UNIVERSITY OF AMSTERDAM

Conversational Question Answering at Scale

Svitlana Vakulenko

Communication

UNIVERSITY OF AMSTERDAM

Knowledge

Telecommunication

- ✖ Computer-mediated communication
  - ✖ Real-time long-distance
- ✖ Information repository
  - ✖ Distributed system
  - ✖ Individual navigation

Information Retrieval

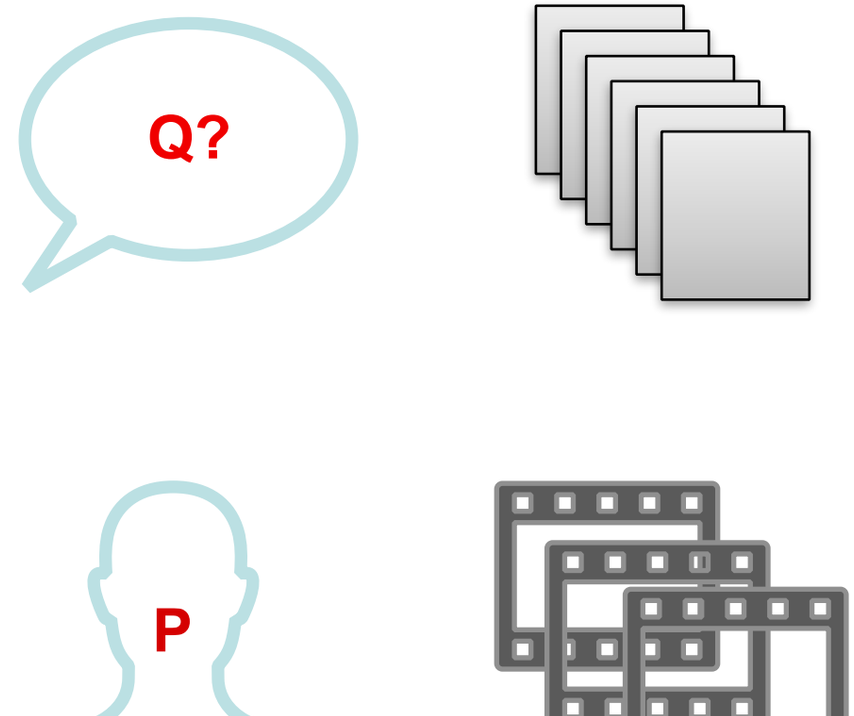
- ✖ Search
- ✖ Recommendation

Ranking Task

✖ Top-k matches by

✖ Relevance

✖ Similarity



Human-AI Collaboration

✖ Communication

Human-AI Collaboration

- ✖ Communication
  - ✖ Needs

Human-AI Collaboration

- ✖ Communication
  - ✖ Needs
  - ✖ Values/Preferences

Human-AI Collaboration

- ✖ Communication
  - ✖ Needs
  - ✖ Values/Preferences
  - ✖ State of knowledge

Human-AI Collaboration

- ✖ Communication
  - ✖ Needs
    - ✖ Values/Preferences
    - ✖ State of knowledge
    - ✖ Context

Human-AI Collaboration

- ✖ Communication
- ✖ Needs
- ✖ Values/Preferences
- ✖ State of knowledge
- ✖ Context
- ✖ News & Social Media
- ✖ Education
- ✖ Science
- ✖ eCommerce
- ✖ Healthcare

Mooers' Law

"An information retrieval system will tend not to be used whenever it is more **painful** and **troublesome** for a customer to have information than for him not to have it. "

Mooers, C. N. (1959). Mooers' Law, or Why Some Retrieval Systems Are Used and Others Are Not. American Documentation, 11(3).

Search Interfaces

Information Seeking Dialogue

- ✖ Negotiation
- ✖ Information Need
- ✖ Anomalous State of Knowledge
- ✖ Dynamic process

Summary

- ✖ Human-AI **collaboration**
  - ✖ requires efficient **communication**
- ✖ Natural language **dialogue**
  - ✖ -> Intuitive & powerful **interface**

Conversational Search

- ✖ Automate information-seeking dialogues
- ✖ Interactive IR using natural language
  - ✖ Natural Language Understanding

Conversational QA

Conversational QA

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%

Conversational QA

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%

Anaphora

Ellipsis

Conversational QA

$$f(Q_i, H_i, D) \rightarrow < s, e >$$

Conversational QA

Input for turn i :

$$f(Q_i, H_i, D) \rightarrow < s, e >$$

- question Q_i
- history $H_i = [A_{i-1}, Q_{i-1}, \ldots, Q_{i-j}]$
- document D

Output for turn i :

- start position $s \in \mathbb{Z} : s \in [0, n]$
- end position $e \in \mathbb{Z} : e \in [0, n]$

UNIVERSITY OF AMSTERDAM

Conversational QA at Scale

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%

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/…

Question Rewriting

Q: Where is Xi'an?

A: Shaanxi, China

Q: What is its GDP?

A: 932.12 billion yuan

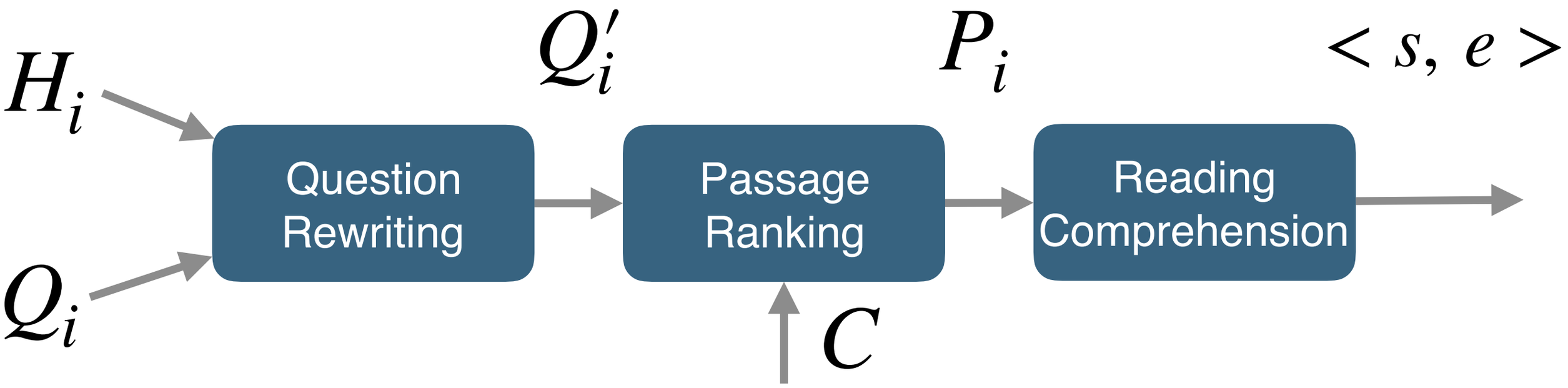
Q: What is the share in the province GDP?

