



Question Rewriting for Conversational Question Answering

Svitlana Vakulenko, Shayne Longpre, Zhucheng Tu, Raviteja Anantha

Task

Conversational Question Answering

Q Where is Xi'an?

A Shaanxi, China

Q What is its GDP?

A 95 Billion USD

Q What is the share in the province GDP?

A 41.8%

Task

Conversational Question Answering

Q Where is Xi'an?

A Shaanxi, China

Q What is **its** GDP?

A 95 Billion USD

Q What **is the share in the province** GDP?

A 41.8%

Anaphora

Ellipsis

Approach

Question Rewriting

Q Where is Xi'an?

A Shaanxi, China

Q What is **its** GDP?

A 95 Billion USD

Q What **is the share in the province** GDP?

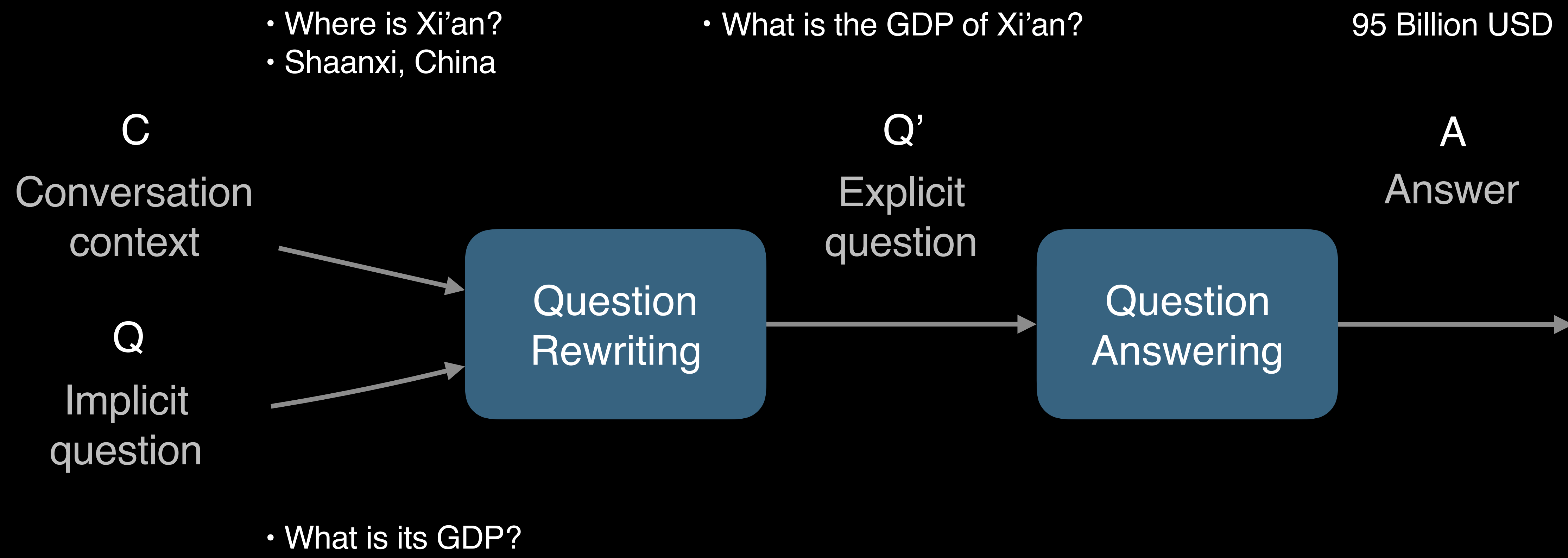
A 41.8%

QR What is **Xi'an's** GDP?

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

Approach

Question Rewriting



Related Work: Unsupervised Approach

[Mele et al., 2020]

- co-reference resolution + heuristics
- identify topics as nouns in dependency parses
- identify topic shifts via lexical cues

Related Work: Supervised Classification

QuReTeC [Voskarides et al., 2020]

Input	Where is Xi'an? <SEP> Shaanxi, China <SEP> What is its GDP?				
Output	0	0	1	0	0
QR	What is its GDP? Xi'an				

- CANARD 35K conversational questions + rewrites
- fine-tune BERT

Our Approach: Supervised Sequence Generation

Transformer++

Input Where is Xi'an? <SEP> Shaanxi, China <SEP> What is its GDP?

