasets (contextual)

Conversational Question Answering (CoQA)



- Multi-turn conversation, each turn is a question and an answer
- Questions and answers in free-form text
- Conversation is grounded in Passage
 - Concrete eval unlike chit-chat
- 127,000 questions and answers
 - 8K conversations (avg. 15 turns)
 - 7 diverse domains
 - Children stories, literature, exams, cnn news, Wikipedia
 - Hidden domains : reddit, science

CoQA Dataset

The Virginia governor's race, billed as the marquee battle of an otherwise anticlimactic 2013 election cycle, is shaping up to be a foregone conclusion. Democrat Terry McAuliffe, the longtime political fixer and moneymen, hasn't trailed in a poll since May. Barring a political miracle, Republican Ken Cuccinelli will be delivering a concession speech on Tuesday evening in Richmond. In recent ...

Q₁: What are the candidates **running** for?

A₁: Governor, R₁: The Virginia governor's race

Q₂: **Where**?

A₂: Virginia, R₂: The Virginia governor's race

Q₃: Who is the democratic candidate?

A₃: **Terry McAuliffe**, R₃: Democrat Terry McAuliffe

Q₄: Who is **his** opponent?

A₄: **Ken Cuccinelli**, R₄: Republican Ken Cuccinelli

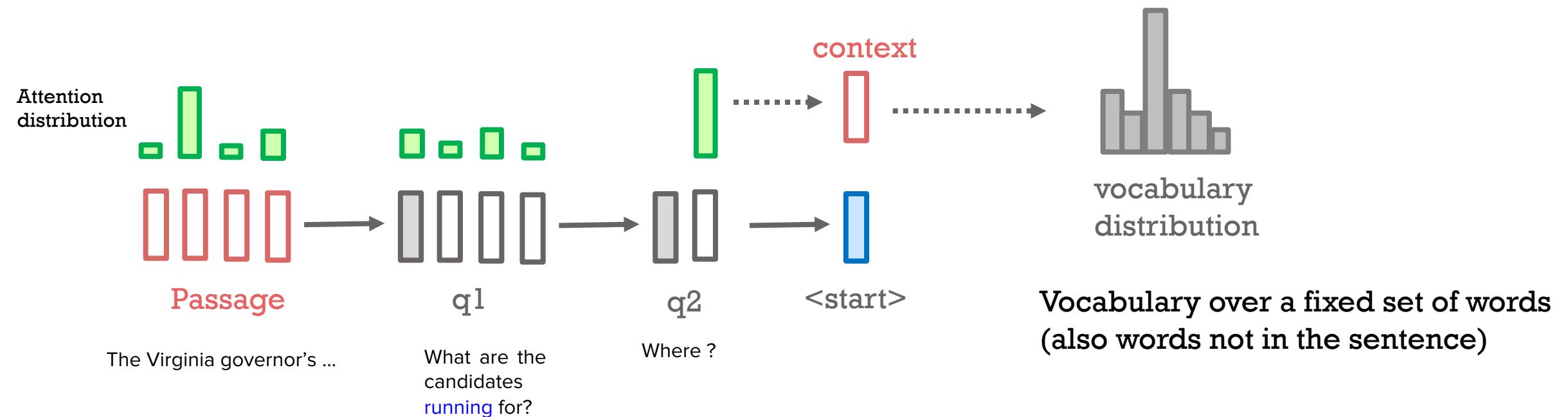
Q₅: What party does **he** belong to?