A: 41.8%

⇨ What is Xi'an's GDP?

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

Architecture



- rewritten question Q_i'
- passage collection C
- top-k relevant passages P_i

Question Rewriting Model

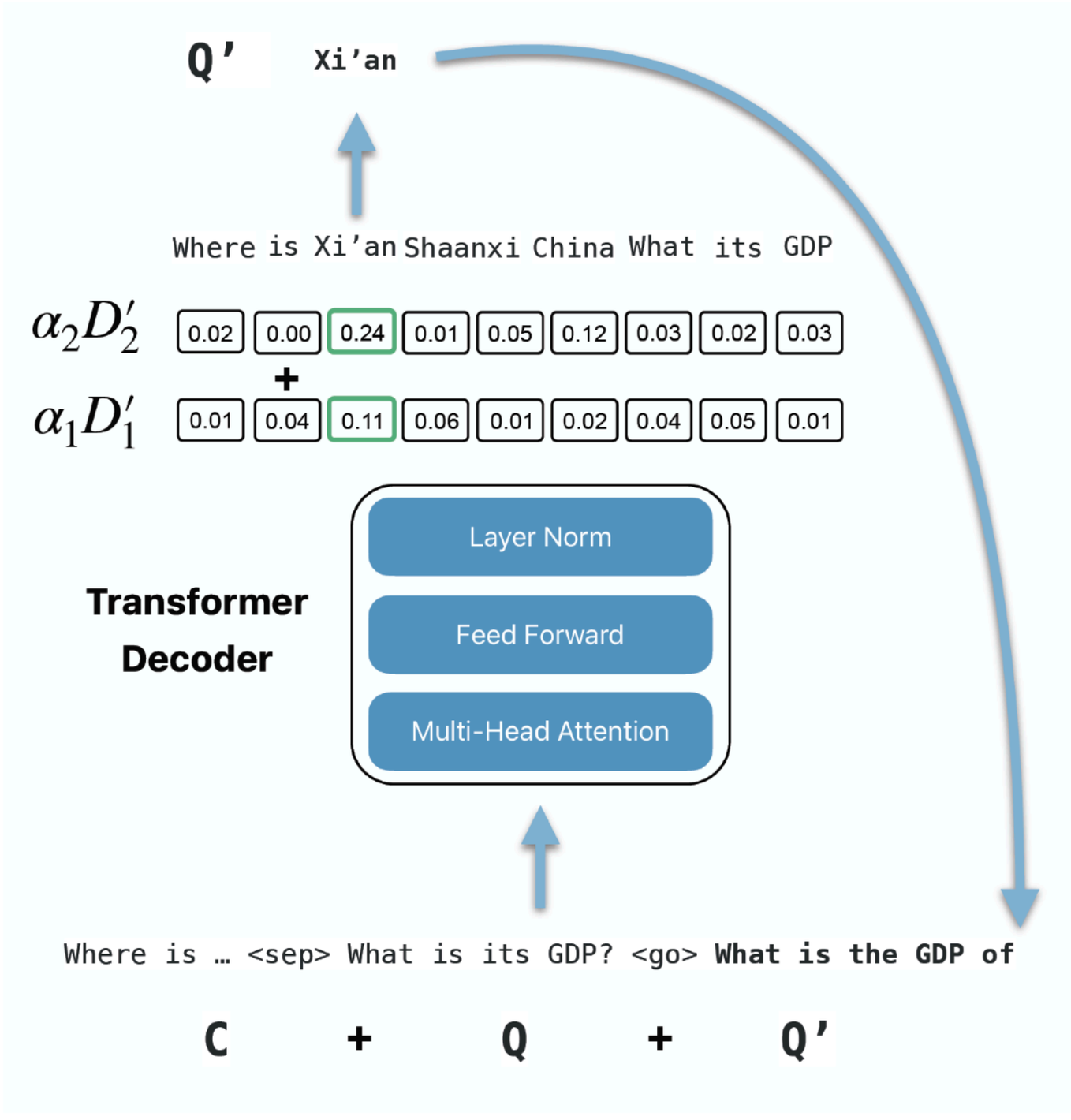
✖ Generative model (GPT2)

✖ NLL loss

✖ Recursively generate Q_i' token by token

✖ Teacher forcing (ground truth)

