

Neural Approaches to Medical Question Understanding

Asma Ben Abacha, PhD
Research Scientist

Yassine Mrabet, PhD
Solutions Architect

PI: Dina Demner-Fushman, MD, PhD

LISTER HILL CENTER (LHC)
NATIONAL LIBRARY OF MEDICINE
NATIONAL INSTITUTES OF HEALTH

NIH.AI Workshop on Natural Language Processing

May 9, 2019



U.S. National Library of Medicine

Consumer Health Questions

Actual examples sent to the NLM:

- A. treatment for dry mouth caused by necessary medicine. My provider can't help (I asked.) I am intolerant of all the sugar alcohols such as maltitol, sorbitol, xylitol, etc. and need something for dry mouth caused by med which I have to take. Biotene products help for only about two minutes.
- B. I'm a mother of a 13 year old girl that has been diagnosed with endometriosis. I know its not spelled right. The doctor stated it was severe and uncommon for children her age. she was 12 when diagnosed is there anything going on for her age? she has not had any treatments and hurts constantly?
- C. I suffered a massive stroke on [DATE] with paralysis on my left side of my body, I'm home and conduct searches on the internet to find help with recovery, and always this product called neuroaid appears claiming to restore function. to my knowledge it isn't approved by the FDA, but it sounds so promising. do you know anything about it and id there anything approved by our FDA, that does help



Consumer Health Questions

Actual examples sent to the NLM:

- A. **treatment** for **dry mouth** **caused** by necessary medicine. My provider can't help (I asked.) I am intolerant of all the **sugar alcohols** such as **maltitol**, **sorbitol**, **xylitol**, etc. and **need something for dry mouth caused by med** which I have to take. **Biotene** products help for only about two minutes.
- B. I'm a mother of a 13 year old girl that has been diagnosed with endometriosis. I know its not spelled right. The doctor stated it was severe and uncommon for children her age. she was 12 when diagnosed is there anything going on for her age? she has not had any treatments and hurts constantly?
- C. I suffered a massive stroke on [DATE] with paralysis on my left side of my body, I'm home and conduct searches on the internet to find help with recovery, and always this product called neuroaid appears claiming to restore function. to my knowledge it isn't approved by the FDA, but it sounds so promising. do you know anything about it and id there anything approved by our FDA, that does help



CHiQA

Consumer Health QA Project (CHiQA) @ the LHC / NLM



First prototype implementing
some of the research outcomes

<https://chiqa.nlm.nih.gov>



CHiQA
beta

CHiQA is an experimental AI system that is learning how to answer health-related questions using reliable sources for patients. Please note that we provide information. We cannot respond to questions about individual medical cases, provide second opinions or make specific recommendations regarding therapy. Those issues should be addressed directly with your health care provider.



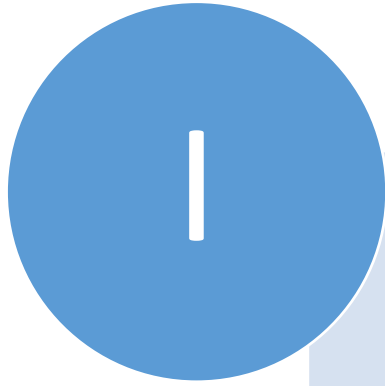
Secure | <https://chiqa.nlm.nih.gov/answer?question=How+to+stop+taking+bisoprolol>

NIH NLM CHiQA beta How to stop taking bisoprolol About API

How to stop taking bisoprolol

Bisoprolol comes as a tablet to take by mouth. It is usually taken once a day. To help you remember to take **bisoprolol**, take it around the same time every day. Follow the directions on your prescription label carefully, and ask your doctor or pharmacist to explain any part you do not understand. Take **bisoprolol** exactly as directed. Do not take more or less of it or take it more often than prescribed by your doctor. Your doctor will probably start you on a low dose of **bisoprolol** and gradually increase your dose. **Bisoprolol** controls high blood pressure but does not cure it. It may take a few weeks before you feel the full benefit of **bisoprolol**. Continue to take **bisoprolol** even if you feel well. Do not **stop taking bisoprolol** without talking to your doctor. Suddenly stopping **bisoprolol** may cause angina (chest pain), heart attack, or irregular heartbeat. Your doctor will probably decrease your dose gradually.

[- show less](#) [Bisoprolol \(How should this medicine be used?\)](#)



