UNIVERSITY OF AMSTERDAM

Conversational Question Answering at Scale

Svitlana Vakulenko

Conversational QA

Q1: Where is Xi’an?

A1: Shaanxi, China

Q2: What is its GDP?

A2: 932.12 billion yuan

Q3: What is the share in the province GDP?

A3: 41.8%

Conversational QA

Q1: Where is Xi'an?

A1: Shaanxi, China

Q2: What is its GDP?

A2: 932.12 billion yuan

Q3: What is the share in the province GDP?

A3: 41.8%

Anaphora

Ellipsis

Conversational QA

Q1: Where is Xi'an?

A1: Shaanxi, China

Q2: What is its GDP?

A2: 932.12 billion yuan

Q3: What is the share in the province GDP?

A3: 41.8%

✖ QuAC (Choi et al. EMNLP'18)

✖ CoQA (Reddy et al. TACL'19)

✖ ORConvQA (Qu et al. SIGIR'20)

UNIVERSITY OF AMSTERDAM

Conversational QA at Scale

Q1: Where is Xi'an?

A1: Shaanxi, China

Q2: What is its GDP?

A2: 932.12 billion yuan

Q3: What is the share in the province GDP?

A3: 41.8%

Xian (西安 Xī'ān, pron. SHE-ahn), is a historic city in **Shaanxi, China**.

https://wikitravel.org/en/Xian

Last year, **Xi'an**'s annual gross domestic product (**GDP**) hit **932.12 billion yuan**

https://govt.chinadaily.com.cn/s/202003/25/…

Xi'an is the largest economy of the Shaanxi province, with **GDP** of RMB 324.1 billion in 2010, up 14.5 percent year-on-year, and accounting for approximately **41.8%** of the **province's** total.

https://www.ucanews.com/directory/dioceses/…

Conversational QA at Scale

Input: Where is X'ian [SEP] Shaanxi, China [SEP] What is its GDP [CLS]

Web Page: … Last year, Xi'an's annual gross domestic product (GDP) hit 932.12 billion yuan ..

https://govt.chinadaily.com.cn/s/202003/25/…

Answer: 932.12 billion yuan

Question Rewriting

Input: Where is X'ian [SEP] Shaanxi, China [SEP] What is its GDP [CLS]

QR: What is Xi'an’s GDP

Web Page: … Last year, Xi'an's annual gross domestic product (GDP) hit 932.12 billion yuan ..

https://govt.chinadaily.com.cn/s/202003/25/…

Answer: 932.12 billion yuan

QReCC Dataset

- 14K self-dialogs with 81K question-answer pairs
- questions: QuAC, TREC CAsT, Natural Questions
- dialog length: AVG=6 turns
- brief answers (AVG=17 words)

Raviteja Anantha*, Svitlana **Vakulenko***, Zhucheng Tu, Shayne Longpre, Stephen Pulman, Srinivas Chappidi: Open-Domain Question Answering Goes Conversational via Question Rewriting. NAACL 2021.

QReCC Dataset

{
"Context": [
"where was the hallmark movie valentine ever after
"Valentine Ever After was mainly filmed in Ontario (
Colorado and Wyoming."
],
"Question": "which scenes in the film were filmed not
"**Rewrite**": "which scenes in valentine ever after were
"**Answer**": "In Valentine Ever After, the downtown stre
Cowboy Bar were filmed in the USA.",
"**Answer_URL**": "https://www.imdb.com/title/tt351552
"Conversation_no": 8352,
"Turn_no": 2,
"**Conversation_source**": "nq"
}

Valentine Ever After (2016 TV Movie)

Filming & Production

Showing all 3 items

Jump to: Filming Locations (3)

Filming Locations

Telluride, Colorado, USA
(downtown street scene)

6 of 6 found this interesting

25 N Cache St, Jackson, Wyoming, USA
(The Million Dollar Cowboy Bar)

5 of 5 found this interesting

Ontario, Canada
(main location)

