

# Consumer Health Question Answering Using Off-the-Shelf Components

April 2-6, 2023



Alexander Pugachev<sup>1</sup>



Ekaterina Artemova<sup>2</sup>



Alexander Bondarenko<sup>3</sup>



Pavel Braslavski<sup>1,4</sup>

<sup>1</sup> HSE University, Moscow, Russia

<sup>2</sup> Center for Information and Language Processing (CIS), MaiNLP Lab., LMU Munich, Munich, Germany

<sup>3</sup> Friedrich-Schiller-Universität Jena, Jena, Germany

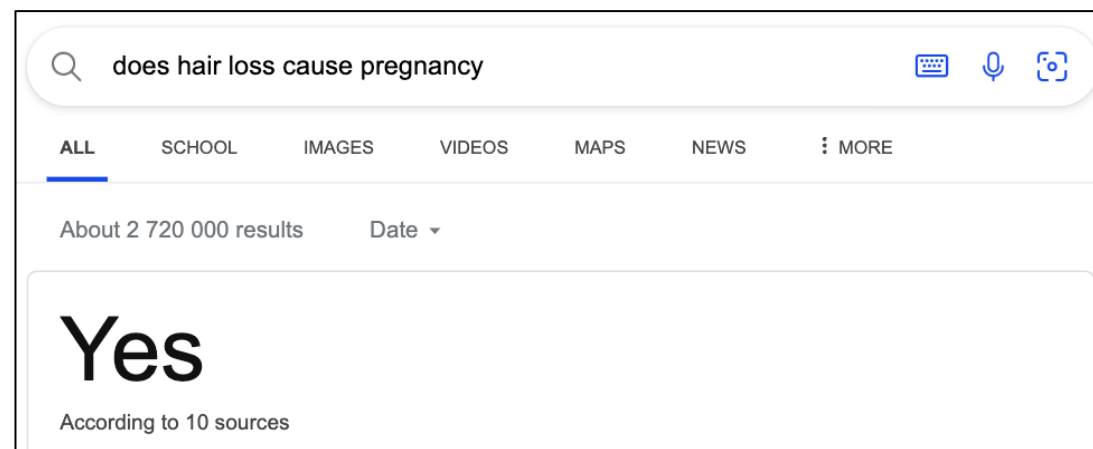
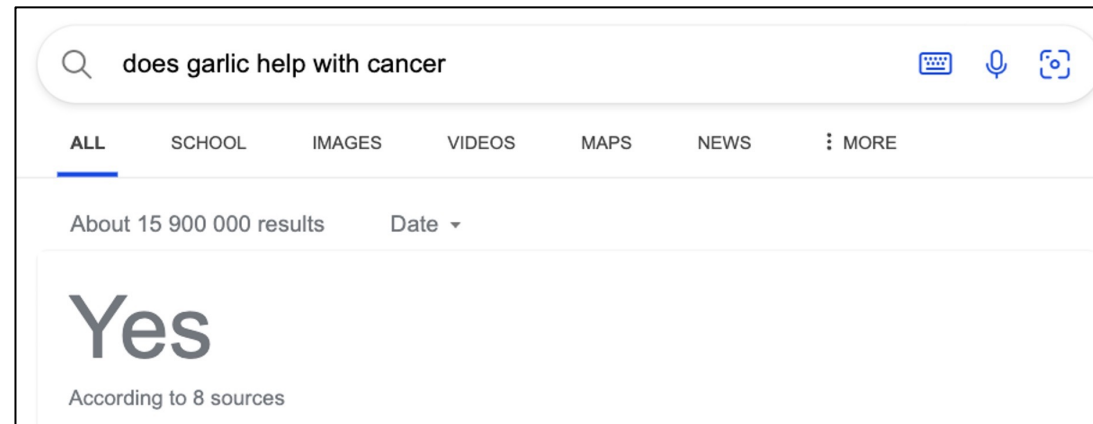
<sup>4</sup> Ural Federal University, Yekaterinburg, Russia



