MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190176

Bridging the Gap Between Consumers' Medication Questions and Trusted Answers

Asma Ben Abacha, Yassine Mrabet, Mark Sharp,

Travis R. Goodwin, Sonya E. Shooshan and Dina Demner-Fushman

Lister Hill National Center of Biomedical Communications
National Library of Medicine, Bethesda, MD

Abstract

This paper addresses the task of answering consumer health questions about medications. To better understand the challenge and needs in terms of methods and resources, we first introduce a gold standard corpus for Medication Question Answering created using real consumer questions. The gold standard[1] consists of six hundred and seventy-four question-answer pairs with annotations of the question focus and type and the answer source. We first present the manual annotation and answering process. In the second part of this paper, we test the performance of recurrent and convolutional neural networks in question type identification and focus recognition. Finally, we discuss the research insights from both the dataset creation process and our experiments. This study provides new resources and experiments on answering consumers' medication questions and discusses the limitations and directions for future research efforts.

Keywords:

Health Informatics, Data Collection, Natural Language Processing

Introduction

Digital information and the World Wide Web make it possible to find information on demand, helping both patients and professionals to find relevant and valuable medical information. With these benefits come growing concerns about the potential misinformation of patients and non-expert consumers. Harmful consequences include self-diagnosis and unfounded anxiety about common symptoms after reviewing online health information (i.e., cyberchondria) [1]. The Pew Research Center reports that 72% of U.S. internet users (over 200 million) have gone online in the past year specifically for health-related information. Of that group, 77% say that their research started with search engines such as Google, while only 13% say they began at a specialized website such as WebMD[2].

Safe and efficient search alternatives are needed to protect non-expert consumers from misleading health information found online. Specialized Question Answering (QA) systems are a potential solution for the medical domain and the consumers' need for reliable health information [2]. Several research efforts tackled the problem of QA in the (bio)medical domain and highlighted the challenges related to question understanding, answer retrieval, and ranking [3-6]. To the best of our knowledge, no studies were previously dedicated to QA about medications.

Figure 1– Word Cloud Representing the Consumer Questions about Drugs that We Used to Create the Gold Standard Corpus.

While several previous studies included such questions as a part of larger and more heterogeneous QA tasks, we argue that questions about medications have specific characteristics that need to be analyzed and evaluated in a focused study. These aspects include (i) different question types (e.g. action, storage, usage time, stopping, and tapering), (ii) more conditional answers (e.g. ingredients and appearance depend often on the manufacturer), (iii) different answer sources having a specific data structure (e.g. websites recommended by the FDA such as DailyMed[3]), and (iv) potentially more difficult questions (e.g. drug alternatives, questions about not formalized/written medical knowledge such as usage time), and a higher sensitivity to error (e.g. drug interactions that cannot be confirmed online).

Several efforts focused on methods for medical question answering [7-10] as well as the creation of relevant datasets. For instance, Kilicoglu et al. [11] introduced a dataset of 2,614 medical questions annotated with the question focus, type and triggers of the question types. In that dataset, "Problems" are the most frequent named entities, while "treatment" and "information" are the most common question types. Ben Abacha et al. [12] organized a medical QA task at the TREC 2017 LiveQA track and published training and testing datasets including consumer health questions received by the NLM, annotations and reference answers retrieved by medical informatics experts. The coverage of questions about medications in the above collections, however, was insufficient to train and develop efficient systems to answer questions about

[1] https://github.com/abachaa/Medication_QA_MedInfo2019
[2] http://www.pewinternet.org/2013/01/15/health-online-2013
[3] https://dailymed.nlm.nih.gov/

medications. In this paper, we present a study of consumers' medication questions with the following contributions:

1. The development and publication of a manually annotated dataset of medication question-answer pairs based on real consumer questions submitted to MedlinePlus.
2. A synthesis of the manual annotation effort summarizing the insights obtained by a variety of experts from annotating and answering those questions.
3. New experiments using deep learning networks trained specifically for the tasks of identifying the main focus in the user's question, and the question type.