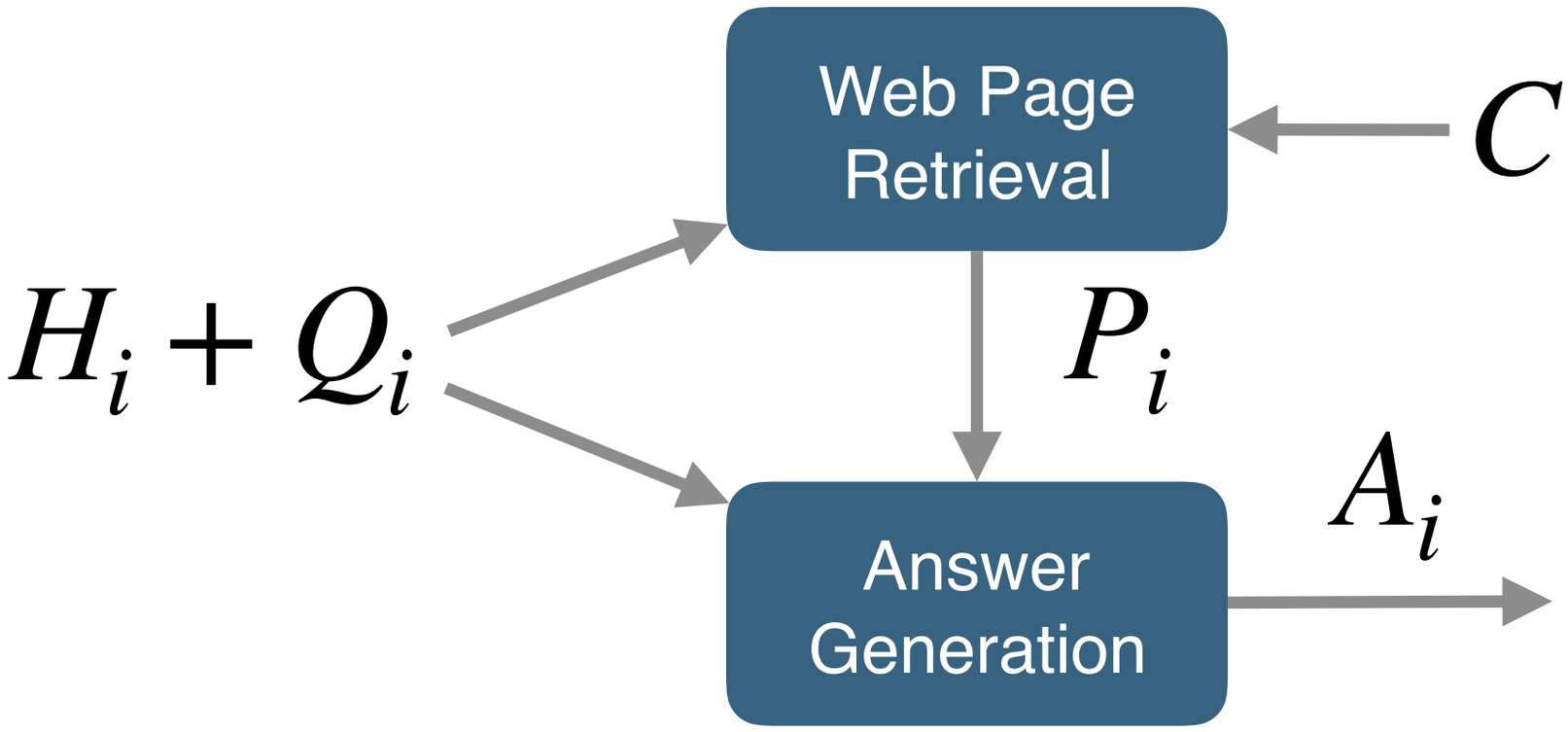
3 of 3 found this interesting

QReCC Dataset

- ✖ 70% dialogs require several pages (AVG=3)
- ✖ 14K relevant pages + 10M random pages
- ✖ 1% of Common Crawl

Raviteja Anantha*, Svitlana **Vakulenko***, Zhucheng Tu, Shayne Longpre, Stephen Pulman, Srinivas Chappidi: Open-Domain Question Answering Goes Conversational via Question Rewriting. NAACL 2021.

Conversational QA at Scale



- conversation history H_i
- current question Q_i
- web collection C

- relevant page P_i
- answer text A_i

Challenges

- ✖ contextual embeddings (+semantics -scalability)
  - ✖ long documents
    - ✖ large collection
- ✖ bag-of-words (+scalability -semantics)
  - ✖ long queries: context understanding
- ✖ evaluation: alternative answers

Conversational QA Approaches

1. Question rewriting

1. Question Rewriting

Q1: Where is Xi'an?

A1: Shaanxi, China

Q2: What is its GDP?

A2: 932.12 billion yuan

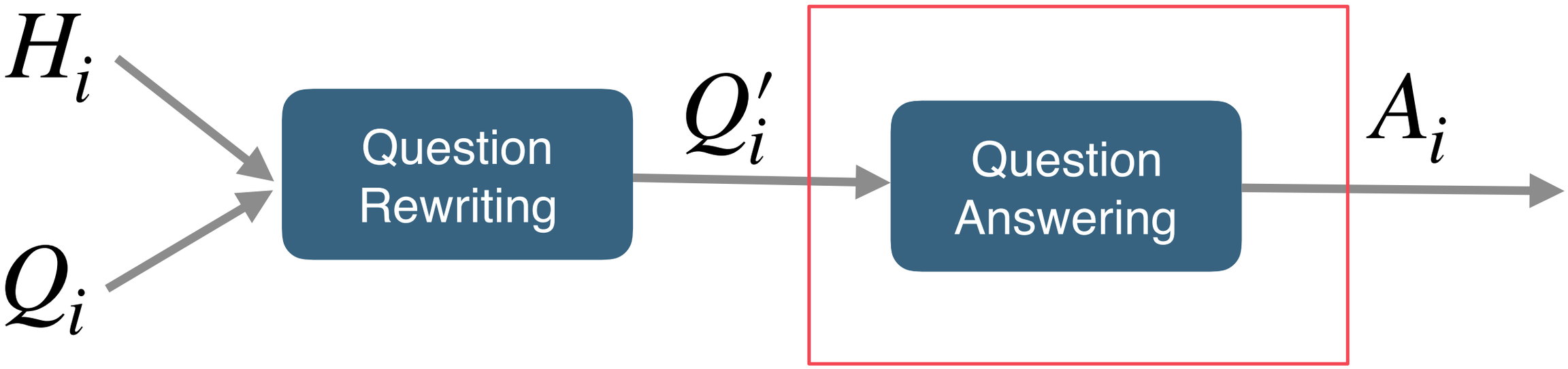
Q3: What is the share in the province GDP?