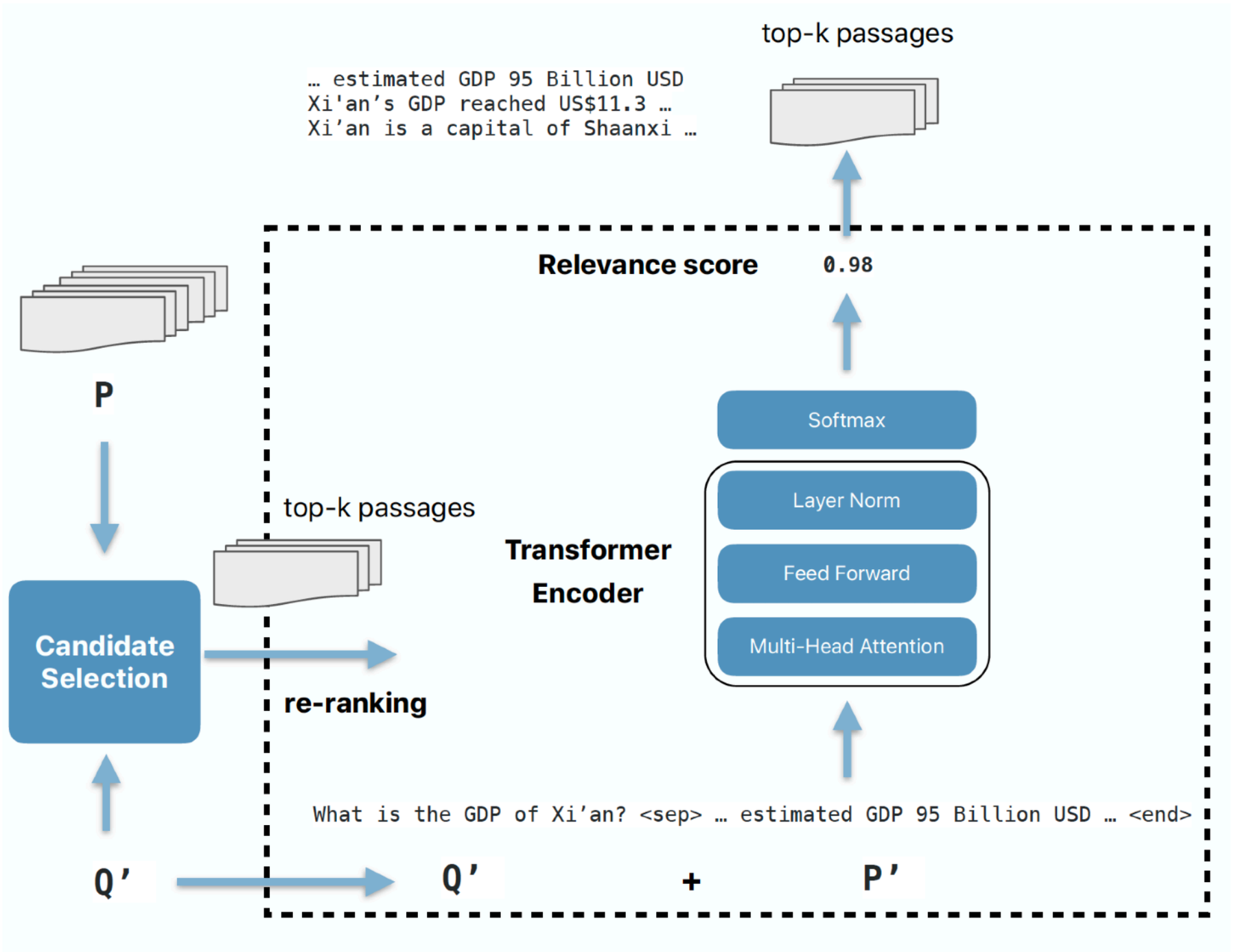
Question Rewriting Model



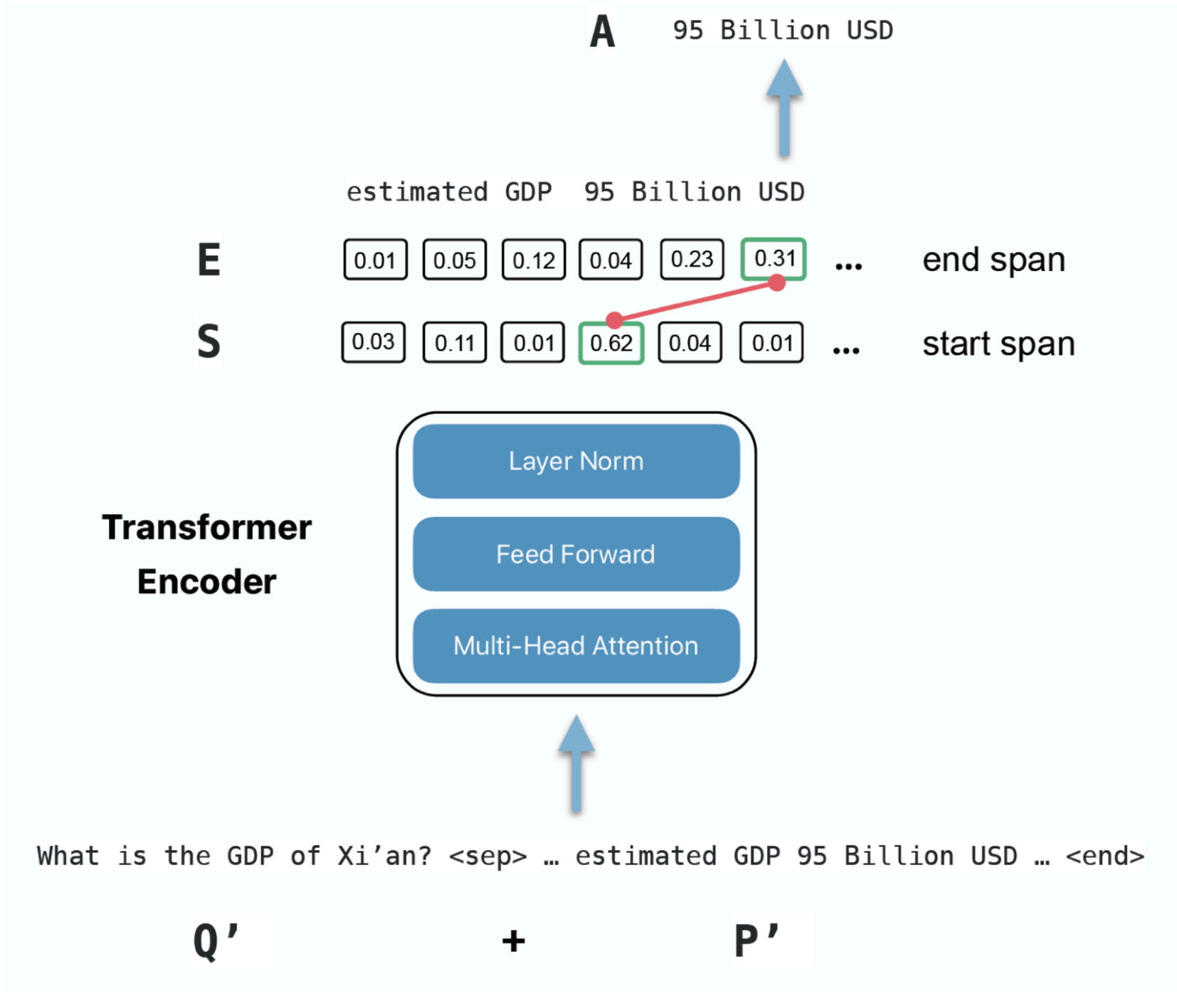
Passage Ranking Model

- ✖ Initial retriever (BM25)
  - ✖ Unsupervised
- ✖ Re-ranker (BERT)
  - ✖ Classification model

Passage Ranking Model



Reading Comprehension Model



Experimental Evaluation

- ✖ QR: CANARD (Elgohary et al., 2019): **35K** Qs
- ✖ Conversational Passage Ranking
  - ✖ TREC CAsT 2019 (Dalton et al., 2019): **173** Qs
- ✖ Conversational Reading Comprehension
  - ✖ QuAC (Choi et al., 2018): **35K** Qs

TREC CAsT 2019

QA Input	QA Model	MAP	MRR	NDCG@3
Original	Anserini	0,172	0,403	0,265
Original +1-DT*	+BERT	0,230	0,535	0,378
Original +2-DT*		0,245	0,576	0,404
Original +3-DT*		0,238	0,575	0,401
Co-reference		0,201	0,473	0,316
PointerGenerator		0,183	0,451	0,298
CopyTransformer		0,284	0,628	0,440
Transformer++		0,341	0,716	0,529
Human		0,405	0,879	0,589

QuAC

QA Input	QA Model	EM	F1	NA Acc
Original	MultiQA —>	41,32	54,97	65,84
Original +1-DT	CANARD-H	43,15	57,03	68,64
Original +2-DT		42,20	57,33	69,42
Original +3-DT		43,29	57,87	71,50
Co-reference		42,70	57,59	66,20
PointerGenerator		4193	57,37	63,16
CopyTransformer		42,67	57,62	68,02
Transformer++		43,39	58,16	68,29
Human		45,40	60,48	70,55