In what follows, we first describe our annotation methodology and the baseline approaches in the Methods section. We present the statistics and characteristics of the developed dataset and the empirical results in the Results section.

Methods

In this section, we describe the guidelines used in the manual annotation process and our first empirical evaluation methods based on the dataset.

Data Creation

Selecting Consumer Questions about Drugs. We selected anonymized consumer questions submitted to MedlinePlus[4]. We first performed Medical Entity Recognition using MetaMapLite [13]. We restricted the recognized entities to the following UMLS semantic types associated with medications: Antibiotic [antb], Clinical Drug [clnd], Neuroreactive Substance or Biogenic Amine [nsba], Pharmacologic Substance [phsu], Steroid [strd], and Vitamin [vita]. Finally, we manually selected the questions that (i) were deemed understandable and potentially answerable and (ii) have a drug name as focus. Figure 1 presents a word cloud of the most frequent terms in the selected consumer health questions.

Annotating the Questions. In a study conducted recently on consumer health question answering, Deardorff et al. [14] showed that for 62% of the questions, it was possible for librarians to find an answer in the top 5 search results in MedlinePlus using only the focus and question type. Given the importance of these two elements for QA, we, therefore, focused our efforts on manually annotating each question with a:

- Question focus (always a Drug name in this dataset),
- Question type (e.g. Dose, Interaction, Side effects).

Searching for Reference Answers. For each answerable question, annotators had to retrieve manually a correct and complete reference answer (with its URL and section title):

- **Correct** with regards to the question's explicit and implicit information (e.g. a question about a drug in the UK, a specific form or dose), and extracted from **reliable** websites or scientific papers.
- **Complete** with regards to all possible answers (e.g. all doses, ingredient lists from all manufacturers), and preferably written in **a consumer-friendly** language.

To select our answer sources, we followed the FDA recommendations suggesting[5] MedlinePlus-Drugs as a consumer-friendly website for consumer drug information and DailyMed for trustworthy information about FDA-approved and marketed drugs in the United States. Our final guideline was to search the following sources sequentially until an answer is retrieved:

1. MedlinePlus and DailyMed.
2. Other NIH or U.S. government websites.
3. Other trustworthy websites (e.g., the Mayo Clinic) or academic institutions' websites.
4. Other websites returned by a Google search.

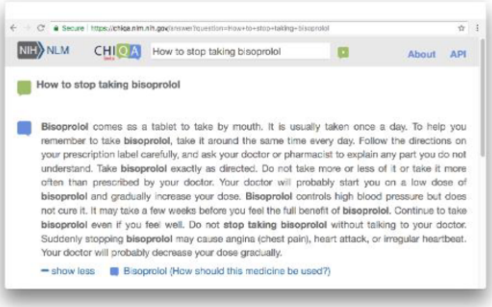
Four annotators participated in the manual annotation and answering process. Then, a medical doctor and an expert in question answering reconciled the annotations of the question types and validated the retrieved answers.

Baseline Methods

Focus Recognition. We adapted, extended and evaluated Bi-directional Long Short-Term Memory (Bi-LSTM) networks[6] on the task of recognizing the question focus (i.e., main drug name) according to the state-of-the-art architecture proposed by Xuezhe and Hovy [15]. The network includes a first Bi-LSTM network to build character-level embeddings, and a second Bi-LSTM taking as input both word embeddings built from the UMLS and pre-trained embeddings built with GloVe [16] and the character-level embeddings built during training with the first Bi-LSTM layer. The token labels were generated with a final Conditional Random Fields (CRF) layer. Our UMLS embeddings consist of binary vectors, where each word is tagged as the Beginning, Inside, Outside, End, or Single token (BIOES) of each semantic concept in the UMLS.

Question Type Identification. We implemented and evaluated a Convolutional Neural Network (CNN) on the task of identifying the question type (e.g., dosage, usage, contraindications). The input embeddings include UMLS BIOES embeddings and a randomly initialized vector of 128 dimensions updated with back-propagation during training.

