Annotating Question Types for Consumer Health Questions

Kirk Roberts, Kate Masterton, Marcelo Fiszman, Halil Kilicoglu, and Dina Demner-Fushman

National Library of Medicine
National Institutes of Health
Bethesda MD, USA
`kirk.roberts@nih.gov; ddemner@mail.nih.gov`

Abstract

This paper presents a question classification scheme and a corresponding annotated corpus of consumer health questions. While most medical question classification methods have targeted medical professionals, the 13 question types we present are targeted toward disease questions posed by consumers. The corpus consists of 1,467 consumer-generated requests for disease information, containing a total of 2,937 questions. The goal of question type classification is to indicate the best strategy for selecting answers from publicly available health information resources, such as medical encyclopedias and medical research websites. The annotated question corpus, along with detailed annotation guidelines, are publicly available.

Keywords: question classification, question answering, medical language processing

1. Introduction

Most research in automatic question answering has focused on well-formed factoid questions in the style of the TREC Question Answering evaluations (Prager, 2006). However, real-life questions that automatic systems must handle are typically not well-formed or explicit in their information needs. This is particularly the case for consumer health questions. Zhang (2010) focuses on consumer health questions submitted to Yahoo Answers, finding these questions were largely concerned with diseases and symptoms, contained many abbreviations and misspellings, and often contained more than one actual question.

Furthermore, the type of answer to be provided to a consumer is qualitatively different than an appropriate answer for a medical professional. Many questions that the general public might have about diseases are answerable by reliable consumer health information resources, such as MedlinePlus[1] or Genetics Home Reference[2]. Even these consumer-oriented resources might be difficult to navigate for users with little medical knowledge. Thus, a question answering system that could point a user to the most relevant encyclopedic articles or sections could prove useful. Some information providers, such as the NIH Genetic and Rare Diseases Information Center[3] (GARD) not only answer consumers' questions, but also make the questions and the answers publicly available.

In our own consumer question answering system, we have found a wide variety in the styles and types of questions even in a relatively focused area of questions about specific diseases, especially their treatments and outlook. Our rule-based classification of these questions into traditional treatment, cause, prognosis, diagnosis, manifestation, and general information types is sufficient for understanding and structuring simple questions, but it is insufficient for most of the questions submitted to our system. It is also clear that it is practically impossible to develop rules for understanding the many types of complex questions, and thus a corpus-based machine learning solution is applicable. To our knowledge, there are few publicly available collections of consumer health questions, and none annotated at the granularity needed for our task. GARD questions are a good intermediate step from the simple questions, such as "*is there any help availabe,for fibromalgia*" that our system translates to "*What are treatments for fibromyalgia?*" to complex questions, such as the one in Figure 1.

My wife has been having shortness of breath in the mornings (mornings only). She has, what I think, excessive heart rate as well. Could this be anxiety attacks? I don't think it would be a heart issue. She certainly isn't overweight. SHe is 5'2"; and 100-105 pounds. if she is having anxiety attack... what is the best course of action? She seems to feel better when she lies down and rests.....while either watching TV...or sleeps. SHe has no history in her family for heart disease either. WHat are your thoughts?

Figure 1: Consumer health question.

GARD questions, for example in Figure 2, are well-formed with few misspellings and are primarily focused on a specific condition. Questions in the GARD corpus are obtained from the public and answered by experienced information specialists. We have annotated these questions with more detailed question types as the next step towards understanding a wide range of consumer health questions.

Do children with Cat Eye Syndrome generally experience a decline in physical abilities as they reach adulthood? Is there any shortening of the lifespan associated with this condition?

Figure 2: Consumer health question from GARD.

For clarity, the remainder of this paper refers to the multi-sentence, possibly multi-question spans found in Figures 1 and 2 as *requests*. A *question* refers to a clause asking for a single piece of information. Section 2 of this paper describes previous work in question classification; Section 3 briefly outlines our proposed question types; Section 4 describes the process of annotating the GARD corpus; Section 5 provides numerous examples of each question type; and Section 6 describes our corpus.