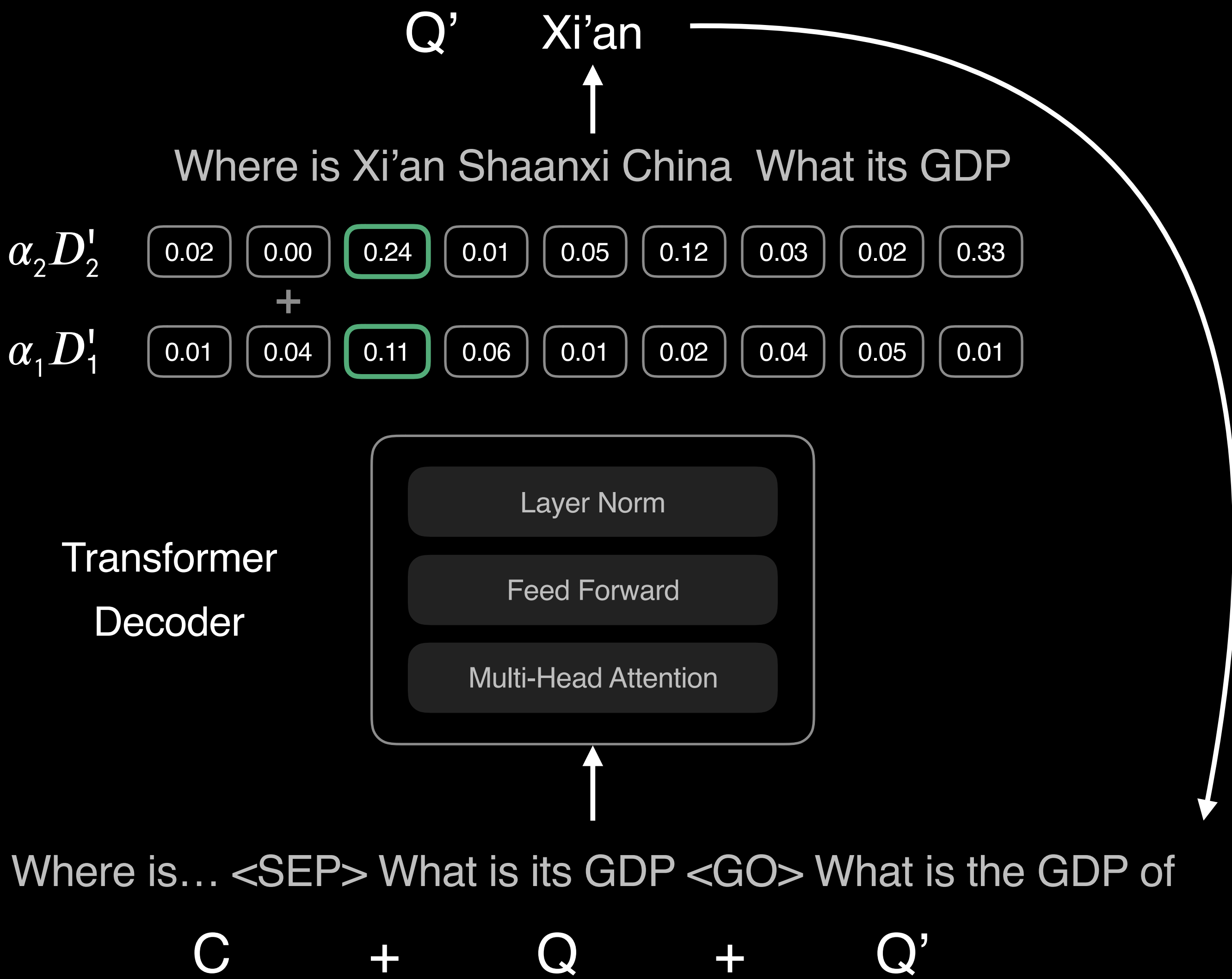
Output What is Xi'an's GDP?