Answer Retrieval. We conduct a first qualitative study on answer retrieval using twenty questions randomly selected from our dataset and the CHiQA question answering system[7]. CHiQA is the first online medical QA system for consumer health questions from reliable sources such as NIH websites (e.g., MedlinePlus, GARD, NCI), Mayo Clinic and DailyMed. The system relies on different machine learning and knowledge-based methods to recognize the question's focus and type [17-18] and uses the extracted information to retrieve answers with the Lucene search engine and a question-entailment recognition approach [19-20]. Figures 2 and 3 present the first answers returned by CHiQA to two questions selected randomly from the gold standard.



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

How to stop taking bisoprolol

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.

Bisoprolol (How should this medicine be used?)

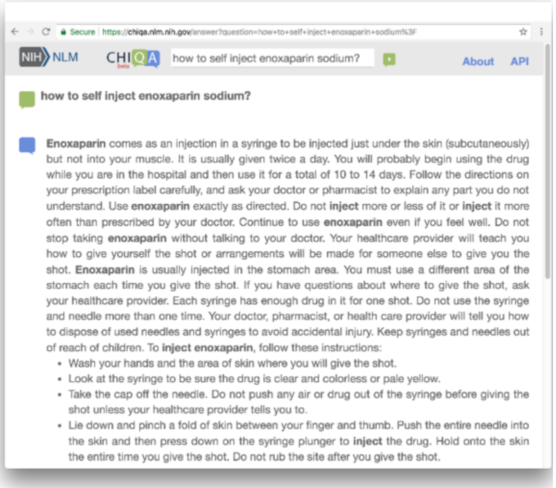
Figure 2– First Answer Returned by CHiQA to the Question: "How to Stop Taking Bisoprolol?"

[4] https://medlineplus.gov

[5] https://www.fda.gov/drugs/resourcesforyou/consumers

[6] https://github.com/guillaumegenthial/sequence_tagging

[7] https://chiqa.nlm.nih.gov


Secure | https://chiqa.nlm.nih.gov/answer?question=how+to+self+inject+enoxaparin+sodium%3F
NIH NLM
CHIQA beta
how to self inject enoxaparin sodium?
About
API

how to self inject enoxaparin sodium?

Enoxaparin comes as an injection in a syringe to be injected just under the skin (subcutaneously) but not into your muscle. It is usually given twice a day. You will probably begin using the drug while you are in the hospital and then use it for a total of 10 to 14 days. Follow the directions on your prescription label carefully, and ask your doctor or pharmacist to explain any part you do not understand. Use **enoxaparin** exactly as directed. Do not **inject** more or less of it or **inject** it more often than prescribed by your doctor. Continue to use **enoxaparin** even if you feel well. Do not stop taking **enoxaparin** without talking to your doctor. Your healthcare provider will teach you how to give yourself the shot or arrangements will be made for someone else to give you the shot. **Enoxaparin** is usually injected in the stomach area. You must use a different area of the stomach each time you give the shot. If you have questions about where to give the shot, ask your healthcare provider. Each syringe has enough drug in it for one shot. Do not use the syringe and needle more than one time. Your doctor, pharmacist, or health care provider will tell you how to dispose of used needles and syringes to avoid accidental injury. Keep syringes and needles out of reach of children. To **inject enoxaparin**, follow these instructions:

- Wash your hands and the area of skin where you will give the shot.
- Look at the syringe to be sure the drug is clear and colorless or pale yellow.
- Take the cap off the needle. Do not push any air or drug out of the syringe before giving the shot unless your healthcare provider tells you to.
- Lie down and pinch a fold of skin between your finger and thumb. Push the entire needle into the skin and then press down on the syringe plunger to **inject** the drug. Hold onto the skin the entire time you give the shot. Do not rub the site after you give the shot.

Figure 3– First Answer Returned by CHiQA to the Question: "How to Self Inject Enoxaparin Sodium?"