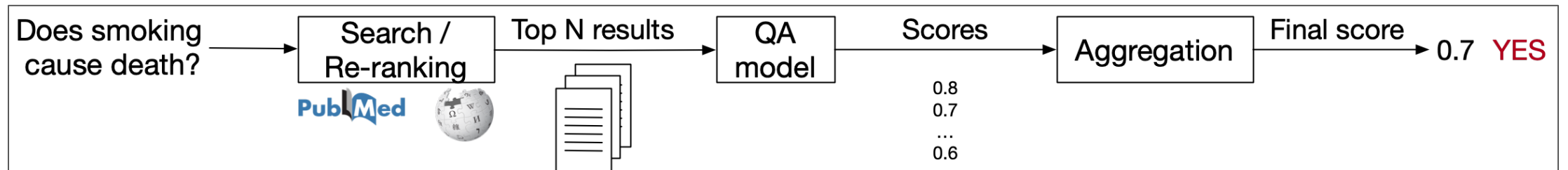
45th European Conference on Information Retrieval (ECIR 2023)

# Motivation

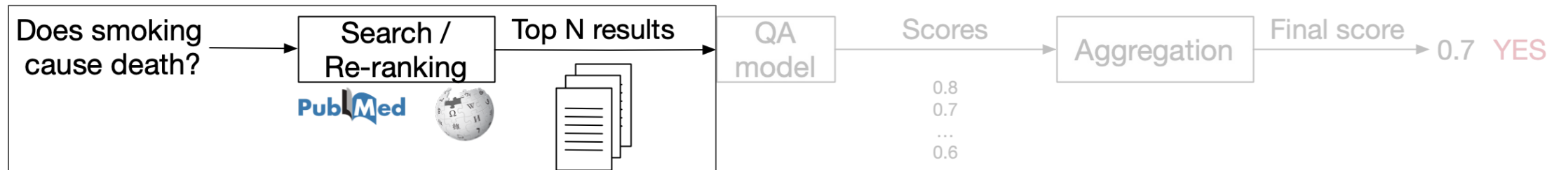
- About half of top-ranked search engines' results may provide **incorrect answers** to health-related questions [White, R.; SIGIR'13]
- It is important to provide the answers based on **reliable sources**
- We build our system from **ready-to-use** third-party blocks