- CANARD 35K conversational questions + rewrites
- fine-tune GPT2

Our Approach: Supervised Sequence Generation

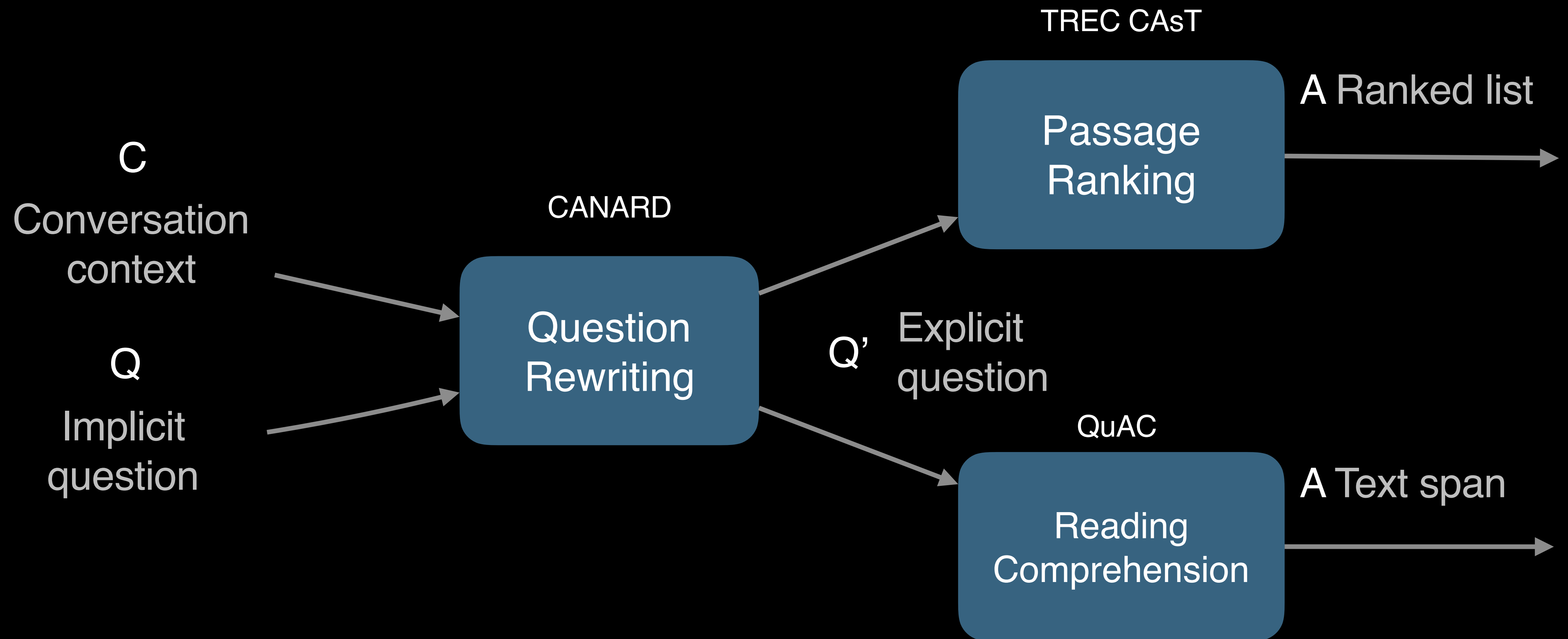
Transformer++

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



Experiments

Question Rewriting



Passage Ranking

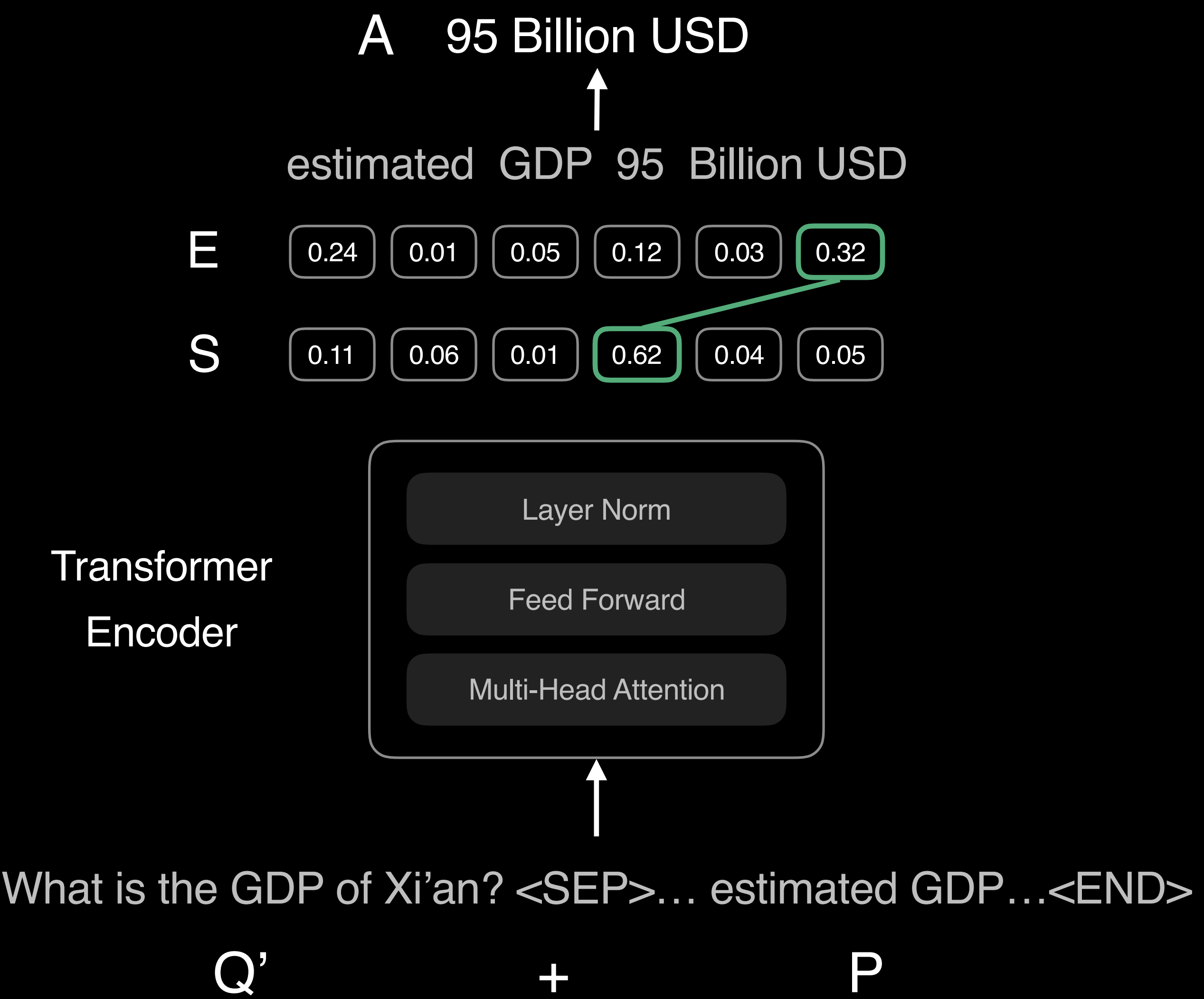
1. Candidate selection: BM25
 - initial top-k passages
2. Re-ranking: fine-tune BERT
 - re-ranked top-k passages