A3: 41.8%

⇨ Q2': What is Xi'an's GDP?

⇨ Q3': What is the share of Xi'an in the Shaanxi province GDP?

1. Question Rewriting



- conversation history H_i
- current question Q_i
- rewritten question Q_i'
- answer text A_i

1. Question Rewriting

✖ seq2seq task (translation/summarization)

Input: Where is X'ian [SEP] Shaanxi, China [SEP] What is its GDP [CLS]

Output: What is Xian’s GDP

Svitlana **Vakulenko**, Zhucheng Tu, Shayne Longpre: Question Rewriting for Conversational Question Answering. WSDM. 2021.

1. Question Rewriting

✖ Training: teacher forcing

✖ Cross-entropy loss (softmax)

$$-\sum_{t_i \in A} \sum_{c_j \in V} y_{ij} log(p_{ij}) = -\sum_{t_i \in A} log(p_{c_j = t_i})$$

1. Question Rewriting

✖ Inference: greedy decoding

$$until\ t_i = [STOP]\ do\ t_i = argmax\ M(< t_1 \ldots t_{i-1} >)$$

TREC CAsT 2019

- Conversational Passage Retrieval
  - **QR**: fine-tuned GPT2
  - **QA**: Anserini BM25 + BERT reranker
- # dialogues: train 30 test 50

https://www.treccast.ai

UNIVERSITY OF AMSTERDAM

TREC CAsT 2019

Run	Group	MAP	MRR	NDCG@3
UMASS_DMN_V2	UMass	0.082	0.300	0.100
ict_wrfml	ICTNET	0.105	0.373	0.165
UNH-trema-ecn	TREMA-UNH	0.073	0.505	0.222
unh-trema-relco	TREMA-UNH	0.077	0.533	0.239
UNH-trema-ent	TREMA-UNH	0.076	0.534	0.242
topicturnsort	ADAPT-DCU	0.136	0.555	0.259
rerankingorder	ADAPT-DCU	0.137	0.564	0.259
combination	ADAPT-DCU	0.130	0.539	0.259
datasetreorder	ADAPT-DCU	0.135	0.550	0.260
VESBERT	VES	0.124	0.541	0.291
VESBERT1000	VES	0.204	0.555	0.304
manual_indri_ql	-	0.309	0.660	0.361
clacMagic	WaterlooClarke	0.302	0.687	0.411
clacMagicRerank	WaterlooClarke	0.301	0.732	0.411
RUCIR-run1	RUCIR	0.163	0.725	0.415
ug_cur_sdm	uogTr	0.334	0.715	0.421
CFDA_CLIP_RUN1	CFDA_CLIP	0.224	0.772	0.460
h2oloo_RUN4	h2oloo	0.319	0.811	0.529
h2oloo_RUN3	h2oloo	0.322	0.810	0.531
CFDA_CLIP_RUN8	CFDA_CLIP	0.361	0.854	0.560
h2oloo_RUN5	h2oloo	0.352	0.864	0.561
CFDA_CLIP_RUN6	CFDA_CLIP	0.392	0.861	0.572
humanbert	ATeam	0.405	0.879	0.589
mpi-d5_cqw	mpi-inf-d5	0.185	0.591	0.286
mpi-d5_igraph	mpi-inf-d5	0.187	0.597	0.287
mpi-d5_intu	mpi-inf-d5	0.240	0.596	0.289
ensemble	CMU	0.258	0.587	0.294
bertrr_rel_q	USI	0.141	0.516	0.298
bertrr_rel_1st	USI	0.146	0.539	0.308
UDInfoC_BL	udel_fang	0.075	0.596	0.316
mpi_bert	mpii	0.166	0.597	0.319
ug_cont_lin	uogTr	0.275	0.584	0.325
ug_1stprev3_sdm	uogTr	0.253	0.585	0.328
clacBaseRerank	WaterlooClarke	0.244	0.629	0.343
BM25_BERT_RANKF	RUIR	0.158	0.597	0.350
ilps-bert-feat2	UAmsterdam	0.256	0.603	0.352
BM25_BERT_FC	RUIR	0.158	0.601	0.354
ug_cedr_rerank	uogTr	0.216	0.643	0.356
clacBase	WaterlooClarke	0.246	0.640	0.360
ilps-bert-featq	UAmsterdam	0.262	0.653	0.365
ilps-bert-feat1	UAmsterdam	0.260	0.614	0.377
pg2bert	ATeam	0.258	0.641	0.389
pgbert	ATeam	0.269	0.665	0.413
h2oloo_RUN2	h2oloo	0.273	0.714	0.434
CFDA_CLIP_RUN7	CFDA_CLIP	0.267	0.715	0.436