A₅: Republican, R₅: Republican Ken Cuccinelli

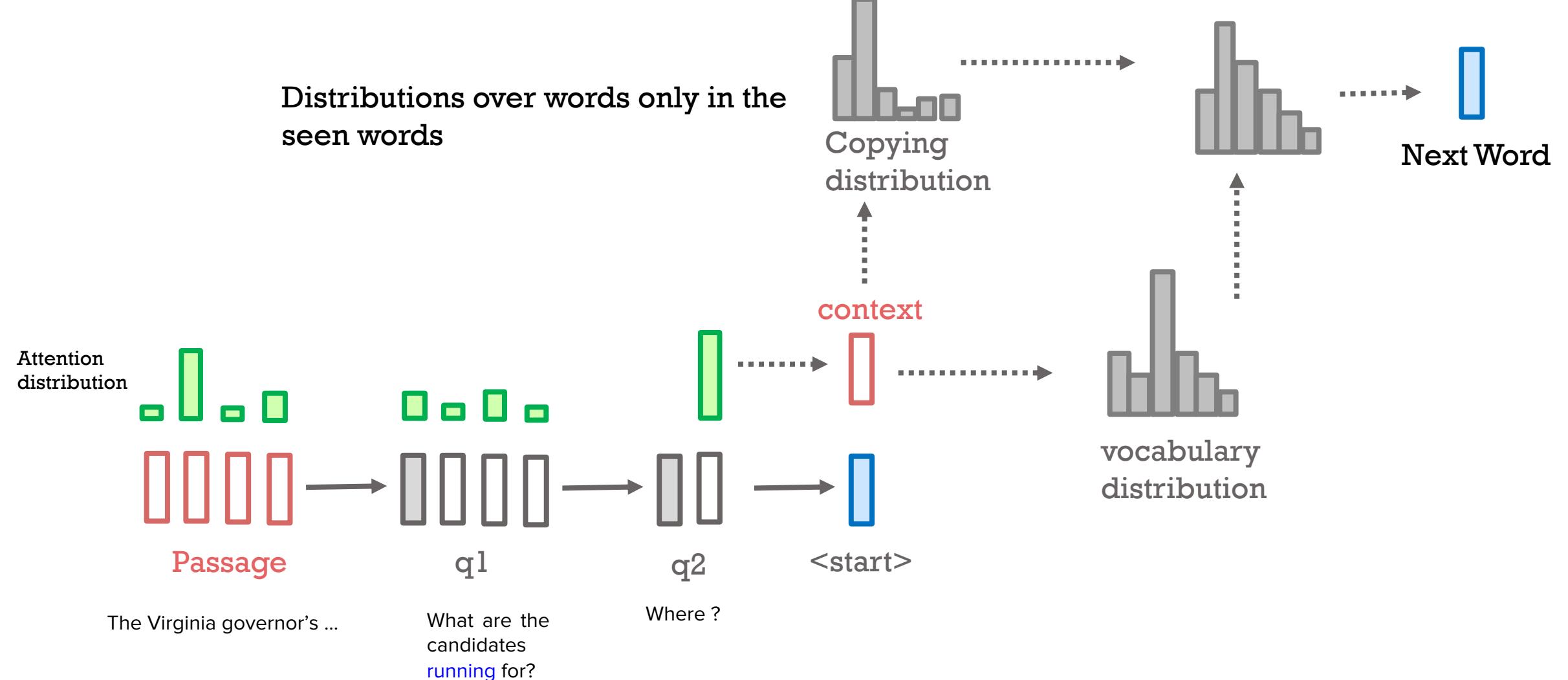
Q₆: Which of **them** is winning?

A₆: **Terry McAuliffe**, R₆: Democrat Terry McAuliffe, the longtime political fixer and moneymen, hasn't trailed in a poll since May

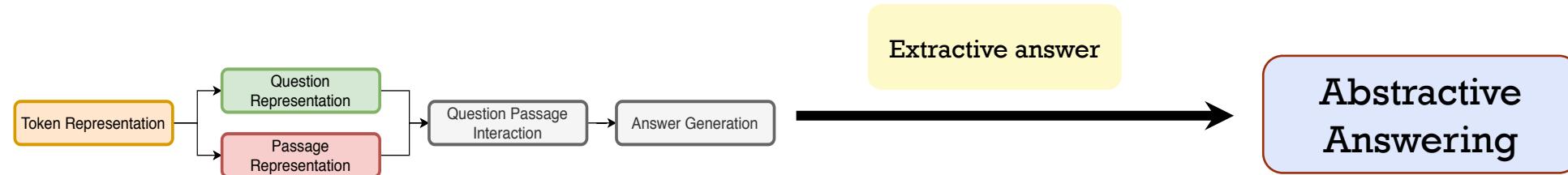
Seq2Seq Abstractive Response Generation



PGNet



Hybrid Model



**Seq2seq
Model**

Type	Seq2seq	PGNet	DrQA	Augmt. DrQA	DrQA+ PGNet	Human
Answer Type						
Answerable	27.5	45.4	54.7	67.3	66.3	89.9
Unanswerable	33.9	38.2	55.0	49.1	51.2	72.3
Span found	20.2	43.6	69.8	71.0	70.5	91.1
No span found	43.1	49.0	22.7	59.4	57.0	86.8
Named Entity	21.9	43.0	72.6	73.5	72.2	92.2
Noun Phrase	17.2	37.2	64.9	65.3	64.1	88.6
Yes	69.6	69.9	7.9	75.7	72.7	95.6
No	60.2	60.3	18.4	59.6	58.7	95.7
Number	15.0	48.6	66.3	69.0	71.7	91.2
Date/Time	13.7	50.2	79.0	83.3	79.1	91.5
Other	14.1	33.7	53.5	55.6	55.2	80.8