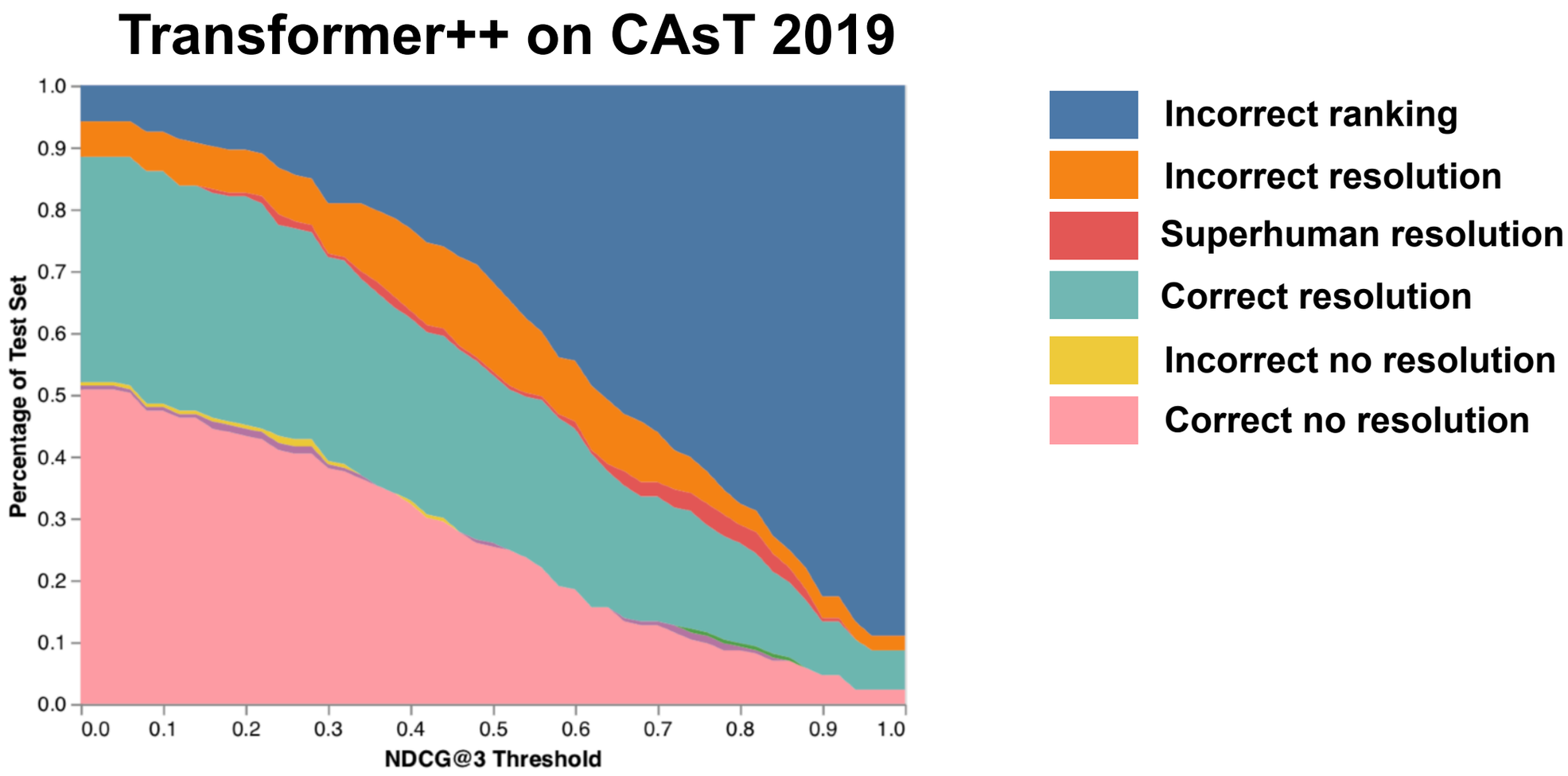
QReCC Dataset

- ✖ 14K conversations with 81K question-answer pairs
- ✖ 10M web pages (split into 54M passages)
- ✖ instructed to produce conversational answers

QReCC Dataset

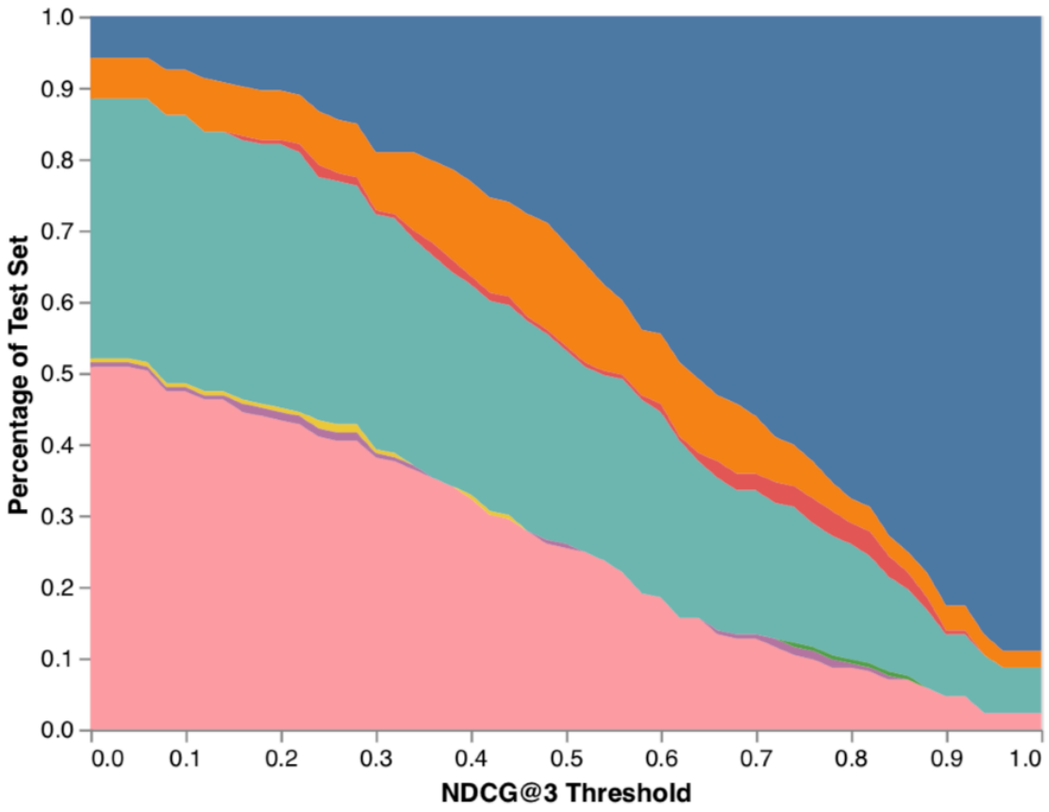
Setting	Rewrite Type	F1	EM
End-to-End	Original	11.78	0.49
	Transformer++	19.07	0.94
	Human	21.81	1.19
Known Context	Original	17.24	1.90
	Transformer++	32.34	4.04
	Human	36.42	4.70
Extractive Upper Bound		74.47	24.42