---

[1] http://www.nlm.nih.gov/medlineplus/
[2] http://ghr.nlm.nih.gov/
[3] https://rarediseases.info.nih.gov/GARD/

2. Background

In much of the literature, question classification is synonymous with answer type detection (Hermjakob, 2001). An answer type corresponds to the answer's expected class, typically a named entity type for TREC QA style factoid questions. For instance "*Who is Tom Cruise married to?*" has a PERSON answer type, while "*What was the capital of the Holy Roman Empire?*" has a CITY answer type. Li and Roth (2002) provide one of the first statistical methods for classifying answer types, along with a corpus on which many other methods have been judged. Their answer types form a 2-level hierarchy with six "coarse" types and fifty "fine" types. Roberts and Hickl (2008) demonstrate how such an answer type hierarchy can be arbitrarily extended as the need arises. The automatic classification of answer types has been shown to be a highly semantic task (Metzler and Croft, 2005) due to the short length and predictable syntactic structures of TREC-style questions. The single most important semantic feature involves finding and typing the answer type term (*capital* in the above question) to the answer type (Krishnan et al., 2005; Hickl et al., 2007).

Despite the focus on answer type detection, there are other important forms of question classification, especially when question answering systems are designed to answer nonfactoid questions or questions that require domain knowledge. For instance, the question "*List the cities of the Holy Roman Empire.*" still has a CITY answer type, but also requires classifying the question's function (Bu et al., 2010). Additionally, it was useful to identify a question's topic or domain (Duan et al., 2008; Roberts and Harabagiu, 2012; Chan et al., 2013), especially when further domain knowledge is necessary to answer the question. In addition to the question classification tasks listed above, innumerable other tasks exist for the many classes of questions found in natural language (Chali and Hasan, 2012; Banerjee and Bandyopadhyay, 2012).

In contrast to the forms of question classification in TREC QA style questions, medical questions have presented very different classification methods. Ely and colleagues (1999; 2000; 2005) collected several sets of medical questions posed by physicians. These questions are available as part of the NLM Clinical Questions Collection[4] (CQ). On these well-studied question sets, automatic approaches have been proposed to classify question answerability (Yu and Sable, 2005), taxonomy (Kobayashi and Shyu, 2006), and topic (Yu and Cao, 2008). The diversity of possible ways medical questions may be classified speaks to their potential complexity. Because of this complexity, medical question answering does not appear to be a promising field for purely factoid question answering approaches. For this reason, many medical question answering systems utilize the PICO structure (Demner-Fushman and Lin, 2007; Schardt et al., 2007), which encodes a far wider range of information-seeking queries.

The complexity of medical questions is further compounded when answering questions posed by consumers, as shown in Figures 1 and 2. While answer types are certainly important in returning appropriate medical information to

---

[4]http://clinques.nlm.nih.gov/

the consumer, we have found it is more appropriate to determine the general type of information the user is looking for (treatment, prognosis, symptoms, etc.). By classifying questions along these lines, then, our question answering system may choose the most appropriate passage selection strategy.

In this paper we present a set of question types and a corresponding annotated corpus designed for selecting an appropriate strategy for consumer health questions. While previous work has demonstrated the value of deep question taxonomies, we have focused only on an appropriate first-level set of types for now. The specific impact of dealing with consumer health questions (as opposed to questions posed by medical professionals) is two-fold. First, certain ambiguities often present in consumer questions result in ontological choices about how best to organize the first level of our type scheme. Second, the distribution of question types can differ greatly between professional and consumer questions. While medical question answering systems for professionals often focus on the PICO structure, consumer requests often contain more questions for general information, causes, and prognosis (Liu et al., 2011; Kilicoglu et al., 2013). Our types thus focus on classifying the most common classes of consumer questions, and ignore the PICO structure. While the corpus we present focuses on genetic and rare diseases, we believe the following question types generalize well to other disease questions posed by consumers.