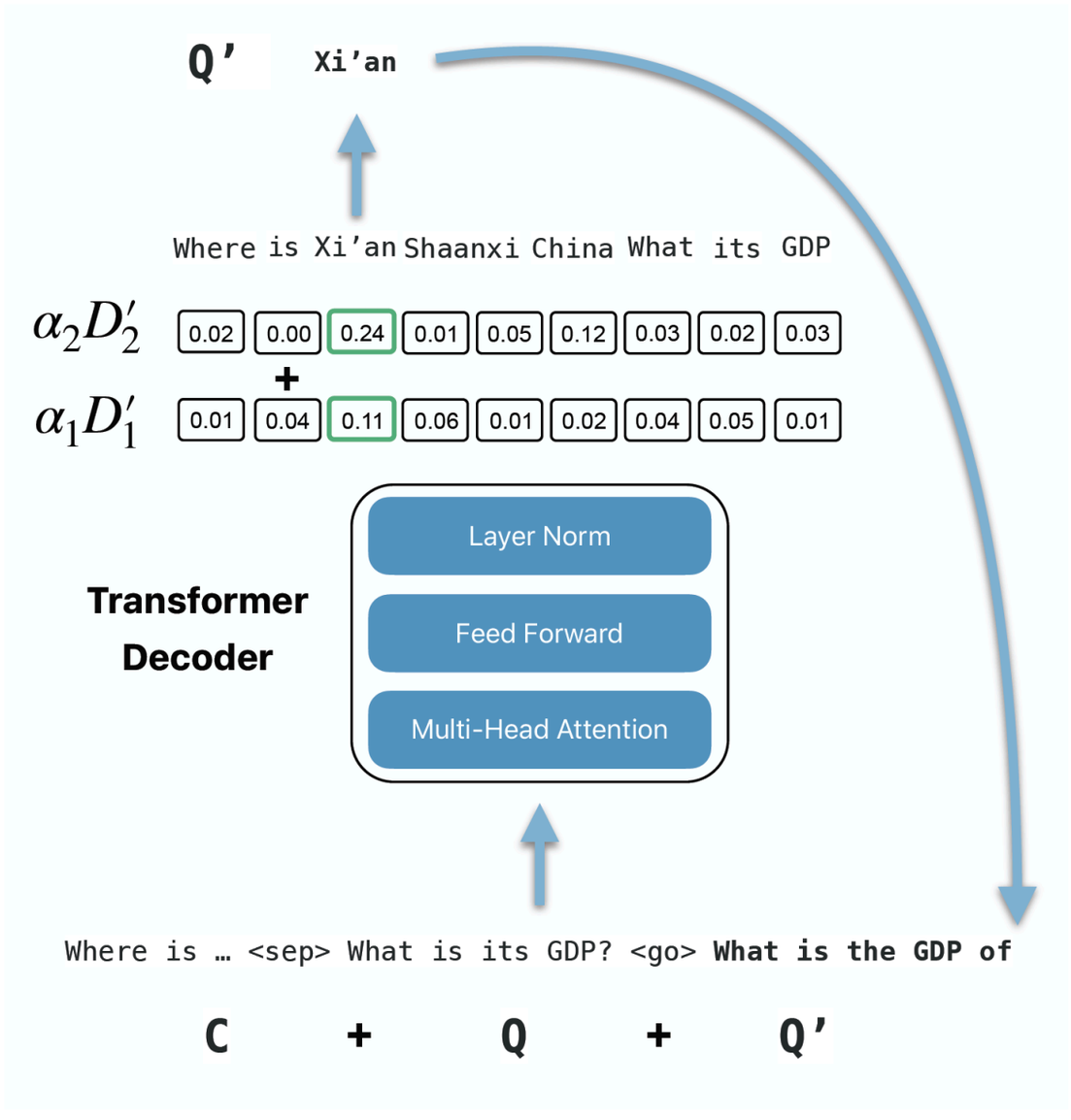
Results QuAC & CAsT'19

✖ CANARD (Elgohary et al. EMNLP'19)

QA Input	EM	F1	MAP	MRR	NDCG@3
Original	41.32	54.97	0.172	0.403	0.265
Original + 1-DT	43.15	57.03	0.230	0.535	0.378
Original + 2-DT	42.20	57.33	0.245	0.576	0.404
Original + 3-DT	43.29	57.87	0.238	0.575	0.401
Co-reference	42.70	57.59	0.201	0.473	0.316
PointerGenerator	41.93	57.37	0.183	0.451	0.298
CopyTransformer	42.67	57.62	0.284	0.628	0.440
Transformer++	**43.39**	**58.16**	**0.341**	**0.716**	**0.529**
Human	45.40	60.48	0.405	0.879	0.589

Svitlana **Vakulenko**, Zhucheng Tu, Shayne Longpre: Question Rewriting for Conversational Question Answering. WSDM. 2021.