Results

Characteristics of the Dataset

The final gold standard contains **674** question-answer pairs with their associated annotations. These annotations include 25 question types, reported with examples in Table 1, and the answer sources, summarized in Figure 4. In particular, on the 674 answers, DailyMed was used to answer 290 questions, MedlinePlus for 128 questions, and other websites were used to answer 256 questions (e.g. cdc.gov, mayoclinic.org, health.harvard.edu, and PubMed abstracts and articles). Table 2 presents token-and-sentence-level statistics about the questions and the answers in the dataset.

Table 1 – Question Types in the Gold Standard

Question Type	#	Example
Information	112	what type of drug is amphetamine?
Dose	70	what is a daily amount of prednisolone eye drops to take?
Usage	61	how to self inject enoxaparin sodium?
Side Effects	60	does benazepril aggravate hepatitis?
Indication	55	why is pyridostigmine prescribed?
Interaction	51	can i drink cataflam when i drink medrol?
Action	39	how xarelto affects in the process of homeostasis?
Appearance	38	what color is 30mg prednisone?
Usage/time	36	when is the best time to take lotensin?
Stopping/tapering	31	how to come off citalopram?
Ingredient	28	what opioid is in the bupropion patch?
Action/time	23	how soon does losartan afffect blood pressure?
Storage and disposal	13	in how much temp bcg vaccine should store?
Comparison	11	why is losartin prescribed rather than a calcium channel blocker?
Contraindication	11	if i am allergic to sufa can i take glipizide?
Overdose	10	what happens if your child ate a tylenol tablet?
Alternatives	8	what medicine besides statins lower cholesterol?
Usage/duration	7	how long should i take dutasteride?
Time *(other time-related types)*	6	how long are you protected after taking the hep b vaccine?
Brand names	3	what is brand name of acetaminophen?
Combination	3	how to combine dapaliflozin with metformin?
Pronunciation	3	how do you pronounce humira?
Manufacturer	2	who makes this drug nitrofurantoin?
Availability	1	has lisinopril been taken off the market?
Long term consequences	1	what are the side effects and long term consequences of using nicotine?

Table 2 – Statistics about the Questions and Answers

Average number of tokens per question	**7.16**
Average number of tokens per answer	**69.06**
Average number of sentences per answer	**3.23**
Percentage of questions tokens present in answers	**34.72%**

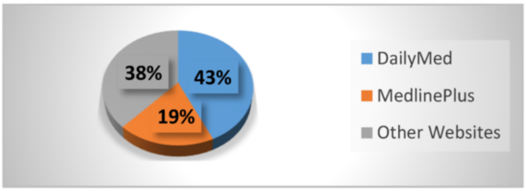


Figure 4 – Websites Used to Answer the Drug Questions

Evaluation Results

We report the average performance on 5 runs and the variation range for focus recognition and question type identification.

Focus Recognition. The Bi-LSTM-CRF network was implemented with Python 3.5 and Tensorflow 1.4. We trained the network with a random subset containing 80% of the data, using 10% for hyperparameter tuning and the remaining 10% for the final test. All data were represented in the BIOES token annotation format. The results on the test set are summarized in Table 3. We used a learning rate of 0.01, a dropout of 0.5, a batch size of 40, and the Adam optimizer [21] to minimize the CRF-based loss function.

Table 3 – Bi-LSTM-CRF Results for Focus Recognition

Results (%)	**F1**	**P**	**R**
Exact entity match	74.07 [+/-2.1]	78.12	70.42
Partial entity match	90.37 [+/- 3.4]	95.31	85.92

Question Type Identification. Our CNN network was also implemented using Python 3.5 and Tensorflow 1.4. Data were split according to same 80/10/10 percent subsets for training, development and test. To have enough training samples per class, we reduced the question types into the 14 most common types by aggregating the subtypes into their hypernym types: Information, Dose, Usage, Tapering, Interaction, Side effects, Indication, Action, Ingredient, Alternatives, Contraindication, Comparison, Manufacturing, and Appearance. We used a learning rate of 0.005, a dropout of 0.6, a batch size of 100, and the Adam optimizer to minimize the softmax-based loss function. The CNN network achieved an average accuracy of 75.7% on 5 runs with a variation in the [0, 2.5%] range. Despite using fixed random seeds for Tensorflow in both experiments, the relatively small size of the training data made them more susceptible to the non-deterministic implementation of GPU reduction operations and the Adam optimizer in TensorFlow[8].