Recognizing
Question
Entailment

Question
Summarization



Extraction of
Question Focus
& Type

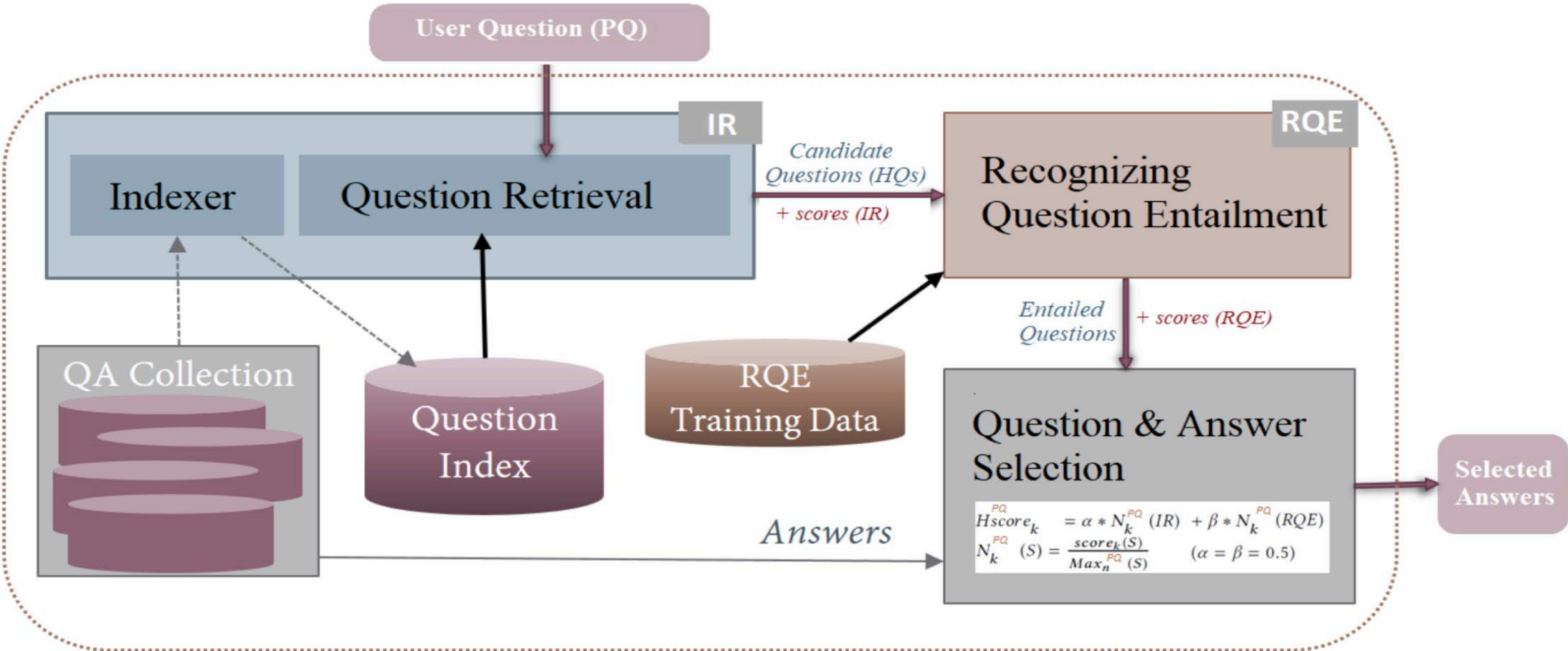
Answer
Retrieval



RQE-based Approach

→ Answering new questions by retrieving entailed questions with existing answers.

Ben Abacha & Demner-Fushman. AMIA 2016



RQE Models

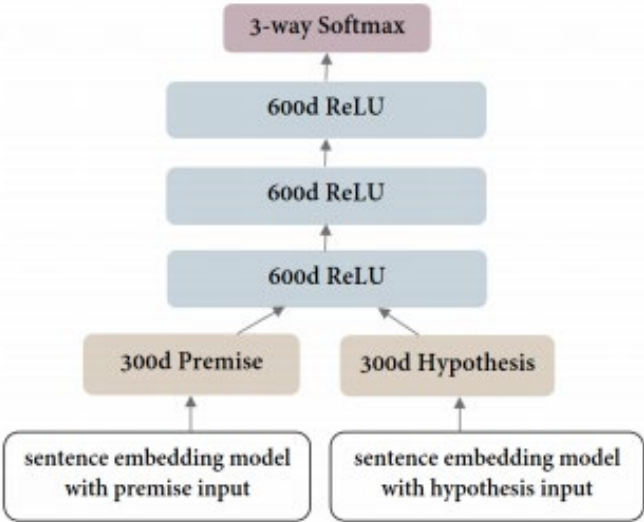
Methods	Training Datasets			
	SNLI	MultiNLI	Quora	Clinical-QE
Neural Network (NN)	48.94	54.59	52.35	48.71
NN + GloVe embeddings	49.41	54.82	52.82	57.18
Logistic Regression + Features	67.05	64.94	52.11	73.18

Ben Abacha & Demner-Fushman. ArXiv 2019

Observations:

→ Logistic Regression outperformed neural networks when trained with traditional word embeddings such as Glove and Word2Vec.

→ Relying on recent language models such as Bert for pre-training led to a better performance.



MedQuAD Dataset for Medical QA

□ Medical question-answer collection created from trusted resources.

Ben Abacha & Demner-Fushman. ArXiv 2019

- Contains **47k question-answer pairs** extracted from **12 NIH websites** (e.g. MedlinePlus, NCI, GARD, NIDDK).
- Covers 16 question types about **Diseases** (e.g. Treatment, Susceptibility), 20 types about **Drugs** (e.g. Usage, Interaction), and an additional type (Information) for all possible question foci (e.g. **Procedure, Medical exam**).

Link: <https://github.com/abachaa/MedQuAD>



U.S. National Library of Medicine

Question Summarization

➔ Tackling the complexity of consumer questions by automatic summarization.

Question: **polymicrogyria**. My 16 month old son has this. Does not sit up or crawl yet but still trying and is improving in grabbing things etc etc. Have read about other cases that seem 10000 time worse. It's it possible for this post of his brain to grown to normal and he grow out of it?