Transformer++



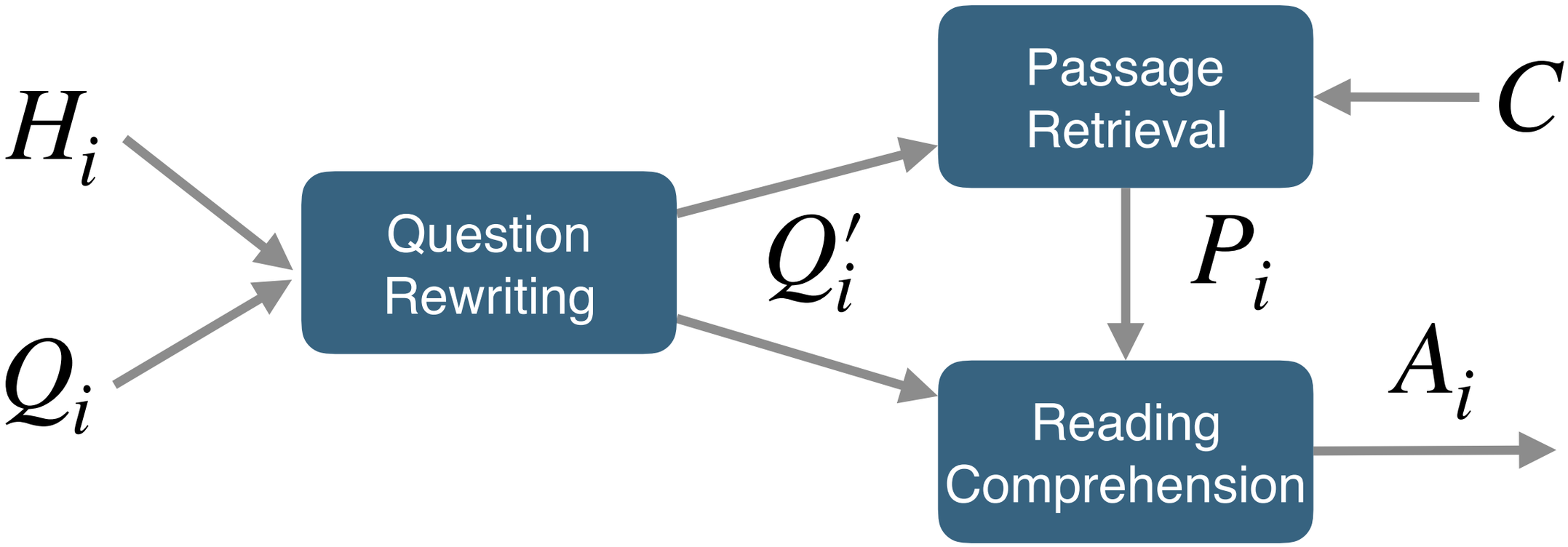
$$D'_i = W_i^H h + b^H$$

$$D' = \sum_{i=0}^{m} \alpha_i D'_i$$

$$\alpha_i = W_i^G norm(G) + W_i^X X + b_i^\alpha$$

Svitlana **Vakulenko**, Zhucheng Tu, Shayne Longpre: Question Rewriting for Conversational Question Answering. WSDM. 2021.

QReCC Baseline



- passage collection C
- top-k relevant passages P_i

QReCC Results

Setting	Rewrite Type	F1	EM
End-to-End	Original	11.78	0.49
	Transformer++	19.07	0.94
	Human	21.81	1.19
Known Context	Original	17.24	1.90
	Transformer++	32.34	4.04
	Human	36.42	4.70
Extractive Upper Bound		74.47	24.42

Raviteja Anantha*, Svitlana **Vakulenko***, Zhucheng Tu, Shayne Longpre, Stephen Pulman, Srinivas Chappidi: Open-Domain Question Answering Goes Conversational via Question Rewriting. NAACL 2021.

Search-Oriented Conversational AI

Online Event

8 October 2021

TIRA Forum

SCAI QReCC 21 Conversational Question Answering Challenge

Evaluation Results

Public Results | My Software | Task Page | Admin

Datasets

users

scai-qrecc21-test-dataset-2021-05-15 4

Evaluations on *scai-qrecc21-test-dataset-2021-05-15*

User	Actions	Software	Run	Input run	ROUGE1-R	MRR	F1	Exact match	Runtime
scai-qrecc21-naacl-baseline		software1	2021-07-04-19-39-23	2021-05-25-09-36-23	0.919	0.314	0.209	0.011	
scai-qrecc21-simple-baseline		software1	2021-07-04-19-43-59	2021-05-17-09-49-29	0.571	0.065	0.067	0.001	10:49:36

Conversational QA Approaches

1. Question rewriting

2. **Query expansion**

2. Query Expansion

Q: Where is Xi'an?

A: Shaanxi, China

Q: What is its GDP?

A: 932.12 billion yuan

Q: What is the share in the province GDP?