3. Question Types

We propose the following question types. Examples and annotation rules are provided in Section 5. As an example of how the question types correspond to encyclopedic sections, for each question type we list the sections of MedlinePlus where answers are most likely to be found (though our system uses several other sources as well).

(1) ANATOMY: Identifies questions asking about a particular part of the body, such as the location affected by a disease. Answers to ANATOMY questions are typically found in the "Anatomy/Physiology" section.

(2) CAUSE: Identifies questions asking about the cause of a disease. This includes both direct and indirect causes, such as factors that might increase the risk of developing a disease. Answers to CAUSE questions are typically found in the "Common Causes" section.

(3) COMPLICATION: Identifies questions asking about the problems a particular disease causes. This primarily focuses on the risks faced by patients with the disease and does not include the signs/symptoms of a disease. Answers to COMPLICATION questions are typically found in the "Related Issues" section.

(4) DIAGNOSIS: Identifies questions asking for help making a diagnosis. Our question answering system is not designed to provide a direct diagnosis. However, the DIAGNOSIS type does include questions asking about diagnostic tests, or methods for determining the difference between possible diagnoses (differential diagnosis). Answers to DIAGNOSIS questions are typically found in the "Diagnosis" and "Testing" sections.

(5) INFORMATION: Identifies questions asking for general information about a disease. This includes type information about diseases (e.g., whether two disease names represent the same disease, or if one disease is a type of another disease). Answers to INFORMATION questions are typically found in the "Definition" and "Description" sections.

(6) MANAGEMENT: Identifies questions asking about the management, treatment, cure, or prevention of a disease. Answers to MANAGEMENT questions are typically found in the "Treatment" and "Prevention" sections.

(7) MANIFESTATION: Identifies questions asking about signs or symptoms of a disease. Answers to MANIFESTATION questions are typically found in the "Symptoms" section.

(8) OTHEREFFECT: Identifies questions asking about the effects of a disease, excluding signs/symptoms (MANIFESTATION) or predispositions (COMPLICATION). When the question requires medical knowledge to understand a given effect is actually a MANIFESTATION or COMPLICATION, it is instead classified as an OTHEREFFECT, the reasoning for which is provided in Section 4. Answers to OTHEREFFECT questions are typically found in the "Symptoms" and "Related Issues" sections.

(9) PERSONORG: Identifies any question asking for a person or organization involved with a disease. This can include medical specialists, hospitals, research teams, or support groups for a particular disease. Answers to PERSONORG questions are typically not found in MedlinePlus articles, but may be found in links on MedlinePlus pages (especially for support groups).

(10) PROGNOSIS: Identifies questions asking about life expectancy, quality of life, or the probability of success of a given treatment. Answers to PROGNOSIS questions are typically found in the "Expectations" section.

(11) SUSCEPTIBILITY: Identifies questions asking how a disease is spread or distributed in a population. This includes inheritance patterns for genetic diseases and transmission patterns for infectious diseases. Answers to SUSCEPTIBILITY questions are typically found in the "Mode of Inheritance" and "Prevalence" sections.

(12) OTHER: Identifies disease questions that do not belong to the above types. This includes non-medical questions about a disease, such as requests for financial assistance or the history of a disease. Answers to OTHER questions are typically found in the "Definition" and "Description" sections, though by definition these answers may be found anywhere in the encyclopedic entry.

(13) NOTDISEASE: Identifies questions that aren't about a disease and thus not handled by our question answering system.

Additionally, we propose one question attribute that could apply to any question of the above types:

(14) RESEARCH: Identifies questions asking for the latest research, such as medical publications or clinical trials. Answers to RESEARCH questions are typically found using specialized search engines such as `PubMed`[5] for medical publications or `ClinicalTrials.gov`[6] for clinical trials.

The choice of these question types started with an initial set (Van Der Volgen et al., 2013), but was altered over the course of the annotation process described in the next section. The final set of types is thus both based on *a priori* medical knowledge and specific issues that arise in consumer health questions.