➔ **Summary:** What is the **prognosis** for **polymicrogyria**?

Question: I have had many **gout attacks** since i have been 30 years old now 70. I take **allopurinol** and **blood pressure meds**. Before i took **allopurinol** i never had **high blood pressure**. Also i have developed **basal cell skin cancer** i have heard **allopurinol** will cause that. **Reduces acid in your system?**

➔ **Summary:** What are the **side effects** of **allopurinol**?



QA performance w(o)/ Summarization

Measures	M1-TQs	M3-SUM
avgScore(0-3)	0.711	1.125
succ@2+	0.442	0.663
succ@3+	0.192	0.317
succ@4+	0.077	0.144
prec@2+	0.46	0.663
prec@3+	0.2	0.317
prec@4+	0.08	0.144

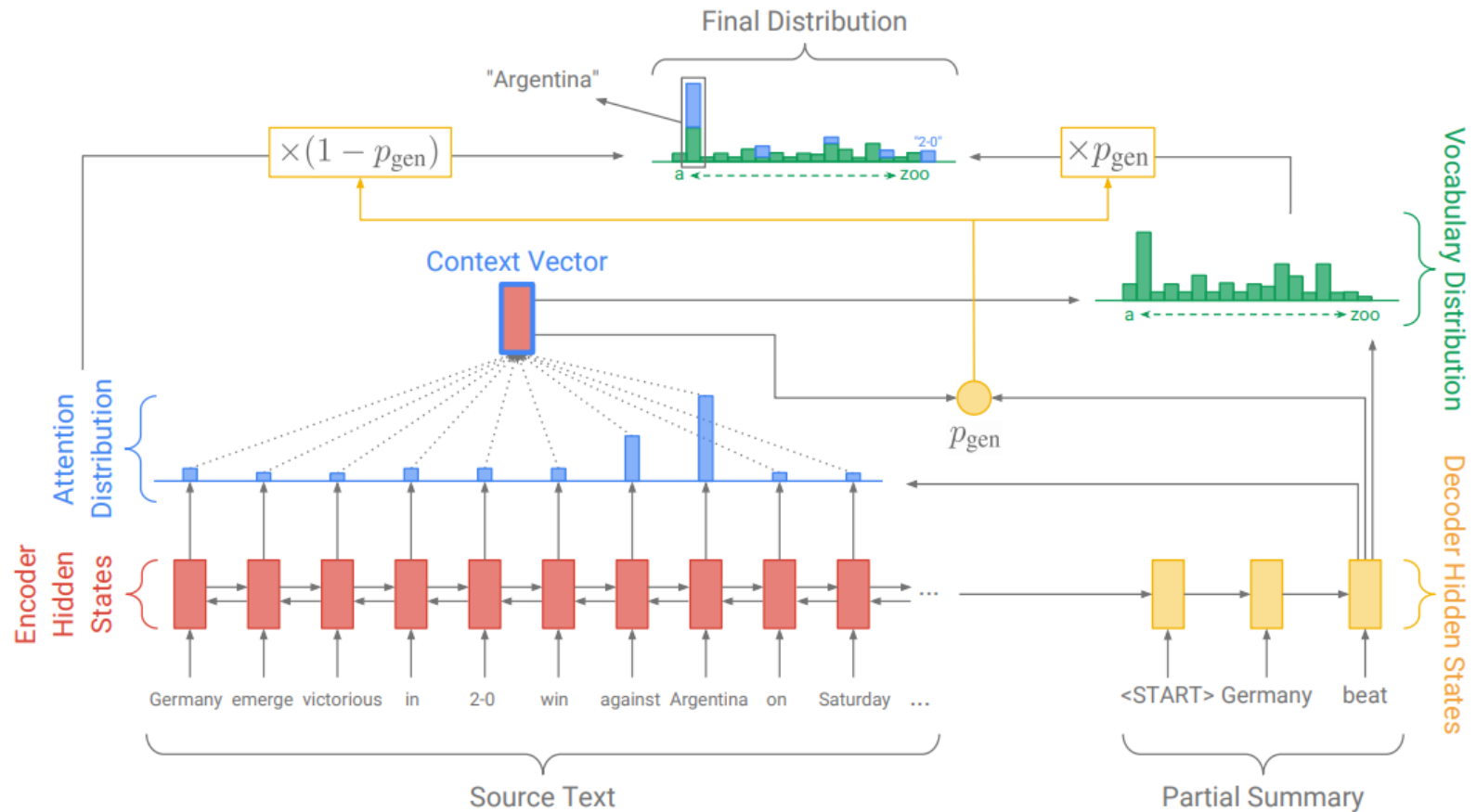
Ben Abacha & Demner-Fushman. AMIA 2019 Informatics Summit



U.S. National Library of Medicine

Pointer-generator network

See et al., ACL 2017



Results

Method	Rouge-1
Seq2seq Attentional Model	27.62
Pointer Generator	44.16

Ben Abacha & Demner-Fushman. ACL 2019

Currently displaying: `attn_vis_data.json`

Question

dvt . can a birth control called ocella cause dvt ? my daughter experiences pains cramping , redness and swelling in her thigh and also really bad huge blood clots during her menstrual cycles after she was prescribed osella for birth control . also these __syntoms__ worsened after she gave birth . this has been happening for a year now should she see discuss this with her doctor right away ?