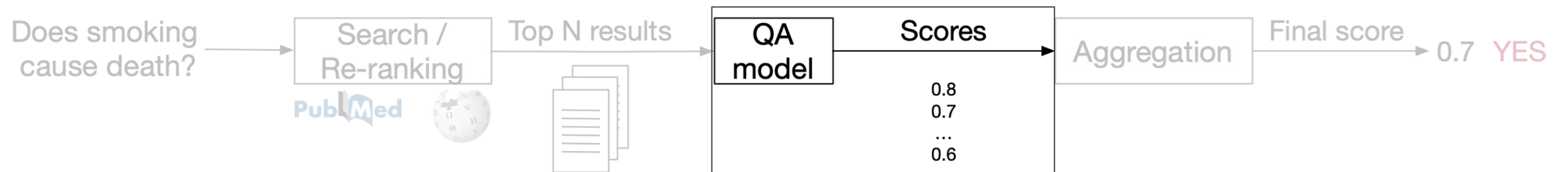
# Proposed Pipeline



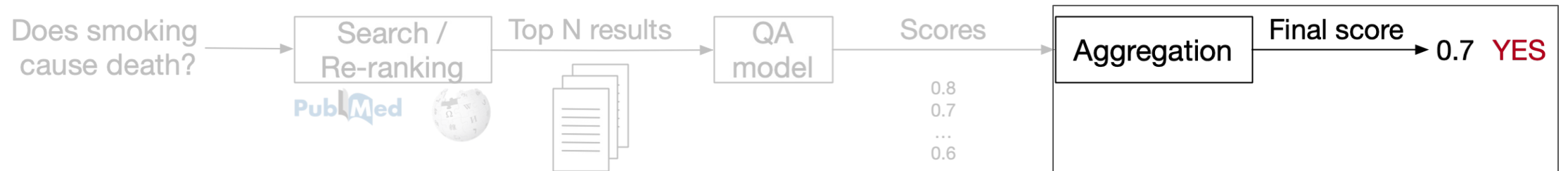
# Proposed Pipeline



# Proposed Pipeline



# Proposed Pipeline



# Experiments

## Retrievers

- PubMed
  - Native PubMed Search
  - Google Search over PubMed
  - Google's BioMed Explorer
- Wikipedia
  - Wikipedia API
  - Google Search over Wikipedia

## QA Models

- RoBERTa-large
  - Fine-tuned on BoolQ
- BioLinkBERT-large
  - Fine-tuned on PubMedQA
  - Fine-tuned on BioASQ

## Score Aggregation

- Top-1 document's score
- Plain average over top 10 documents
- Weighted average over top 10 documents

# Evaluation Data

- We use a collection of 113 yes/no health questions:

Source	# Questions (yes/no)
TREC (2019, 2021)	84 (42/42)
Yandex Log	15 (7/8)
Health Beliefs in Twitter	14 (12/2)

- Examples:

Question	Keyword Query	Label	Evident PubMed ID
Can caffeine help people with asthma?	caffeine asthma	1	20091514
Does aloe help for a runny nose?	aloe help runny nose	0	32924222
Does milk cause acne?	milk causes acne	1	



# Evaluation Results

	Retriever	Query	RoBERTa-large (BoolQ)			BioLinkBERT (PubMedQA)			BioLinkBERT (BioASQ)		
			top 1	avg	wavg	top 1	avg	wavg	top 1	avg	wavg
PubMed	PubMed	keywords	0.65	<u>0.66</u>	<u>0.66</u>	0.61	0.62	0.61	0.56	0.58	0.58
		question	0.61	0.62	0.62	0.57	0.58	0.58	0.58	0.48	0.48
	Google	keywords	<u>0.79</u>	0.73	0.73	0.59	0.66	0.68	0.71	0.65	0.69
		question	0.71	0.76	0.77	0.64	0.64	0.67	0.65	0.57	0.62
BioMed Explorer	keywords	0.72	0.75	0.74	0.63	0.64	0.66	0.58	0.74	0.71	
	question	0.74	<u>0.77</u>	<u>0.77</u>	0.60	0.69	0.69	0.72	0.71	0.73	
Wikipedia	Wikipedia	keywords	<u>0.81</u>	0.77	0.78	0.68	0.65	0.68	0.58	0.57	0.59
		question	0.65	0.68	0.68	0.57	0.56	0.58	0.48	0.48	0.47
	Google	keywords	0.80	0.79	<b>0.82</b>	0.63	0.72	0.68	0.55	0.57	0.57
		question	0.75	0.75	0.75	0.57	0.67	0.64	0.52	0.53	0.56

The values represent ROC-AUC scores

# Conclusion

- Our solution exploits evidence search in PubMed and Wikipedia for open-domain health question answering
- The Google Search over Wikipedia combined with RoBERTa-large model achieves **AUC of 0.82** on 113 test yes/no questions
- The proposed approach **does not use annotated data** directly and **does not require training** on the target data or task
- Our source code and data are publicly available on GitHub<sup>1</sup>



<sup>1</sup> <https://github.com/apugachev/consumer-health-qa>

# Conclusion

- Our solution exploits evidence search in PubMed and Wikipedia for open-domain health question answering
- The Google Search over Wikipedia combined with RoBERTa-large model achieves **AUC of 0.82** on 113 test yes/no questions
- The proposed approach **does not use annotated data** directly and **does not require training** on the target data or task
- Our source code and data are publicly available on GitHub<sup>1</sup>
- A huge gratitude to the organizing committee for providing the student support



<sup>1</sup> <https://github.com/apugachev/consumer-health-qa>