Answer Retrieval. In our qualitative study of the answers returned to 20 random questions from the dataset, CHiQA found the correct answer in the top four results in 35% of the cases, only related answers for 35% of them and irrelevant answers for the remaining 30%. While this limited evaluation must be taken with caution, our independent observations from the annotation process also hint that classical QA systems may not be the best fit for medication questions.

We discuss these insights and potential ways to improve answering questions about medications in the following section.

Table 4– Example Questions and Answers from the Dataset

ID	Question (Q) / Answer (A)
1a	(Q) "*what does prednisone do to the body?"*
1b	(Q) "*what would a normal dose be for valacyclovir?"*
1c	(Q) "*how much gravol to kill you?"*
1d	(Q) "*what time should take memantine?"*
2a	(Q) "*what color is phenytoin?"* (A) [depends on the manufacturer] • PINK • WHITE (/Light Lavender) • ORANGE
2b	(Q) "*why would my urine test be negative for benzodiazepines when i take Ativan?"* (A) "Common limitations exist for screening benzodiazepines when using traditional immunoassay (IA) tests. IA testing for benzodiazepines often targets nordiazepam and oxazepam (…)" **AND** "Some commonly prescribed drugs have limited cross-reactivity (…)"
2c	(Q) "*is it alright touse fluticasone when using oxygen?"* (A) "Pharmacological therapy can influence morbidity and mortality in severe chronic obstructive pulmonary disease (COPD). Long-term domiciliary oxygen therapy (LTOT) improves survival in COPD with chronic hypoxaemia. Oral steroid medication has been associated with improved survival in men and increased mortality in women, while inhaled steroid medication has been associated with a reduction in the exacerbation rate (…)."
2d	(Q) "*how long does vicodin stay in breast milk?"* (A) "Following a 10 mg oral dose of hydrocodone administered to five adult male subjects, the mean peak concentration was 23.6 ± 5.2 ng/mL. Maximum serum levels were achieved at 1.3 ± 0.3 hours and the half-life was determined to be 3.8 ± 0.3 hours."

Discussion

Difficulty and Ambiguity of Medication Questions

In our annotation and answering efforts, questions about medications showed traits common to consumer health questions such as **linguistic ambiguity** due to misspellings, or wrong grammar leading to unclear meaning or multiple interpretations. For instance, in Table 4, example 1a, it is unclear whether the question is about the side effects or the action of "prednisone". In addition, a distinct pattern emerged with many questions lacking essential information or context, making them too **underspecified** [22] to answer. For instance, in example 1b, finding a relevant answer about the right dose of "Valacyclovir" requires additional information about the patient and the condition or purpose of administration. **Circumlocution** [23] was another phenomenon that we encountered, i.e., the use of many words or general terms to express a medical term. For instance, in example 1c, the medical term "toxic dose" or "overdose" is replaced with the expression "how much X to kill you?". In several cases, we also encountered a **knowledge barrier** when the requested medical information/knowledge was not formalized or written online (e.g. Table 4, 1d).

Complexity of Manual Answer Retrieval

Our annotators reported **conditional answers** frequently. These include answers that depend on the manufacturer, on the disease, or on patient information. For instance, in Table 4, 2a, four different answers are possible according to the manufacturer of phenytoin. **Distributed answers** were also encountered in several cases. These are answers that can be formed only by combining different text snippets from different answers and/or sources (e.g., Table 4, 2b). Many answers also required an understanding of **expert terminology**. These include answers that need to be translated to consumer-friendly language and questions that required rephrasing to expert language in order to find relevant answers (e.g., Table 4, 2c). Other answers could not be found without **expert inference** based on background knowledge (e.g., Table 4, 2d). External resources like *eHealthMe* were needed for specific types of questions such as Interaction questions. Answering some of the questions was also time consuming even for medical experts. For example, an hour was necessary to answer the question "is it alright to use fluticasone when using oxygen?" from PubMed.