Reference summary

can birth control drug ocella cause deep vein thrombosis .

Generated summary (highlighted = high generation probability)

can ocella cause dvt ?



U.S. National Library of Medicine

MEDIQA @ ACL-BioNLP 2019

Textual Inference, Question Entailment & Question Answering

ACL-BioNLP'19 Shared Task

MEDIQA 2019

Textual Inference and Question Entailment in the Medical Domain

Introduction

The MEDIQA challenge aims to attract further research efforts in Natural Language Inference (NLI), Recognizing Question Entailment (RQE), and their applications in medical Question Answering (QA). This [ACL-BioNLP 2019](#) shared task is motivated by a need to develop relevant methods, techniques and gold standards for inference and entailment in the medical domain and their application to improve domain specific IR and QA systems.

Tasks

1) NLI: This first task consists in identifying three **inference** relations between two **sentences**: *Entailment*, *Neutral* and *Contradiction* [1]

2) RQE: This task focuses on identifying **entailment** between two **questions** in the context of QA. We use the following definition of question entailment: "a question *A* entails a question *B* if every answer to *B* is also a complete or partial answer to *A*" [2]

3) QA: The objective of this task is to **filter** and **improve the ranking** of automatically retrieved answers. The input ranks are generated by the medical QA system [ChiQA](#). We highly recommend the reuse of RQE and/or NLI systems (first tasks) in the QA task [3-5]

Organizers

- [Asma Ben Abacha](#), *Lister Hill Center, NLM, Bethesda, MD.*
- [Chaitanya Shivade](#), *IBM, Almaden Research Center, San Jose, CA.*
- [Dina Demner-Fushman](#), *Lister Hill Center, NLM, Bethesda, MD.*

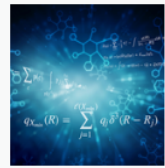
Join our mailing list: <https://groups.google.com/d/forum/bionlp-mediqa>

Important Dates (Tentative)

- February 8, 2019: AICrowd projects go public: NLI@AICrowd, RQE@AICrowd & QA@AICrowd.
- February 28, 2019: Release of the RQE validation set, [run submission open](#)



U.S. National Library of Medicine

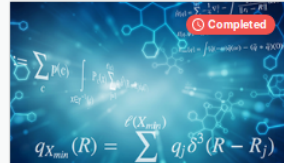


MEDIQA@ACL-BioNLP

Shared Task on Textual Inference and Question Entailment in the Medical Domain

Challenges Members

Create Challenge Edit organizer

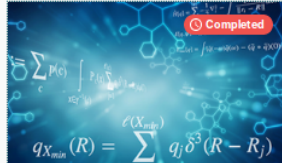


MEDIQA 2019 - Recognizing Question Entailment (RQE)

ACL-BioNLP Shared Task

2837 112 287

By MEDIQA@ACL-BioNLP

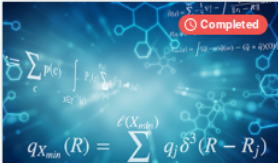


MEDIQA 2019 - Natural Language Inference (NLI)

ACL-BioNLP Shared Task

2607 88 170

By MEDIQA@ACL-BioNLP



MEDIQA 2019 - Question Answering (QA)

ACL-BioNLP Shared Task

1773 61 110

By MEDIQA@ACL-BioNLP

MEDIQA 2019 - Natural Language Inference (NLI)

ACL-BioNLP Shared Task

By MEDIQA@ACL-BioNLP

14 Follow

Overview **Leaderboard** ...

Round 1 **Round 2**

Δ #	Participant	Accuracy	Entries
▲ 01	WTMed	0.980	0.000 3
• 02	qwertyuiop	0.975	0.000 5
• 03	PANLP	0.966	0.000 4
• 04	Your_Baseline	0.938	0.000 5
• 05	Sieg	0.911	0.000 3
▲ 06	MEDIQA	0.909	0.000 5
▲ 07	Surf	0.906	0.000 4
▼ 08	stephen_curry	0.899	0.000 5
• 09	ARS_NITK	0.877	0.000 3
• 10	Pentagon	0.857	0.000 4

MEDIQA 2019 - Recognizing Question Entailment (RQE)

ACL-BioNLP Shared Task

By MEDIQA@ACL-BioNLP

14 Follow

Overview **Leaderboard** ...

Round 1 **Round 2**

Δ #	Participant	Accuracy	Entries
• 01	PANLP	0.749	0.000 5
• 02	Lizimu	0.745	0.000 5
▲ 03	Keke	0.710	0.000 5
• 04	Sieg	0.706	0.000 4
• 05	nlp_gawds	0.684	0.000 5
• 06	IIT-KGP	0.684	0.000 3
• 07	Pentagon	0.671	0.000 4
• 08	ARS_NITK	0.667	0.000 5
• 09	aqldem	0.667	0.000 5
• 10	QEntailment	0.667	0.000 5

MEDIQA 2019 - Question Answering (QA)