Input What is the GDP of Xi'an? <SEP>... estimated GDP...<END>
Output 0.98

Reading Comprehension

$$\max \sum_{i,j=0}^n (S_i + E_j)$$

where $j \geq i$



Results

TREC CAsT

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

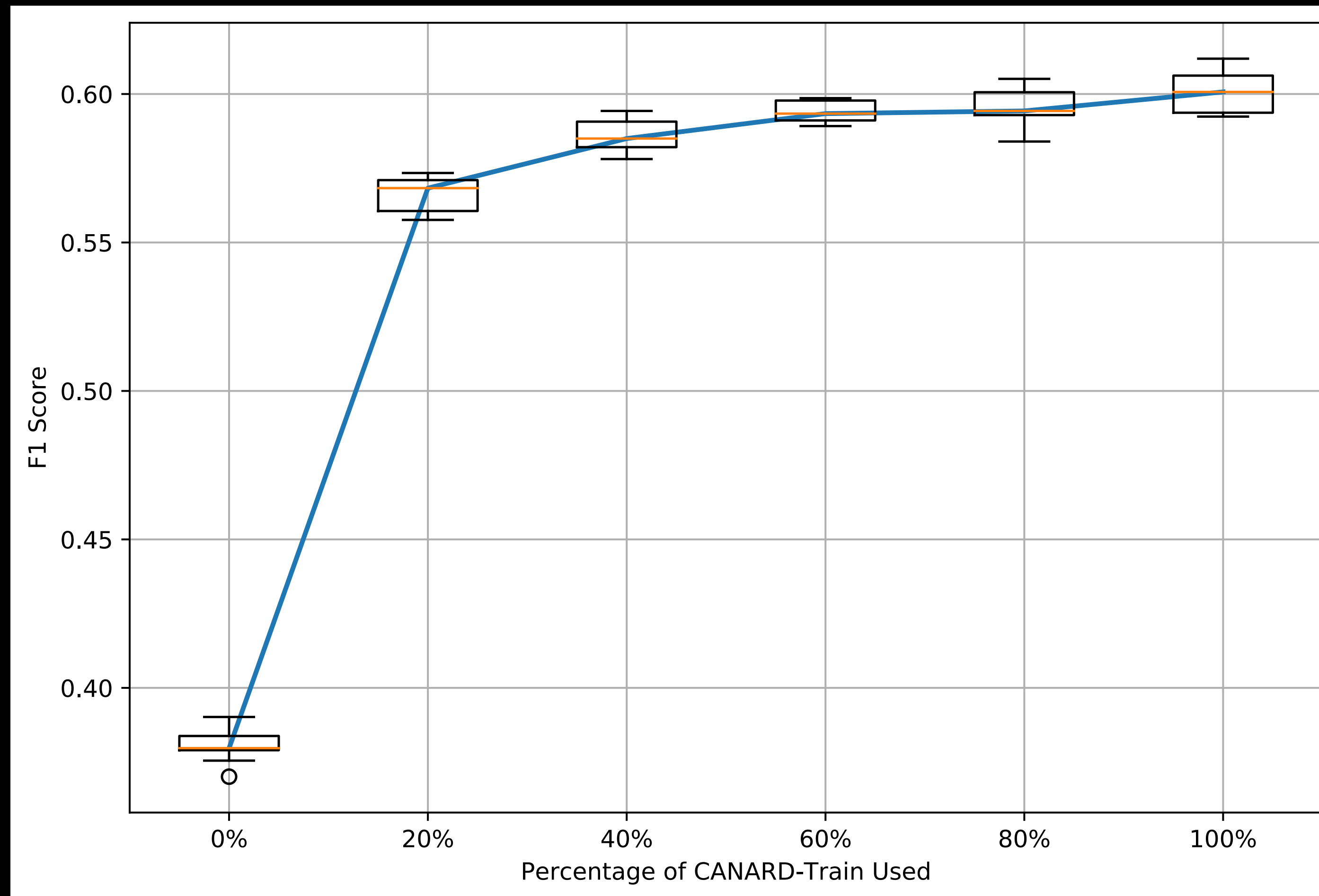
Results

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		41,93	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

Results

QuAC



Results

Error Analysis

		NDCG@3
Original	What types does olive oil contain?	0
QR	What types of fats does olive oil contain?	0.9
Human	What types of unsaturated fats does olive oil contain?	1