To a large extent, the types are generalizable beyond consumer health questions, though the choice of how abstract or how specific the types should be was based on their distribution in consumer health questions. For example, the types overlap quite a bit with those in the CQ set. In some cases, the CQ types are more broad than our types, and in other cases they are more specific.

4. Annotation Process

As a source for consumer health questions about diseases, we used 1,467 publicly available requests on the GARD website. Because these requests contain multiple question sentences, and many of the question sentences contain requests for different types of information, the questions were annotated for syntactic decomposition as described in Roberts et al. (2014). Figure 3 illustrates an example of question decomposition. This decomposition process results in 2,937 questions from the 1,467 GARD requests.

| **Request**: |
| I have been recently diagnosed with antisynthetase syndrome. Could you please provide me with information on antisynthetase syndrome? I am also interested in learning about prognosis, treatment, and clinical trials. |
| **Decomposed Questions**: |
| Q1) Could you please provide me with information on antisynthetase syndrome? |
| Q2) I am also interested in learning about prognosis. |
| Q3) I am also interested in learning about treatment. |
| Q4) I am also interested in learning about clinical trials. |

Figure 3: Question Decomposition Example.

Additionally, these requests are annotated with a FOCUS disease, typically one per request but occasionally more than one. The FOCUS is the disease the consumer is interested in learning more about. In Figure 2, the FOCUS is *Cat Eye Syndrome*. In Figure 3, the FOCUS is *antisynthetase syndrome*. The FOCUS is not only important as a default value for resolving co-reference, but it is also useful for determining the question type. Since our question answering approach primarily utilizes medical encyclopedias, the FOCUS is often the encyclopedic entry, while the question type is the section of the entry in which the answer is found. For cause-and-effect questions, for example, if the FOCUS is the cause, the answer is likely to be found in the effect section, and vice versa.

---

[5]http://www.ncbi.nlm.nih.gov/pubmed
[6]http://clinicaltrials.gov/

A pool of five annotators (all the authors) were used: two MDs (MF, DDF), three computer science PhDs (KR, HK, DDF), and one medical librarian (KM). Annotation was performed with a custom annotation tool. On average, annotators were able to manually classify approximately 10 questions per minute.

The initial set of question types was based on the study of consumer health questions by (Van Der Volgen et al., 2013). The eleven question types in that study were: ANATOMY, CAUSE, COMPLICATION, DIAGNOSIS, INFORMATION, LOCATION, MANIFESTATION, PREVENTION, PROGNOSIS, SUSCEPTIBILITY, and TREATMENT. Using these types as well as the OTHER type as a starting point, three annotators double-annotated the questions from 250 GARD requests. The average inter-annotator agreement (*Kappa*) for this round was 0.802, which is reasonably good agreement. To resolve annotation conflicts, all five annotators met in person to discuss major sources of disagreements. During the resolution process, it was decided that three changes would be made. First, the RESEARCH attribute was added as it was felt this was an important distinction that indicated consumer-oriented medical encyclopedias were not sufficiently detailed to answer these questions. Second, the LOCATION type was re-named PERSON-ORG to emphasize the goal of looking for medical professionals or organizations. Previously, requests for websites had been considered LOCATIONS, but a website can contain any type of information, and would thus only be considered an answer type as opposed to a search strategy. Finally, an ASSOCIATION type was added when a question asked about the relationship between multiple diseases.