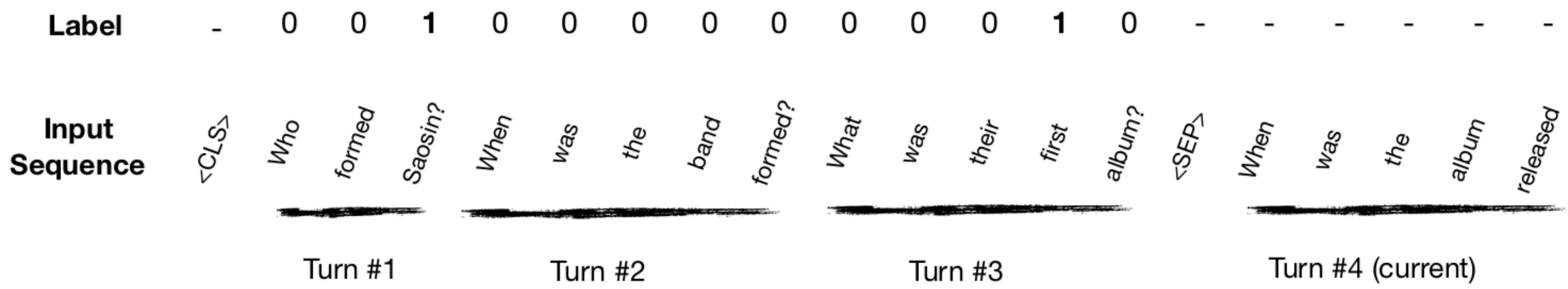
A: 41.8%

⇨ What is its GDP? Xi'an

⇨ What is the share in the province GDP? Xi'an Shaanxi

2. Query Expansion

✖ sequence labeling task (named entity recognition)



Nikos Voskarides, Dan Li, Pengjie Ren, Evangelos Kanoulas, Maarten de Rijke: Query Resolution for Conversational Search with Limited Supervision. SIGIR. 2020.

2. Query Expansion

✖ Binary cross-entropy loss

$$-\sum_{t_i \in H} [y_i log(p_i) + (1 - y_i) log(1 - p_i)]$$

TREC CAsT 2020

Group	Run	NDCG@3	Canonical result	Method	Model size
h2oloo	h2oloo_RUN2	0.494	manual	heuristic rules	770M + 11B
h2oloo	h2oloo_RUN1	0.444	manual	heuristic rules	
UvA.ILPS	quretecQR	0.340	automatic	end-to-end	110M + 336M
HPCLab-CNR	HPCLab-CNR-run3	0.331			
HPCLab-CNR	HPCLab-CNR-run1	0.313			
USI	hist_concat	0.281			
USI	hist_attention	0.214			
UvA.ILPS	quretecNoRerank	0.171			

Svitlana **Vakulenko**, Nikos Voskarides, Zhucheng Tu, Shayne Longpre. Leveraging Query Resolution and Reading Comprehension for Conversational Passage Retrieval. TREC. 2020.

Svitlana **Vakulenko**, Nikos Voskarides, Zhucheng Tu, Shayne Longpre: A Comparison of Question Rewriting Methods for Conversational Passage Retrieval. ECIR. 2021.

TREC CAsT 2020

Q1	What are some **interesting facts about bees**?	**Fun facts about bees**: 1 Honeybees are the only insect that produces food eaten by humans. 2 Worker honeybees are female. 3 The average worker bee produces only 1/12 teaspoon of honey over her lifetime. 4 A worker bee lives about 6 weeks. The queen bee can live to be 5 years old. 5 **Honey never spoils**.
Q2	Why doesn't **it** spoil?	Diana Taliun/iStock/Getty Images. **Honey doesn't spoil** like other foods and even if it has turned cloudy, it's still safe to eat. While you may be used to seeing clear, golden honey, cloudy honey is normal and natural.Clouding is caused by crystallization of the sugars and doesn't affect the taste or safety of the honey and is reversible.If the honey is in a jar, you can warm it on the stove. It doesn't harm honey to store it in the refrigerator, but it isn't necessary and will make honey cloudy and, sometimes, develop graininess as it crystallizes. 2 You can also warm honey in the microwave, as long as it is in a microwave-safe container and you use your microwave's lowest setting.
Q3	Why are so **many dying**?	The group found that **large numbers of bees are dying** during the summer months, when conditions should be more favorable. One in 4 colonies is now dying during summer, which was unheard of several years ago, according to the results. The total number of bee colonies in the United States declined from 6 million during the 1940s to 2.5 million about 10 years ago, but it has remained relatively stable since then.

TREC CAsT 2020

Table 4. Examples where QuReTeC performs better than Human.

qid			NDCG@3
101_9	Human	Does the public pay Jared Kushner?	0
	QuReTeC	And Jared? *ivana donald trump*	0.296
105_3	Human	Why was George Zimmerman acquitted?	0
	QuReTeC	Why was he acquitted? george *trayvon martin* zimmerman	0.202
93_6	Human	What support does the franchise provide?	0
	QuReTeC	What support does it provide? *king* franchise *agreement burger*	0.521
98_7	Human	Can you show me *vegetarian* recipes with almonds?	0
	QuReTeC	Oh *almonds*? Can you show me recipes with it? *almonds*	0.296

Svitlana **Vakulenko**, Nikos Voskarides, Zhucheng Tu, Shayne Longpre. Leveraging Query Resolution and Reading Comprehension for Conversational Passage Retrieval. TREC. 2020.

Conversational QA Approaches

1. Question rewriting

2. Query expansion

3. Dense retrieval

3. Dense Retrieval

✖ end-to-end ranking task

Input: Where is X'ian [SEP] Shaanxi, China [SEP] What is its GDP [CLS]

Output: … Last year, Xi'an's annual gross domestic product (GDP) hit 932.12 billion yuan ..