Results

Error Analysis

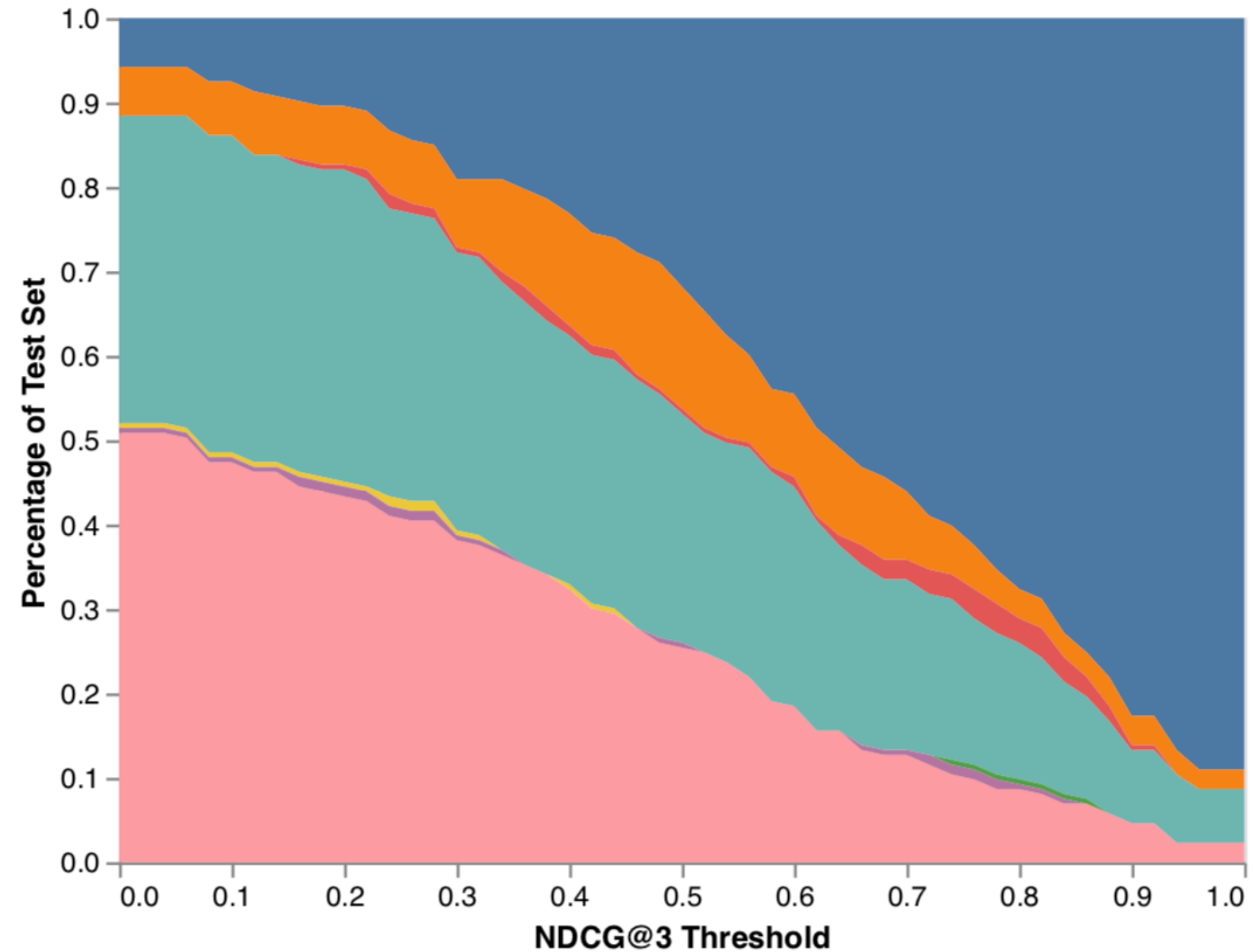
Original	QR	Human
×	×	×
✓	×	×
×	✓	×
✓	✓	×
×	×	✓
✓	×	✓
×	✓	✓
✓	✓	✓

		NDCG@3
Original	What types does olive oil contain?	0
QR	What types of fats does olive oil contain?	0.9
Human	What types of unsaturated fats does olive oil contain?	1