Error Analysis

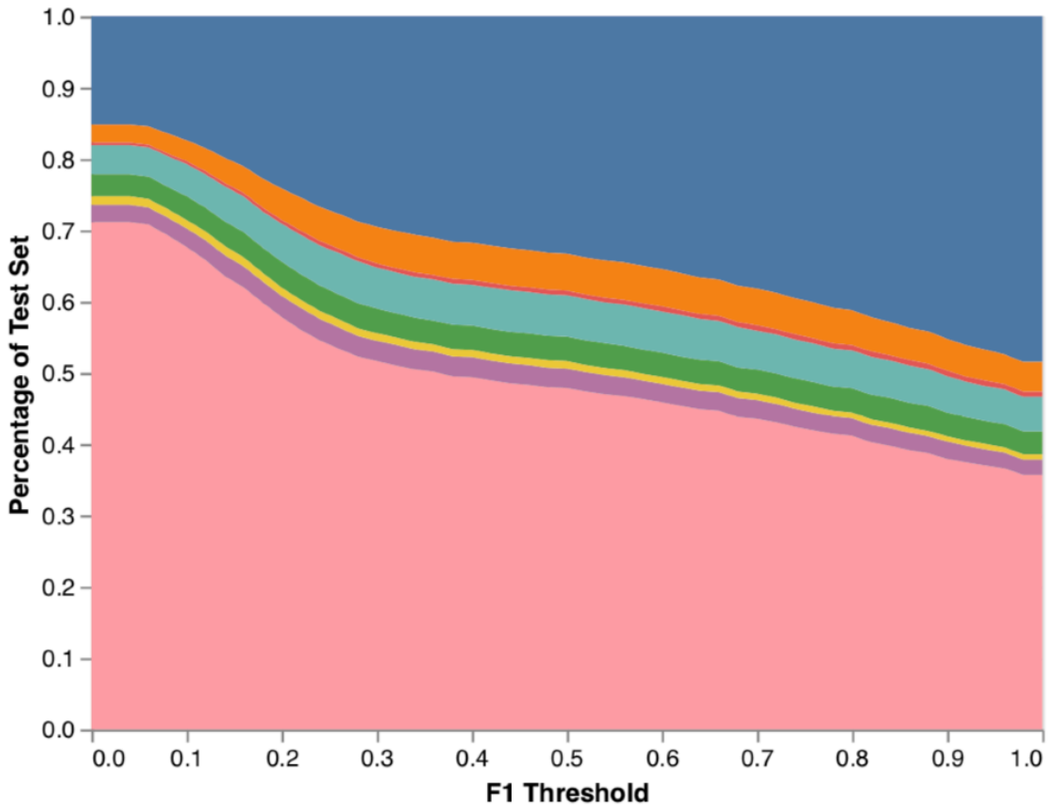


CAsT 2019 vs QuAC

Transformer++ on CAsT 2019

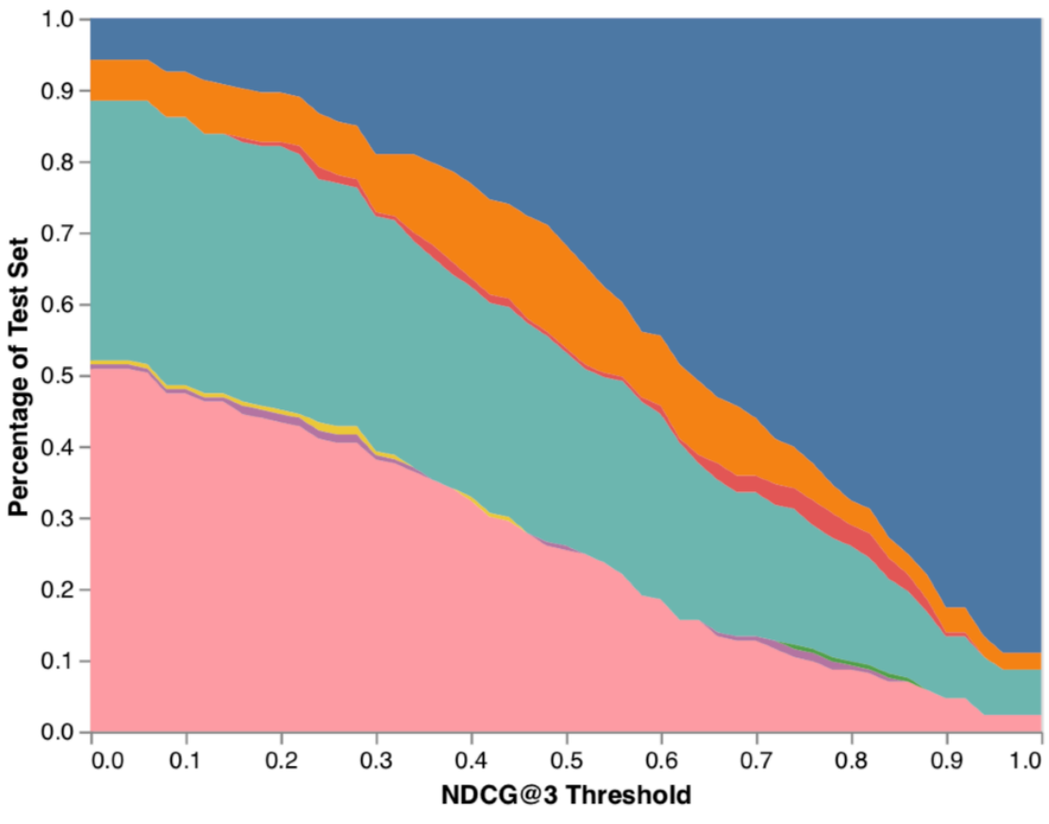


Transformer++ on QuAC

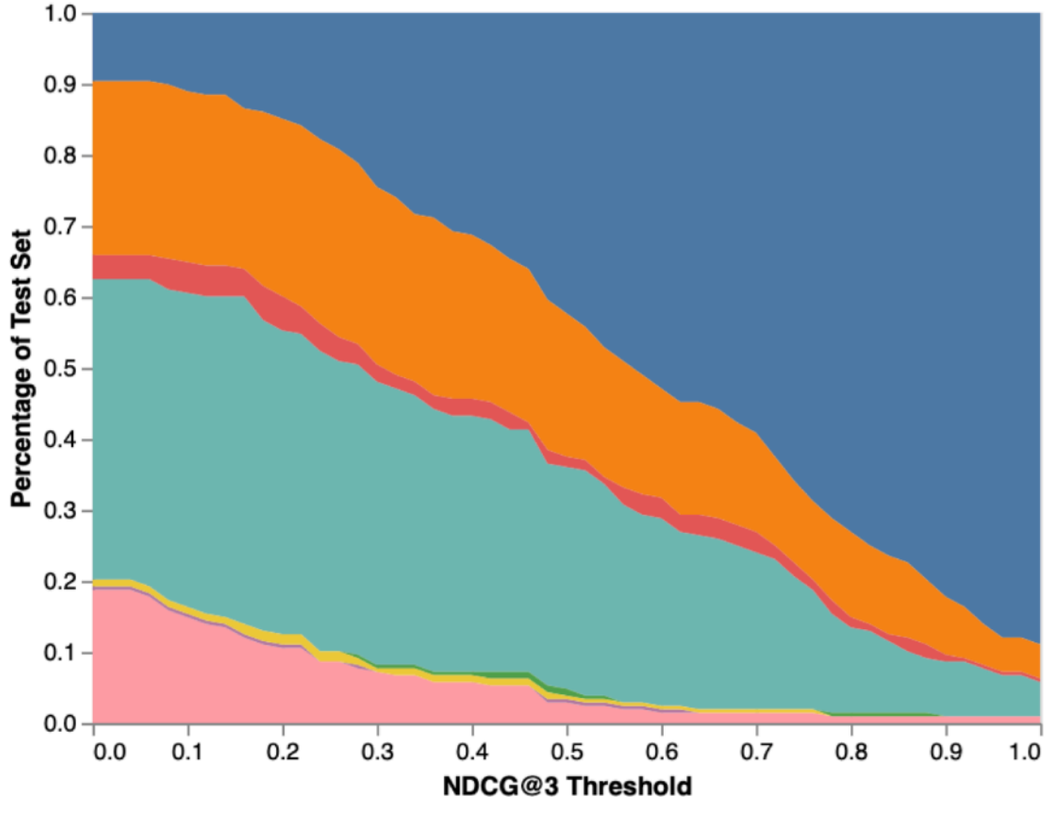


CAsT 2019 vs CAsT 2020

Transformer++ on CAsT 2019



QuReTeC on CAsT 2020



https://github.com/svakulenk0/QRQA

Summary

- ✖ Modular
- ✖ Reusable
- ✖ Cross-platform
- ✖ Debuggable
- ✖ Cheap

QR model

QA model

References

1. Svitlana **Vakulenko**, Zhucheng Tu, Shayne Longpre: Question Rewriting for Conversational Question Answering. WSDM. 2021. (To appear)

2. Svitlana **Vakulenko**, Shayne Longpre, Zhucheng Tu, Raviteja Anantha: A Wrong Answer or a Wrong Question? An Intricate Relationship between Question Reformulation and Answer Selection in Conversational Question Answering. SCAI@EMNLP. 2020. ***Best paper award***.

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

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

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

UNIVERSITY OF AMSTERDAM

My Team

MSc Shayne Longpre
Apple Inc.

MSc Zhucheng Tu
Apple Inc.

MSc Raviteja Anantha
Apple Inc.

Dr Nikos Voskarides
University of Amsterdam

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

UNIVERSITY OF AMSTERDAM

TREC CAsT 2020

Group	Run	Recall	MAP	MRR	NDCG	NDCG@3	Canonical result	Method	Model size