Leaderboard Sneakpeak

Rank	Model	In-domain	Out-of-domain	Overall					
	Human Performance <i>Stanford University</i> (Reddy & Chen et al. TACL '19)	89.4	87.4	88.8					
1	RoBERTa + AT + KD (ensemble) <i>Zhuiyi Technology</i> https://arxiv.org/abs/1909.10772	91.4	89.2	90.7					
1	TR-MT (ensemble) <i>WeChatAI</i>	91.5	88.8	90.7					
2	RoBERTa + AT + KD (single model) <i>Zhuiyi Technology</i> https://arxiv.org/abs/1909.10772	90.9	89.2	90.4	39	DrQA + seq2seq with copy attention (single model) <i>Stanford University</i> https://arxiv.org/abs/1808.07042	67.0	60.4	65.1
3	TR-MT (ensemble) <i>WeChatAI</i>	91.1	87.9	90.2					
4	Google SQuAD 2.0 + MMFT (ensemble) <i>MSRA + SDRG</i>	89.9	88.0	89.4					
5	TR-MT (single model) <i>WeChatAI</i>	90.4	86.8	89.3					
6	XLNet + Augmentation (single model) <i>Xiaoming</i> https://github.com/stevezheng23/xl_net_extension_tf	89.9	86.9	89.0					

Other Conversational Datasets

QuAC [Choi '19]

- Simulating info. seeking dialog
 - About a Wikipedia text
- 11k Dialogs, 98K QA Pairs
- Simple evaluation

QuLAC [Aliannejadi '19]

- Clarifying questions in info. seeking conversations
- Open domain, IR setting
- 198 topics [TREC Web Track]

Section:  Daffy Duck, Origin & History

STUDENT: What is the origin of Daffy Duck?

TEACHER: ↗ first appeared in Porky's Duck Hunt

STUDENT: What was he like in that episode?

TEACHER: ↗ assertive, unrestrained, combative

STUDENT: Was he the star?

TEACHER: ↗ No, barely more than an unnamed bit player in this short

STUDENT: Who was the star?

TEACHER: ↗ No answer

STUDENT: Did he change a lot from that first episode in future episodes?

TEACHER: ↗ Yes, the only aspects of the character that have remained consistent (...) are his voice characterization by Mel Blanc

STUDENT: How has he changed?

TEACHER: ↗ Daffy was less anthropomorphic

STUDENT: In what other ways did he change?

TEACHER: ↗ Daffy's slobbery, exaggerated lisp (...) is barely noticeable in the early cartoons.

STUDENT: Why did they add the lisp?

TEACHER: ↗ One often-repeated "official" story is that it was modeled after producer Leon Schlesinger's tendency to lisp.

STUDENT: Is there an "unofficial" story?

TEACHER: ↗ Yes, Mel Blanc (...) contradicts that conventional belief

...



dinosaur

Information Need (Facet)
I'm looking for the Discovery Channel's dinosaur site, which has pictures of dinosaurs and games.

 Are you looking for dinosaur books?
No, just the discovery channel website.

 Are you looking for meat-eating or plant-eating dinosaurs?
I'm not sure.
 No answer

 Would you like to see pictures or videos of dinosaurs?
I'd like to see pictures of dinosaurs on the discovery channels website.

Others: CSQA (Saha et al., 2018) CQA (Talmor and Berant, 2018) SQA (Iyyer et al., 2017)

Some papers this SIGIR...

Analyzing and Learning from User Interactions for Search Clarification

Hamed Zamani, Bhaskar Mitra, Everest Chen, Gord Lueck, Fernando Diaz, Paul N. Bennett,
Nick Craswell, and Susan T. Dumais

Microsoft

hazamani.bmitra.vuycche.gordonl.fdiaz.pnaben.nickcraswell.sudumais@microsoft.com

Open-Retrieval Conversational Question Answering

Chen Qu¹ Liu Yang¹ Cen Chen² Minghui Qiu³ W. Bruce Croft¹ Mohit Iyyer¹

¹University of Massachusetts Amherst ²Ant Financial ³Alibaba Group

{chenqu,lyang,croft,miyyer}@cs.umass.edu,chencen.cc@antfin.com,minghui.qmh@alibaba-inc.com

Query Resolution for Conversational Search with Limited Supervision

Nikos Voskarides¹ Dan Li¹ Pengjie Ren¹ Evangelos Kanoulas¹ Maarten de Rijke^{1,2}

¹University of Amsterdam, Amsterdam, The Netherlands ²Ahold Delhaize, Zaandam, The Netherlands

nickvosk@gmail.com,d.li@uva.nl,p.ren@uva.nl,e.kanoulas@uva.nl,m.derijke@uva.nl