Results

Error Analysis

Passage ranking

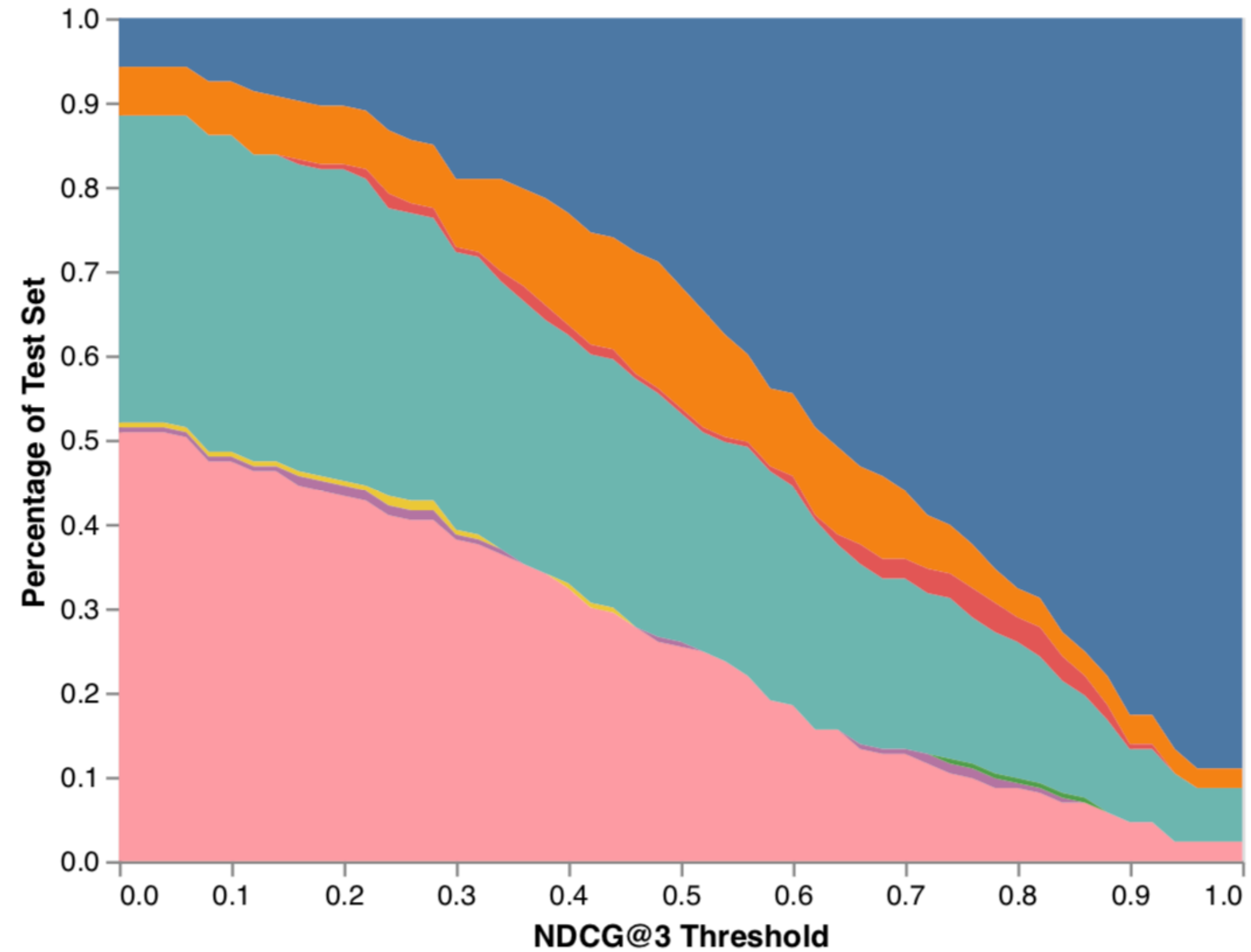


- Incorrect ranking
- Incorrect resolution
- Superhuman resolution
- Correct resolution
- Incorrect no resolution
- Correct no resolution