Challenges of Automating Medication QA

Prior to our study, the lack of gold standard datasets for medication QA was the major bottleneck in automatic QA for drug questions. This new dataset opens new opportunities for both qualitative studies and quantitative evaluation of QA systems. In addition, systems relying on big training data can use it both (i) as a development set for fine-tuning the hyperparameters and testing different architectures and (ii) as a test benchmark.

In question understanding, our baseline networks achieved an encouraging performance despite the limited training data, with (i) 74% F1 score in question focus recognition for exact span matching and 90% for partial span matching, and (ii) 75.7% accuracy in identifying the question type. Several improvements can be considered for future developments, such as a richer set of embeddings, or a relevant language model. A more fine-grained adaptation of the UMLS can also be applied by restricting to the list of relevant semantic types either through pre-filtering or through trainable masks.

In answer retrieval, insights from both our annotation process and our CHiQA-based evaluation provide additional guidance on the relevant solutions and ways of improvement for medication question answering. In particular:

- Medical text translation [24] and simplification [25] are often needed to find relevant answers and make the retrieved answers readable for non-expert users.
- More data and resources are needed to cover information about drug interactions and usage guidelines. Such information can be extracted from both the scientific literature and clinical sources [26].
- Conditional answers require different solutions such as providing a list of answers or interacting with the user in a dialogue-based approach.

[8] https://www.twosigma.com/insights/article/a-workaround-for-non-determinism-in-tensorflow

- Due to the frequency of these conditional answers, fully unsupervised approaches are less likely to succeed than approaches based on (advanced) inference methods that can rely on explicit context semantics.

Conclusions

We studied consumers' questions about medications. We created a new gold standard corpus for question answering about drugs that we shared in the scope of this paper[9]. We presented statistics, insights and conclusions based on the manual annotation process, deep learning experiments, and preliminary evaluation of automatic answer retrieval. We hope that this new benchmark and initial experiments will foster new approaches and additional community efforts in addressing the growing need for reliable information about medications online.

Acknowledgements

This work was supported by the intramural research program at the U.S. National Library of Medicine, NIH.

References

[1] R.W. White and E. Horvitz, Experiences with web search on medical concerns and self diagnosis, *AMIA Annu Symp Proc* **2009** (2009), 696-700.

[2] A.B. Abacha and D. Demner-Fushman, On the Role of Question Summarization and Information Source Restriction in Consumer Health Question Answering, in: *AMIA Informatics Summit*, 2019.

[3] P. Jacquemart and P. Zweigenbaum, Towards a medical question-answering system: a feasibility study, *Stud Health Technol Inform* **95** (2003), 463-468.

[4] D. Moll and J.L. Vicedo, Question Answering in Restricted Domains: An Overview, *Comput. Linguist.* **33** (2007), 41-61.

[5] S.J. Athenikos and H. Han, Biomedical question answering: a survey, *Comput Methods Programs Biomed* **99** (2010), 1-24.

[6] A. Ben Abacha and P. Zweigenbaum, MEANS: A medical question-answering system combining NLP techniques and semantic Web technologies, *Information Processing & Management* **51** (2015), 570-594.

[7] Y. Niu and G. Hirst, Analysis of semantic classes in medical text for question answering, in: *Workshop on Question Answeing in Restricted Domains, ACL*, 2004.

[8] J. Jeon, W.B. Croft, and J.H. Lee, Finding similar questions in large question and answer archives, in: *Proceedings of the 14th ACM international conference on Information and knowledge management*, ACM, Bremen, Germany, 2005, pp. 84-90.