h2oloo	h2oloo_RUN2	0.705	0.326	0.621	0.575	0.494	manual	heuristic rules	770M + 11B
h2oloo	h2oloo_RUN1	0.705	0.284	0.576	0.549	0.444	manual	heuristic rules	
UvA.ILPS	quretecQR	0.264	0.147	0.476	0.283	0.340	automatic	end-to-end	110M + 336M
HPCLab-CNR	HPCLab-CNR-run3	0.561	0.193	0.449	0.422	0.331			
HPCLab-CNR	HPCLab-CNR-run1	0.545	0.181	0.434	0.403	0.313			
USI	hist_concat	0.475	0.160	0.424	0.354	0.281			
USI	hist_attention	0.475	0.125	0.340	0.321	0.214			
UvA.ILPS	quretecNoRerank	0.264	0.081	0.262	0.216	0.171			

https://www.treccast.ai

On-going Work

Question-based Summarization

What is a physician's assistant?

Physician assistants work under the supervision of a physician or surgeon; however, their specific duties and the extent to which they must be supervised differ

A **Physician Assistant** can deliver diagnostic, therapeutic, and preventive healthcare services, as delegated by a physician

Physician assistants, also known as PAs, practice medicine on a team under the supervision of physicians and surgeons.

UNIVERSITY OF AMSTERDAM

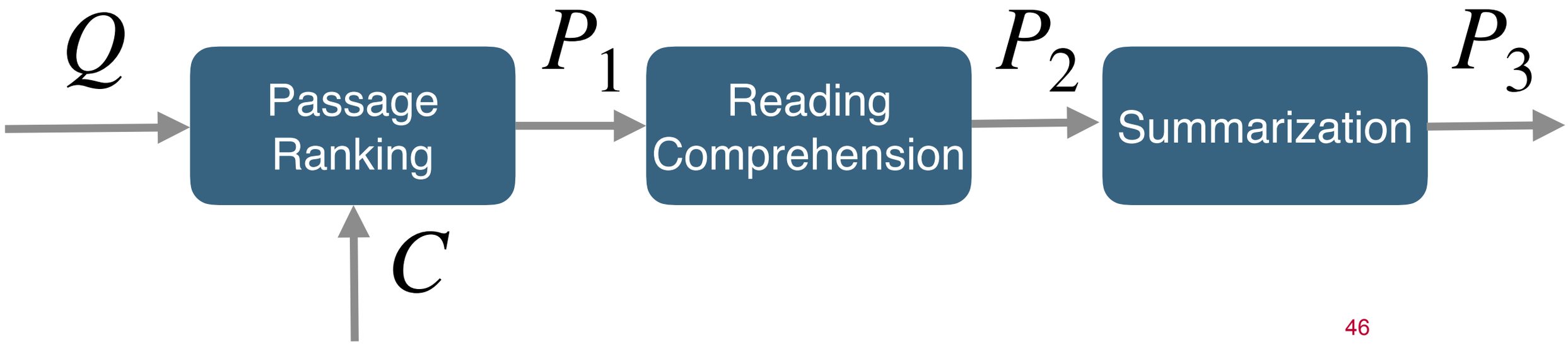
Question-based Summarization

MSc Weijia Zhang
University of Amsterdam

✖ Importance

✖ Relevance

✖ Diversity



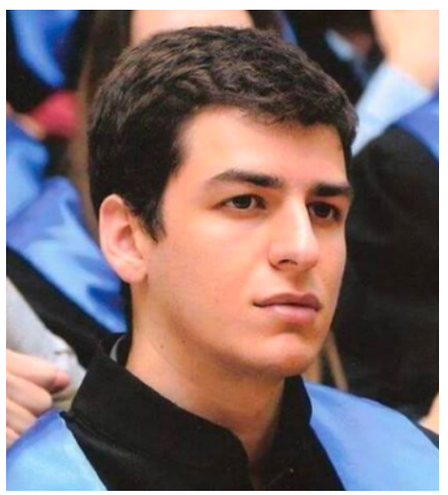
Multi-hop QA

Gunmen from Laredo starred which narrator of "Frontier"?