3. Dense Retrieval

✖ Dual/Siamese encoder (k-NN search)

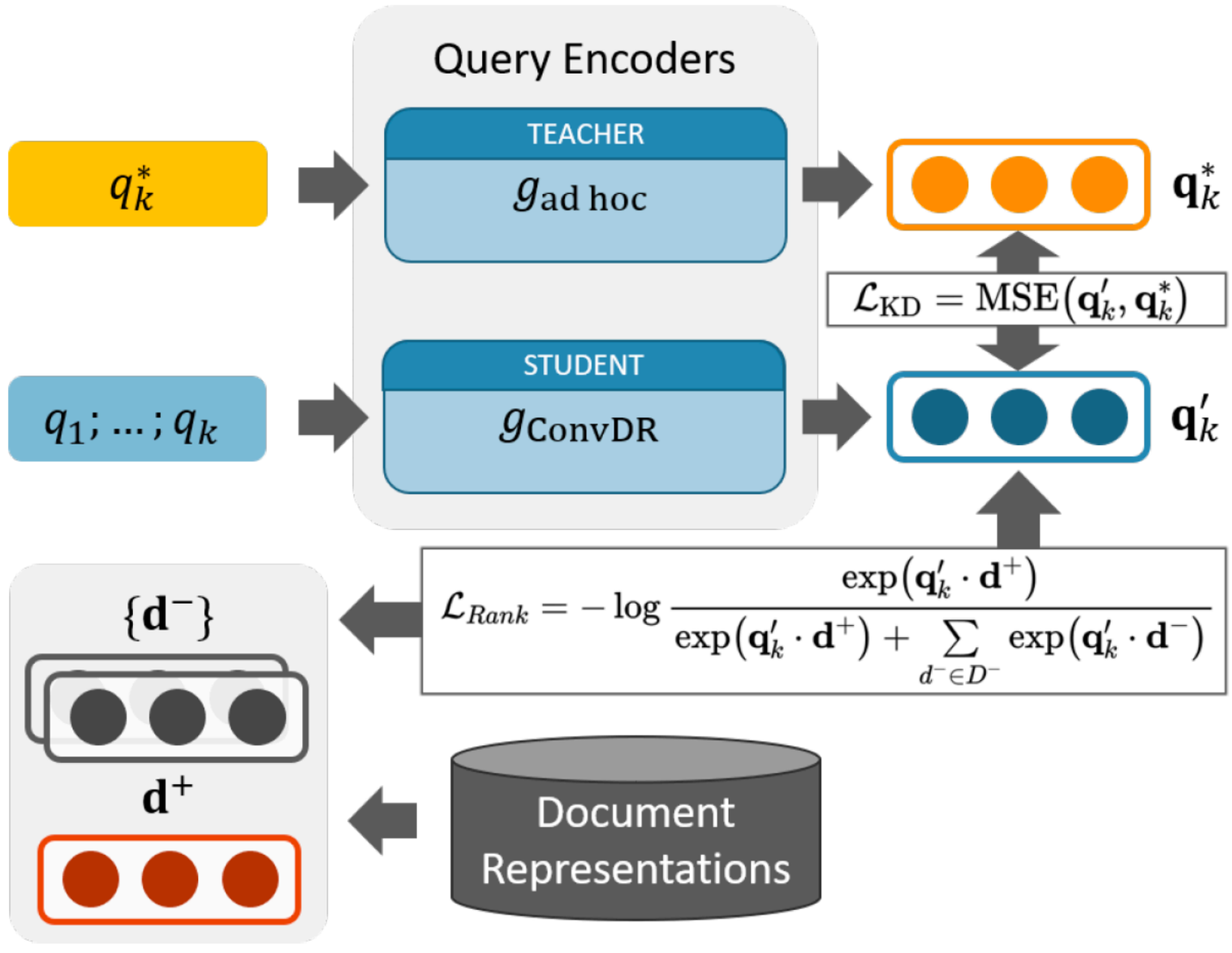
$$argmax_k \; M(< Q_i, H_i >) \cdot M(P_j)$$