[9] H. Yu, C. Sable, and H.R. Zhu, Classifying Medical Questions based on an Evidence, in: *Workshop on Question Answering in Restricted Domains, AAAI*, 2005.

[10] L.C. Llanos, S. Rosset, and P. Zweigenbaum, Automatic classification of doctor-patient questions for a virtual patient record query task, in: *BioNLP*, 2017.

[11] H. Kilicoglu, A. Ben Abacha, Y. Mrabet, S.E. Shooshan, L. Rodriguez, K. Masterton, and D. Demner-Fushman, Semantic annotation of consumer health questions, *BMC Bioinformatics* **19** (2018), 34.

[12] E.A. Asma Ben Abacha, Yuval Pinter, Dina Demner-Fushman, Overview of the Medical Question Answering Task at TREC 2017 LiveQA, in: *TREC*, 2017.

[13] D. Demner-Fushman, W.J. Rogers, and A.R. Aronson, MetaMap Lite: an evaluation of a new Java implementation of MetaMap, *J Am Med Inform Assoc* **24** (2017), 841-844.

[14] A. Deardorff, K. Masterton, K. Roberts, H. Kilicoglu, and D. Demner-Fushman, 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* **68** (2017), 1724-1736.

[15] X. Ma and E. Hovy, End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF, in: *ACl*, 2016.

[16] J. Pennington, Richard Socher, and Christopher Manning, Glove: Global vectors for word representation, in: *EMNLP*, 2014.

[17] Y. Mrabet, H. Kilicoglu, K. Roberts, and D. Demner-Fushman, Combining Open-domain and Biomedical Knowledge for Topic Recognition in Consumer Health Questions, *AMIA Annu Symp Proc* **2016** (2016), 914-923.

[18] K. Roberts, H. Kilicoglu, M. Fiszman, and D. Demner-Fushman, Automatically classifying question types for consumer health questions, *AMIA Annu Symp Proc* **2014** (2014), 1018-1027.

[19] A. Ben Abacha and D. Demner-Fushman, Recognizing Question Entailment for Medical Question Answering, *AMIA Annu Symp Proc* **2016** (2016), 310-318.

[20] A.B. Abacha and D. Demner-Fushman, A Question-Entailment Approach to Question Answering, *arXiv:1901.08079 [cs.CL]* (2019).

[21] D.P. Kingma and J.B. Adam, A Method for Stochastic Optimization, in: *ICLR*, 2015.

[22] D. Schlangen, A. Lascarides, and A. Copestake, Resolving underspecification using discourse information, in: *Perspectives on Dialogue in the New Millennium*, 2003.

[23] I. Stanton, S. Ieong, and N. Mishra, Circumlocution in diagnostic medical queries, in: *Proceedings of the 37th international ACM SIGIR conference on Research &#38; development in information retrieval*, ACM, Gold Coast, Queensland, Australia, 2014, pp. 133-142.

[24] Q. Zeng-Treitler, S. Goryachev, H. Kim, A. Keselman, and D. Rosendale, Making texts in electronic health records comprehensible to consumers: a prototype translator, *AMIA Annu Symp Proc* (2007), 846-850.

[25] E. Abrahamsson, T. Forni, M. Skeppstedt, and M. Kvist, Medical text simplification using synonym replacement : adapting assessment of word difficulty to a compounding language, in: *14th Conference of the European Chapter of the Association for Computational Linguistics,April 27, 2014 Gothenburg, Sweden,* W. Sandra, ed., Association for Computational Linguistics, Stroudsburg, 2014, pp. 57-65.

[26] S. Vilar, C. Friedman, and G. Hripcsak, Detection of drug-drug interactions through data mining studies using clinical sources, scientific literature and social media, *Brief Bioinform* **19** (2018), 863-877.

Address for Correspondence

Asma Ben Abacha, PhD, asma.benabacha@nih.gov
Research Scientist
U.S. National Library of Medicine
8600 Rockville Pike, Bethesda, MD 20894, USA.

[9] https://github.com/abachaa/Medication_QA_MedInfo2019