ACL-BioNLP Shared Task

By MEDIQA@ACL-BioNLP

12 Follow

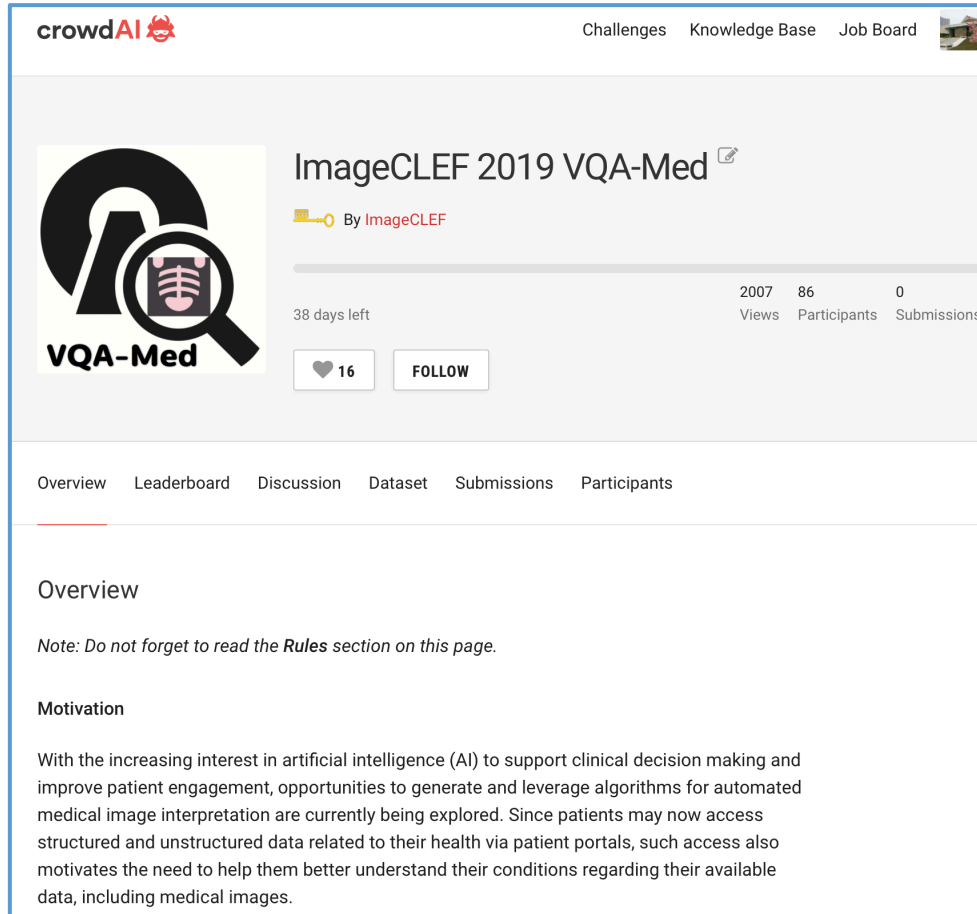
Overview **Leaderboard** ...

Round 1 **Round 2**

Δ #	Participant	Accuracy	Spearman's rho
• 01	nlpHUST	0.783	0.209
▼ 02	Your_Baseline	0.780	0.238
• 03	PANLP	0.777	0.180
• 04	Pentagon	0.765	0.338
• 05	have_fun	0.755	0.109
• 06	unk	0.754	0.342
• 07	threeTreeStar	0.745	0.106
• 08	BigOrange	0.745	0.106
▼ 09	stephen_curry	0.734	0.170
• 10	fjgjjg	0.733	0.354

VQA-Med@ImageCLEF 2019

Visual Question Answering in the Medical Domain



The screenshot shows the challenge page for 'ImageCLEF 2019 VQA-Med' on the crowdAI platform. The page includes a header with navigation links for 'Challenges', 'Knowledge Base', and 'Job Board'. The challenge title is 'ImageCLEF 2019 VQA-Med' by 'ImageCLEF'. It shows 2007 views, 86 participants, and 0 submissions, with 38 days left. There are 16 likes and a 'FOLLOW' button. Below the challenge details are tabs for 'Overview', 'Leaderboard', 'Discussion', 'Dataset', 'Submissions', and 'Participants'. The 'Overview' section contains a note to read the rules and a 'Motivation' paragraph discussing the use of AI in clinical decision making.

crowdAI Challenges Knowledge Base Job Board

ImageCLEF 2019 VQA-Med

By ImageCLEF

38 days left

2007 Views 86 Participants 0 Submissions

16 FOLLOW

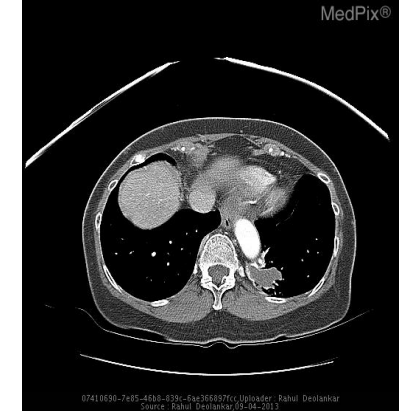
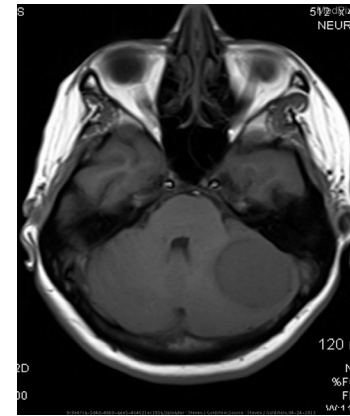
Overview Leaderboard Discussion Dataset Submissions Participants

Overview

Note: Do not forget to read the Rules section on this page.