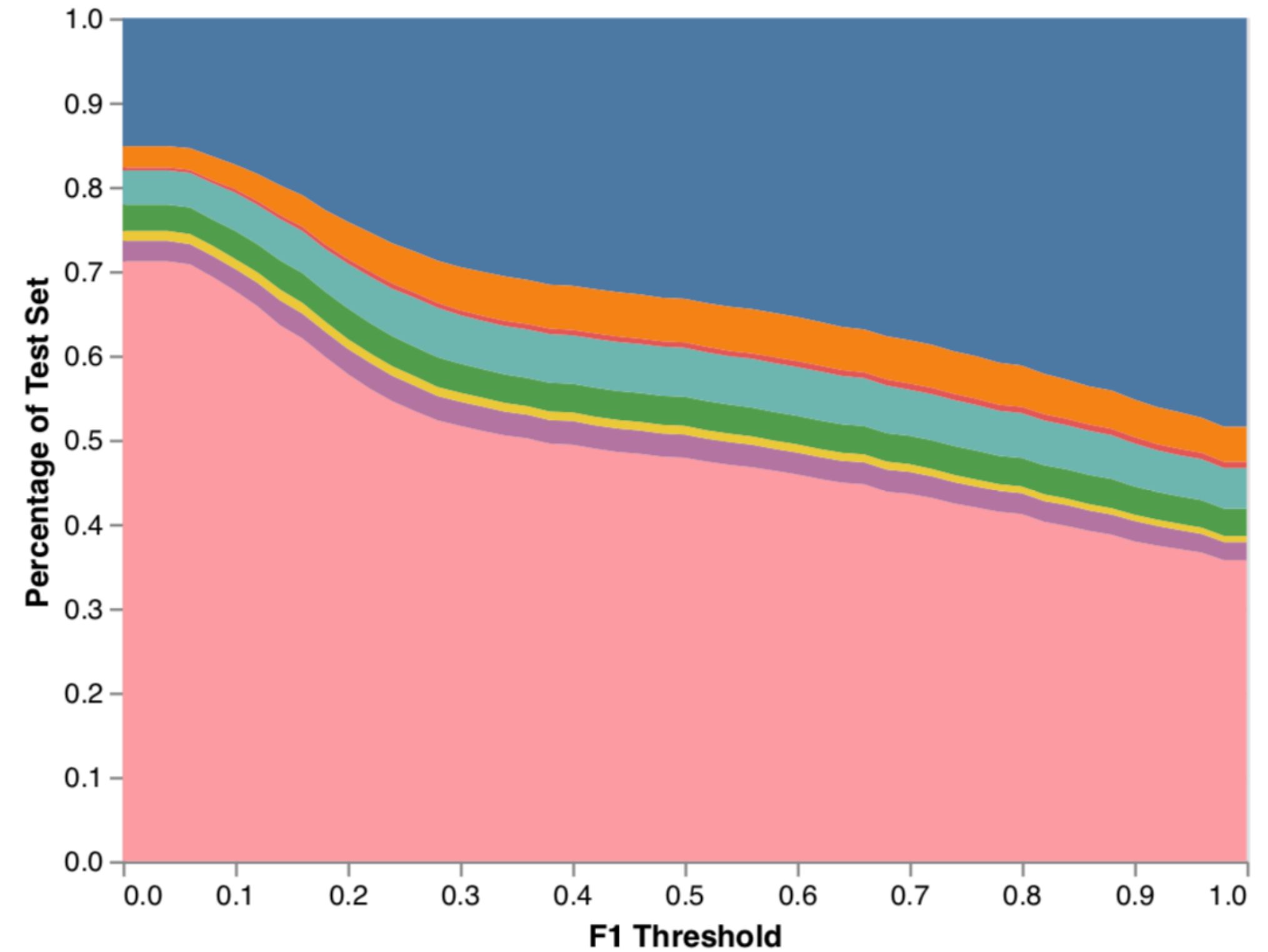
Results

Error Analysis

Passage ranking



Reading comprehension



Conclusion

Question Rewriting

Modular
Reusable
Debuggable
Cheap



QR model



QA model

References

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

Raviteja Anantha, Svitalana Vakulenko, Zhucheng Tu, Shayne Longpre, Stephen Pulman, Srinivas Chappidi. Open-Domain Question Answering Goes Conversational via Question Rewriting. 2021. (Submitted)

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

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*

Nikos Voskarides, Dan Li, Pengjie Ren, Evangelos Kanoulas, Maarten de Rijke: Query Resolution for Conversational Search with Limited Supervision. SIGIR 2020: 921-930.

Shi Yu, Jiahua Liu, Jingqin Yang, Chenyan Xiong, Paul Bennett et al. Few-Shot Generative Conversational Query Rewriting. SIGIR 2020: 1933-1936.

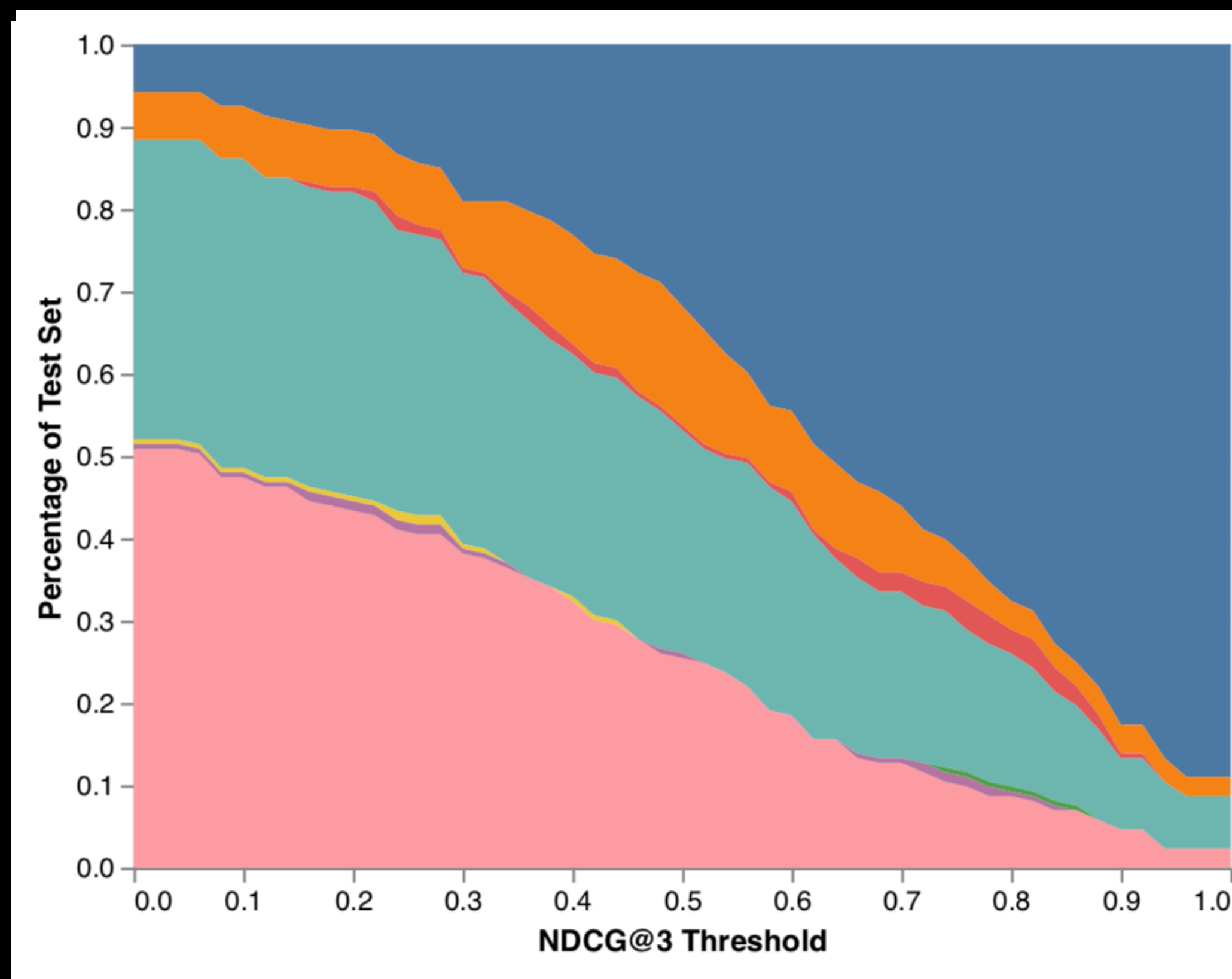
Ida Mele, Cristina Ioana Muntean, Franco Maria Nardini, Raffaele Perego et al. Topic Propagation in Conversational Search. SIGIR 2020: 2057-2060.