Gunmen from Laredo is a 1959 American western film produced and directed by Wallace MacDonald, which stars Robert Knapp, Maureen Hingert, and **Walter Coy**.

Walter Darwin Coy (January 31, 1909 – December 11, 1974) was an American stage, radio, film, and, principally, television actor, originally from Great Falls, Montana. He was best known for narrating the NBC western anthology series, **"Frontier"**, which aired early Sunday evenings in the 1955–1956 season.

Multi-hop QA



MSc Georgios Sidiropoulos
University of Amsterdam

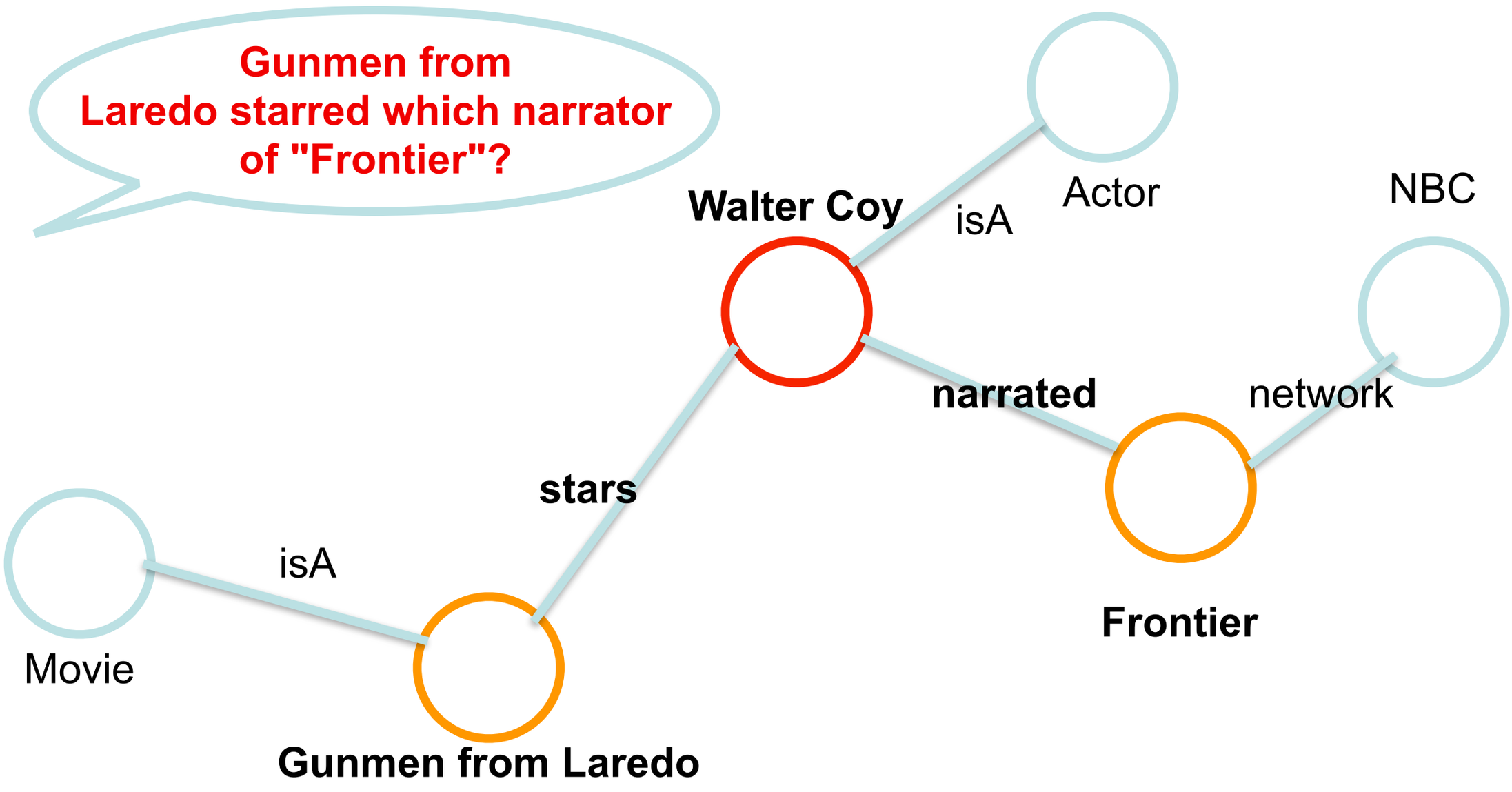
✖ Dense Passage Retrieval

✖ Efficient

✖ Scalable

✖ Generalizable

Knowledge Graph QA



Knowledge Graph QA

Dataset: LC-QuAD v.1

5K QA pairs

Knowledge Graph: DBpedia v. 2016-04 HDT

1B triples > 26M entities > 68K predicates

Rank	System	Precision	Recall	F1 Score	Runtime
1	**QAmp** Vienna University of Economics and Business, Austria Paper	0.25	0.50	0.33	0.72 s/q
2	**WDAqua** Université de Lyon, France Paper	0.22	0.38	0.28	1.50 s/q
-	**Krantikari QA** Smart Data Analytics, Germany Paper | Code	TBD	TBD	TBD	

http://lc-quad.sda.tech/lcquad1.0.html

Knowledge Graph QA

- ✖ natural-language access to structured data sources
- ✖ e.g., DBpedia, Wikidata KGs and Open Data tables

1. Svitlana **Vakulenko**, Javier D. Fernandez, Axel Polleres, Maarten de Rijke and Michael Cochez. Message Passing for Complex Question Answering over Knowledge Graphs. CIKM. 2019.

2. Sophia Keyner, Vadim Savenkov and Svitlana **Vakulenko**. Open Data Chatbot. ESWC. 2019.

3. Svitlana **Vakulenko** and Vadim Savenkov. TableQA: Question Answering on Tabular Data. SEMANTiCS. ***Best paper award nominee (Poster & Demo Track)***. 2017.