Motivation

With the increasing interest in artificial intelligence (AI) to support clinical decision making and improve patient engagement, opportunities to generate and leverage algorithms for automated medical image interpretation are currently being explored. Since patients may now access structured and unstructured data related to their health via patient portals, such access also motivates the need to help them better understand their conditions regarding their available data, including medical images.



Examples:

- is this a t1 weighted, t2 weighted, or flair image?
- in which plane is the mri displayed?
- what is the organ principally shown in this x-ray?
- what is most alarming about this ct scan?



Extraction of
Question
Focus & Type

Answer
Retrieval

Premise

- The question focus (topic) is the main entity in the question [1]
 - E.g., “Alzheimer’s” in “How to diagnose Alzheimer’s disease?”
- The question type is the aspect of interest
 - E.g., diagnosis, treatment, side effect, cause.
- Previous investigations showed that using only the focus and type as search terms allowed to answer 61% of consumer questions [2].
- An additional 12% can be answered using more terms from the question as keywords.

[1] Mrabet, Y., Kilicoglu, H., Roberts, K., & Demner-Fushman, D. (2016).

Combining open-domain and biomedical knowledge for topic recognition in consumer health questions.

In AMIA Annual Symposium Proceedings (Vol. 2016, p. 914). American Medical Informatics Association.

[2] Deardorff A, Masterton K, Roberts K, Kilicoglu H, Demner-Fushman D.

A protocol-driven approach to automatically finding authoritative answers to consumer health questions in online resources.

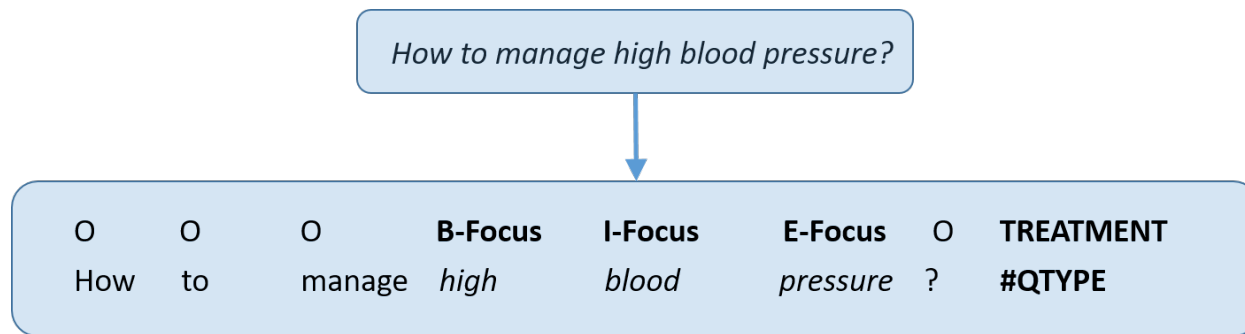
Journal of the Association for Information Science and Technology. 2017 Jul;68(7):1724-36.



U.S. National Library of Medicine

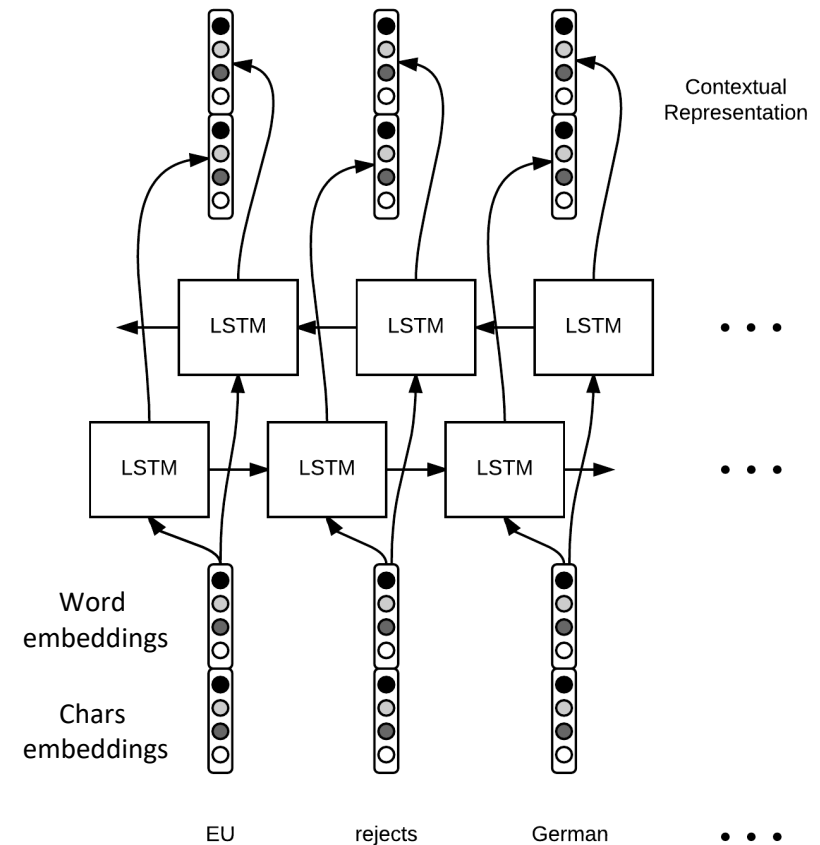
Deep Learning Approach

Joint (Frame) Recognition



Recurrent Bi-directional LSTM networks with CRF prediction:

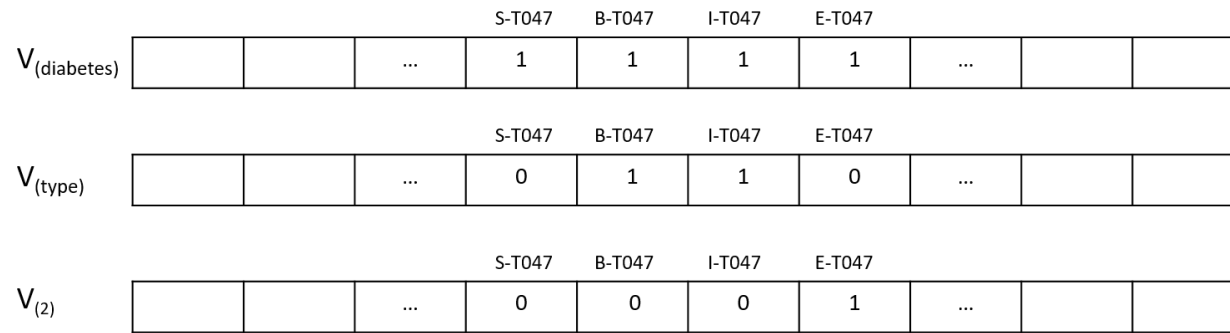
- First bi-LSTM to build character embeddings
- Second word-level bi-LSTM with several embedding sources:
 - Glove word embeddings
 - UMLS embeddings
 - POS embeddings
 - TF-IDF based vectors



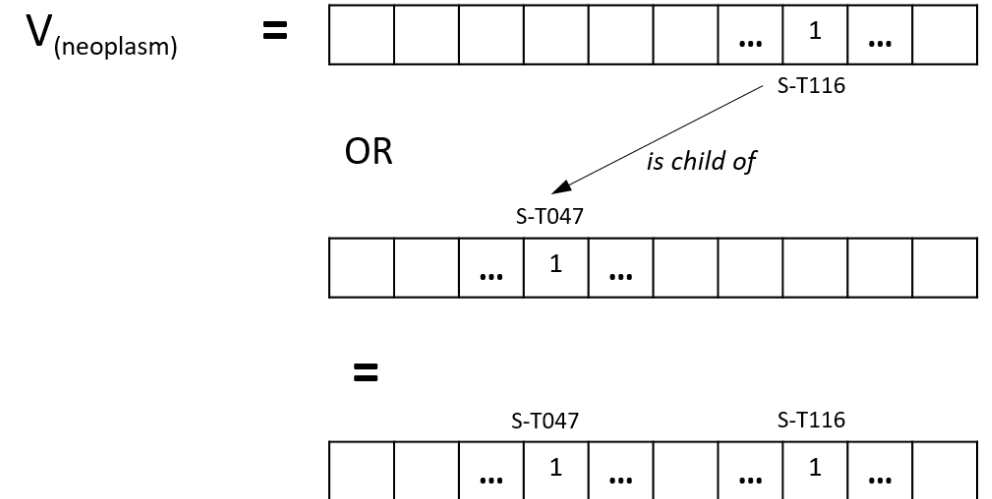
<https://guillaumegethial.github.io/sequence-tagging-with-tensorflow.html>

UMLS Taxonomic Embeddings

A) Binary vector representation in BIOES format



B) Value propagation using taxonomic hierarchy

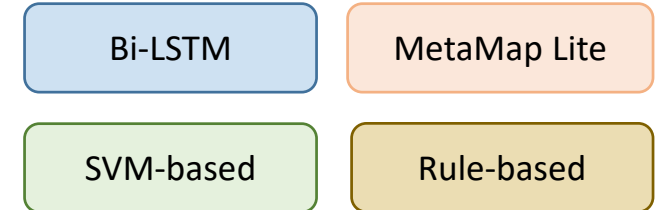


C) Experiments included several terminology-based variants:

- Head words vs all words
- TF-IDF values instead of binary vectors (propagation ratio set to 0.5)
- TF only.

CHiQA Prototype (Ensemble Methods)

- Question Focus Recognition
 - Token-level voting



How to treat **dry mouth** caused by **Sorbitol?** **Biotene** products did not work more than a few **minutes.**

➔ Selected Focus: “dry mouth”