Results

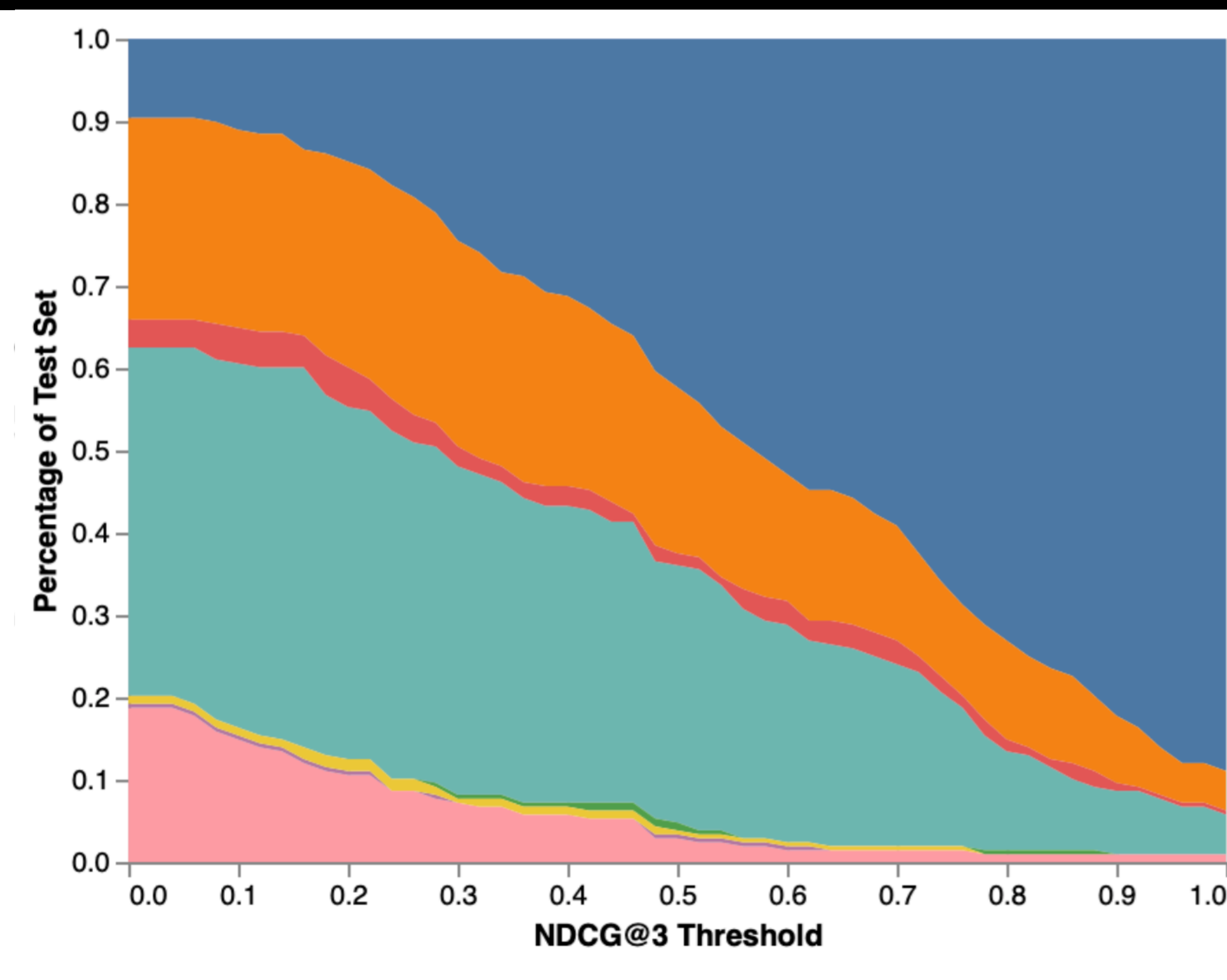
Error Analysis

github.com/svakulenk0/QRQA

Transformer++ on CAsT 2019



QuReTeC on CAsT 2020



QReCC Dataset

- 14K conversations with 81K question-answer pairs
- Collection: 10M web pages (split into 54M passages)
 - question rewrites
 - web page URLs
 - answer spans
- instructed to produce conversational answers
 - non-consecutive spans