4. Sebastian Neumaier, Vadim Savenkov and Svitlana **Vakulenko**. Talking Open Data. ESWC. 2017.

UNIVERSITY OF AMSTERDAM

Knowledge Graph QA

- ✖ Conversational QA
- ✖ Supervised model
- ✖ Learn to traverse KG
- ✖ Sampling vertices/edges

Dr Peter Bloem
Vrije Universiteit Amsterdam

UNIVERSITY OF AMSTERDAM

Inverse QA

Dr Ali Bahrainian
EPFL & Brown University

✖ Produce QA pairs

✖ Bootstrap annotation

✖ Dynamic evaluation

Inverse QA

how much does it cost to subscribe to **amazon prime**
does **amazon prime** offer free shipping
is **amazon prime** unlimited
can you watch **amazon prime** offline
amazon prime instant video
amazon prime cost per month
what is the monthly fee for **amazon prime**
what is **netflix** monthly cost
what is **netflix** tv
what network is **netflix**

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. (Under review)

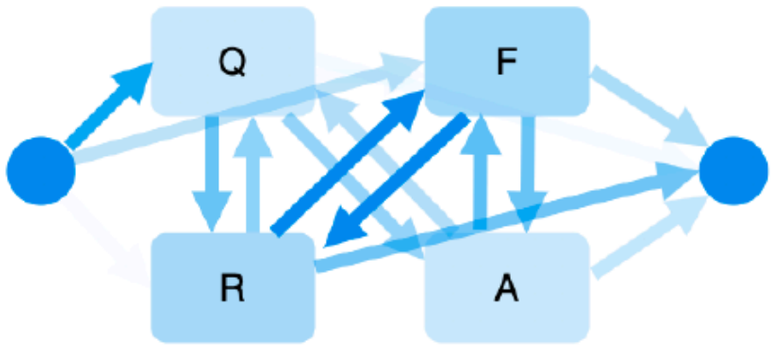
2. Svitlana **Vakulenko**, Evangelos Kanoulas, Maarten de Rijke. An Analysis of Mixed Initiative and Collaboration in Information-Seeking Dialogues. SIGIR. 2020.

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.

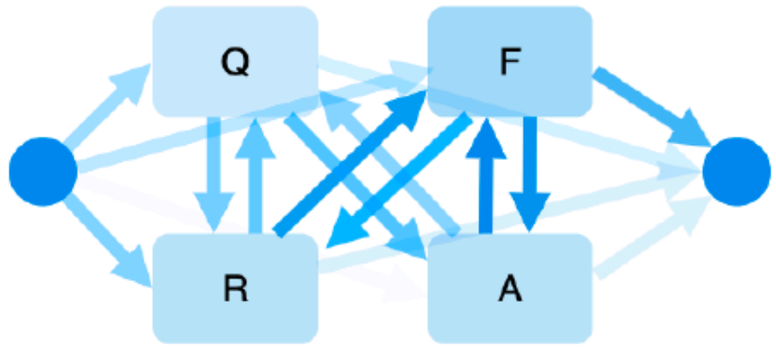
4. Svitlana **Vakulenko**, Maarten de Rijke, Michael Cochez, Vadim Savenkov and Axel Polleres. Measuring Semantic Coherence of a Conversation. ISWC. ***Spotlight paper***. 2018.

Beyond QA

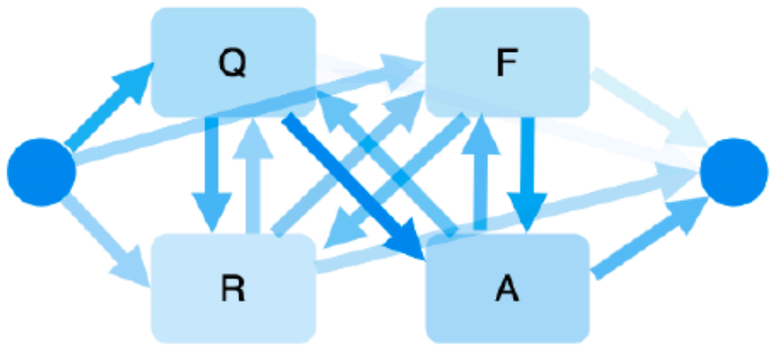
OCLC



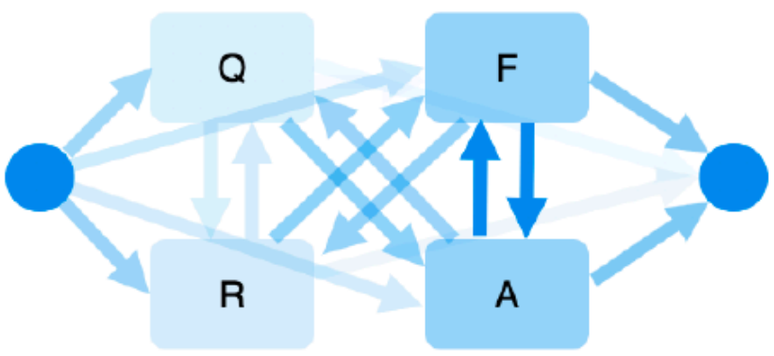
ReDial



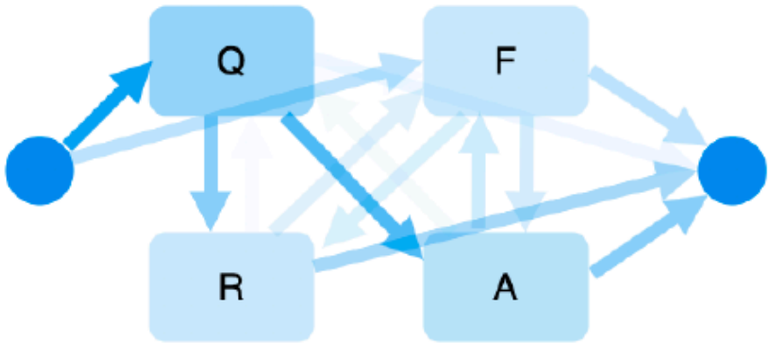
MANTIS



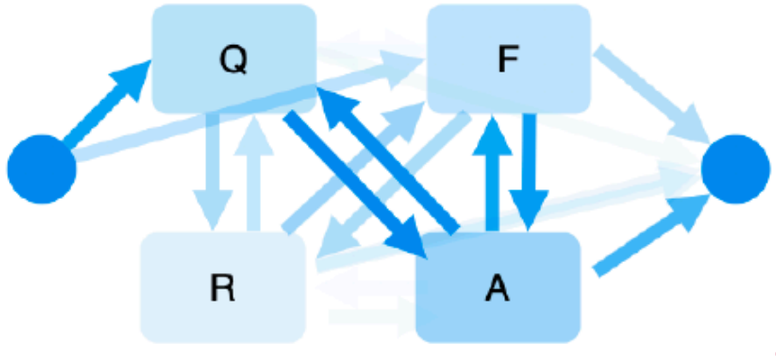
MISC



Ubuntu



SCSdata



Beyond QA

- ✖ Mixed Initiative ≠ Clarifying Questions
  - ✖ 576 **reference interviews** from OCLC
    - ✖ 51 clarifying questions = 3% of all questions
    - ✖ Feedback / Follow-up questions

Summary

- **Conversational QA** is an important step towards intelligent interfaces for information access
- But it is only **the first step**
- The road ahead requires **integration of expertise** from multiple domains, such as IR, NLP, KR & HCI

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

Heloisa Candello
IBM Research