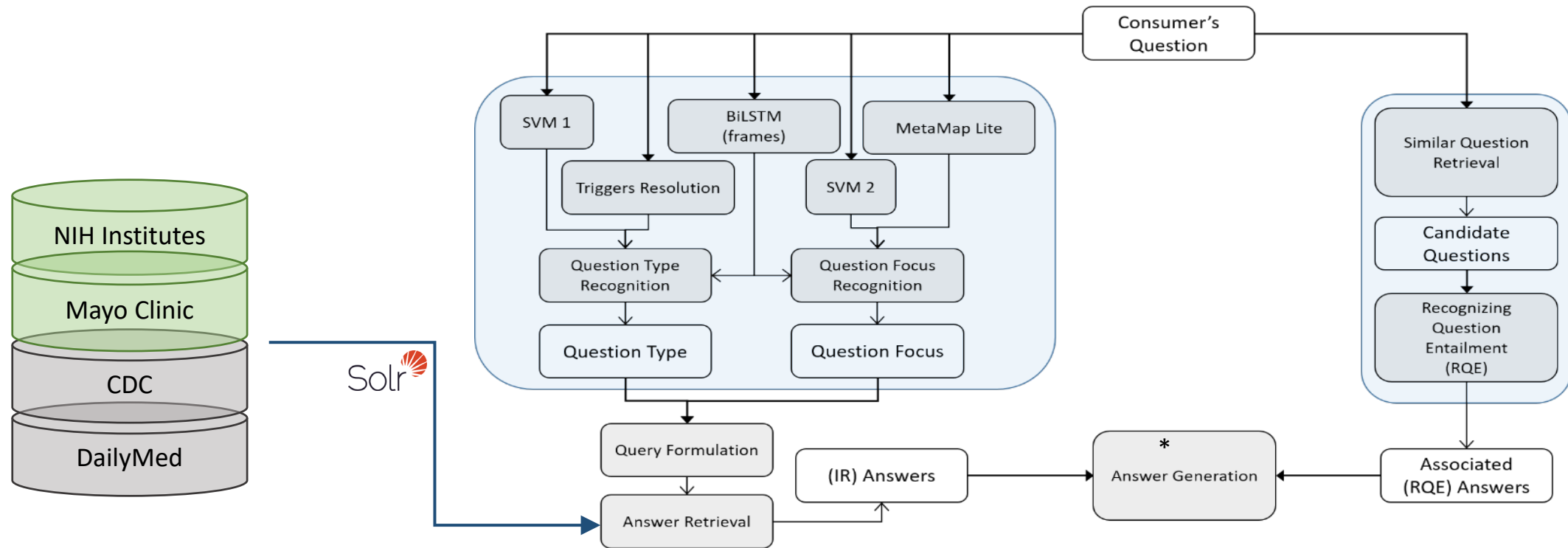
- Question Type Recognition
 - Conventional voting method

How to **treat** dry mouth **caused** by Sorbitol? Biotene products did not **work** more than a few minutes.

➔ Selected Type: TREATMENT



CHiQA Prototype (System Architecture)



[3] Mrabet, Y., Kilicoglu, H., & Demner-Fushman, D. (2017, July).
TextFlow: A text similarity measure based on continuous sequences.
 In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics
 (Volume 1: Long Papers) (pp. 763-772).

* Similar answers filtered out using the **TextFlow** similarity measure [3]

CHiQA Prototype

(Interface and end-to-end evaluation)

The screenshot shows the CHiQA Prototype web interface. The browser address bar displays the URL: <https://chiqa.nlm.nih.gov/answer?question=How+to+diagnose+Alzheimer%27s+Disease>. The page header includes the NIH NLM logo, the CHiQA logo, a search bar with the text "Ask a question?", and links for "About" and "API".

The main content area is titled "How to diagnose Alzheimer's Disease" and features several text-based search results, each with a "show more" link and a small blue square icon. The results include:

- Early-onset, autosomal dominant **Alzheimer disease** is **diagnosed** in families that have more than one member with AD in which the
- A skilled health care provider can often **diagnose** AD with the following steps:
 - Performing a complete physical **exam**,
- The following diseases are related to Familial **Alzheimer disease**. If you have a question about any of these diseases, you can contact
- Familial **Alzheimer disease** (familial AD) is a degenerative **disease** of the brain that causes gradual loss of memory, judgment, and the

On the right side of the page, there is a "Related Questions" section with the following questions:

- What is Alzheimer disease?
- What causes Alzheimer disease?
- How many people are affected by Alzheimer disease?

The footer contains the following text: "Lister Hill National Center for Biomedical Communications • U.S. National Library of Medicine • 8600 Rockville Pike, Bethesda, MD 20894". It also includes logos for NIH, U.S. National Library of Medicine, LHMCC, and USA.gov, along with a copyright notice: "Copyright • Privacy • Accessibility • Freedom of Information Act • Viewers and Players • Contact Us".

Metrics	LiveQA-Med Dataset			
	IR	RQE	COMB	LiveQA Best
Average Score	1.183	0.827	1.308	0.637
MAP@10	0.405	0.311	0.445	--
MRR@10	0.438	0.333	0.516	--

Benchmark from the Medical Question Answering Competition @ TREC 2017 LiveQA



Conclusion

- Relying on the retrieval of entailed questions is a viable strategy to answer consumer health questions.
- Limiting the number of answer sources can enhance the performance of QA systems.
- Paraphrasing and summarizing the questions leads to a substantial improvement.
- Deep learning models for question understanding provided a good performance with a distinct perspective, but we still need external supervision with ensemble methods for end-to-end systems.



Future work

- Relevant techniques for question reformulation and answer selection.
- Relevant methods for answering more contextual questions such as questions about medications.
- Answers summarization and aggregation.
- Enhance answer retrieval with a relevant embedding space.
- Neural ensemble methods.



Thank you for your attention!

asma.benabacha@nih.gov

mrabety@mail.nih.gov

ddemner@mail.nih.gov

Resources:

- Papers: <https://www.researchgate.net/project/Consumer-Health-Question-Answering>
- CHiQA Project: <https://lhncbc.nlm.nih.gov/project/consumer-health-question-answering>
- Data: <https://github.com/abachaa>



U.S. National Library of Medicine