✖ Knowledge distillation ~ Question Reformulation

$$max \; M(Q_i, H_i) \cdot M(Q_i')$$

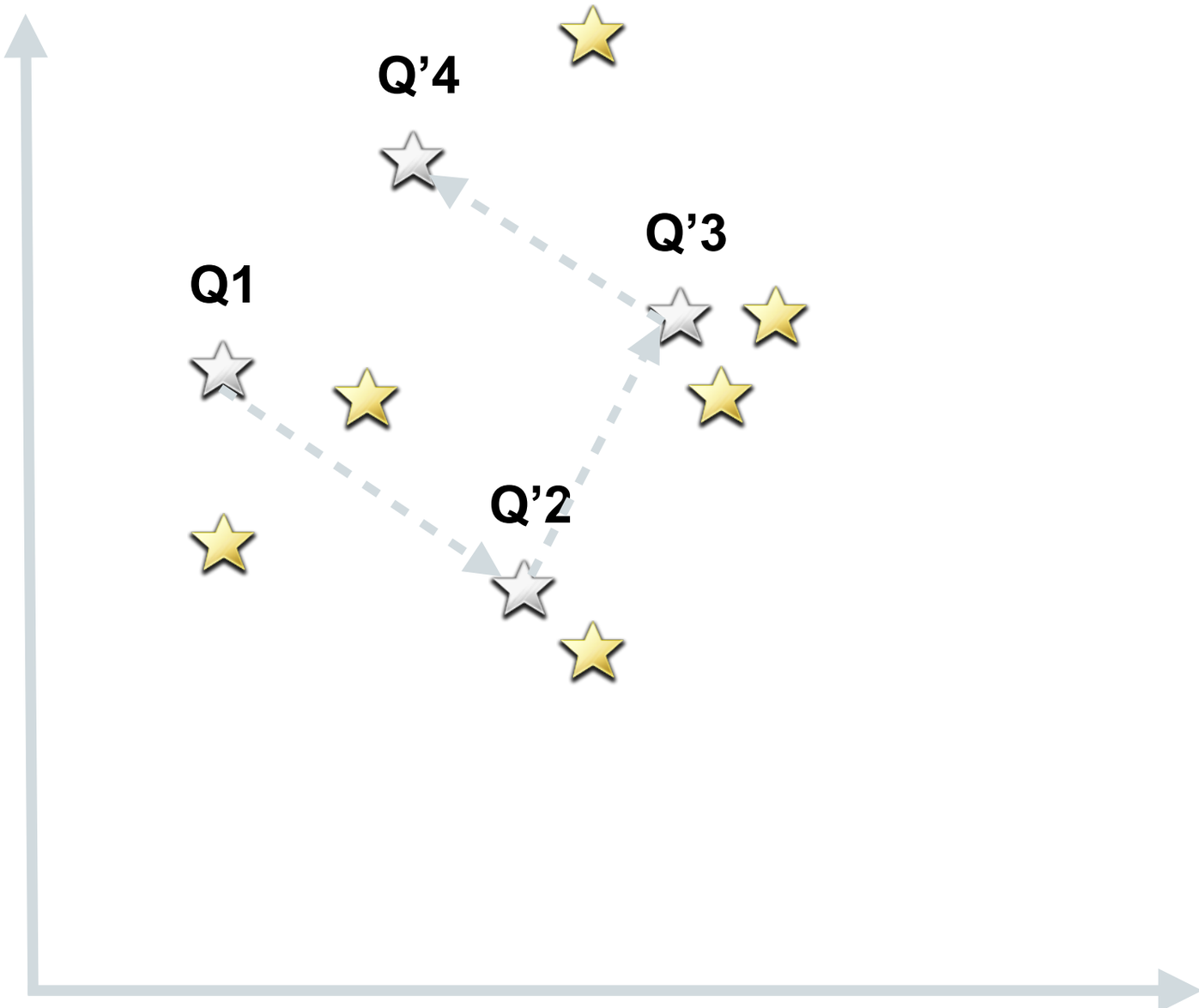
Shi Yu, Zhenghao Liu, Chenyan Xiong, Tao Feng, Zhiyuan Liu: Few-Shot Conversational Dense Retrieval. SIGIR. 2021.

3. Dense Retrieval



Shi Yu, Zhenghao Liu, Chenyan Xiong, Tao Feng, Zhiyuan Liu: Few-Shot Conversational Dense Retrieval. SIGIR. 2021.

3. Dense Retrieval



Conversational QA Approaches

1. Question rewriting

* distributed storage: external QA services / APIs

2. Query expansion

* large contexts: sampling/compression

3. Dense retrieval

* implicit query reformulation

Question Formulation

- ✖ more general framework
- ✖ context understanding/adaptation
- ✖ modular architecture
  - ✖ reusable
  - ✖ cheap

QR model

QA model

UNIVERSITY OF AMSTERDAM

Beyond QA

✖ Understanding interaction by modeling dialogue

1. Svitlana **Vakulenko**, Evangelos Kanoulas, Maarten de Rijke. A Large Scale Analysis of Mixed Initiative in Information-Seeking Dialogues for Conversational Search. TOIS. 2021.

2. Svitlana **Vakulenko**, Evangelos Kanoulas, Maarten de Rijke. An Analysis of Mixed Initiative and Collaboration in Information-Seeking Dialogues. SIGIR. 2020.

3. Svitlana **Vakulenko**, Kate Revoredo, Claudio Di Ciccio and Maarten de Rijke. QRFA: A Data-Driven Model of Information Seeking Dialogues. ECIR. ***Best paper award (User track).*** 2019.

4. Svitlana **Vakulenko**, Maarten de Rijke, Michael Cochez, Vadim Savenkov and Axel Polleres. Measuring Semantic Coherence of a Conversation. ISWC. ***Spotlight paper.*** 2018.

5. Svitlana **Vakulenko**, Ilya Markov, Maarten de Rijke. Conversational Exploratory Search via Interactive Storytelling. SCAI (ICTIR). 2017.

Search-Oriented Conversational AI

Online Event

8 October 2021

This workshop is intended as a **discussion platform on Conversational AI for intelligent information access** bringing together researchers and practitioners across NLP, IR, ML and HCI fields. Among other topics, we will discuss design, evaluation and human factors in relation to automating information-seeking dialogues. The workshop will also feature a shared task on Conversational Question Answering.

Svitlana Vakulenko
University of Amsterdam

Ondřej Dušek
Charles University

Leigh Clark
Swansea University