In the second round, four annotators (two of whom participated in the previous round) double-annotated questions from 450 GARD requests. The inter-annotator agreement for this round was a slightly lower 0.742. This drop can likely be explained by the two additional annotators, but is more likely the result of adding an additional question type as both new annotators had participated in the first round conflict resolution session. During the second resolution process, it was decided to make two additional changes. First, the NOTDISEASE type was split off from OTHER to indicate questions not relevant to a disease question answering system. Second, and more importantly, the OTHEREFFECT type was created. This type subsumed most of the ASSOCIATION questions, where the association is an effect of the FOCUS. All other ASSOCIATION questions involve one disease being the same as, different from, or a sub-type of another disease, and were re-classified as INFORMATION. More notably, though, OTHEREFFECT sits in the ambiguous zone between MANIFESTATION and COMPLICATION. It is difficult for an automatic system to encode sufficient medical knowledge to accurately differentiate between these two types. Indeed, this was a major source of annotator disagreement, as both are considered effects of a disease. Instead of merging both types into a single effect type, MANIFESTATION and COMPLICATION remain but require explicit lexical evidence to classify them (such as reference to *symptoms* or *risks*). When medical knowledge is needed, questions are considered OTHEREFFECT. The strategy for answering OTHEREFFECT ques-

Question Type	Accuracy	κ
ANATOMY	99.6	0.314
CAUSE	98.2	0.784
COMPLICATION	97.4	0.506
DIAGNOSIS	97.8	0.843
INFORMATION	94.3	0.819
MANAGEMENT	98.4	0.925
MANIFESTATION	96.5	0.714
OTHEREFFECT	96.0	0.570
PERSONORG	98.4	0.809
PROGNOSIS	96.2	0.795
SUSCEPTIBILITY	95.5	0.804
OTHER	96.2	0.327
NOTDISEASE	99.6	0.233
Overall	81.0	0.787

Table 1: Inter-annotator agreement broken down by question type. Accuracy = % agreement between annotators, κ = Cohen's Kappa.

tions is then simply the union of the MANIFESTATION and COMPLICATION strategies.

In the final round, all five annotators double-annotated the remaining 767 GARD requests. The inter-annotator agreement for this round was 0.791, an improvement from the previous round. No further changes were made to the annotation standard. Inter-annotator agreement for all rounds is broken down by question type in Table 1. In general, the more frequent types (see Table 2) have higher agreement scores, an exception being OTHEREFFECT questions.

5. Examples

To illustrate the question types presented in Section 3, we present the following examples from the annotated GARD data. For some types, informal sub-classes are described. At this point, however, only the 13 types are annotated.

ANATOMY questions are concerned with the body location of a disease, including where symptoms manifest (example 1) and if specific anatomical locations are possible body sites for a disease (2 & 3):

(1) *Does IP affect any areas inside the body, such as internal organs?*

(2) *I was researching metachondromatosis - is this something that when felt is attached to the bone itself?*

(3) *Can this cancer occur on someone's back?*

CAUSE questions are concerned with the cause of the FOCUS disease, including direct causes (1-6) as well as factors that might increase the susceptibility of a disease (7-10):

(1) *What causes this condition?*

(2) *Can pregnancy trigger such an issue?*

(3) *Could this be caused by hip dysplasia?*

(4) *Can in vitro fertilization cause Duane syndrome?*

(5) *Can you provide me with more information about this condition, including its causes?*

(6) *What are the genes involved?*

(7) *Are there any exogenous (outside) factors that cause Batten disease?*

(8) *Can a father taking an antidepressant contribute to the reason a baby has a trisomy?*

(9) *What increases the risk of males to have nondisjunction during meiosis?*

(10) *Has any research been done to check whether there is any viral involvement?*

COMPLICATION questions are concerned with longer term effects of a disease, usually in the form of risks (1-6), statements of the FOCUS causing some other longer-term disease (7 & 8), or direct queries of complications (9 & 10):

(1) *Are carriers of cystic fibrosis at a higher risk for other health conditions?*

(2) *Is there an increased risk of getting malignant disorders?*

(3) *Does propionic acidemia carry risk factors for hearing loss, especially progressive hearing loss?*

(4) *As an individual with G6PD deficiency, would living in a malaria risk country pose a greater risk to me compared with an individual who does not have this deficiency?*

(5) *Can you tell me if the risk for pre-eclampsia in pregnancy is increased significantly for a woman over the age of 35?*

(6) *Would my diagnosis of nail patella syndrome increase my risk for a miscarriage?*

(7) *Can Charcot-Marie-Tooth disease cause a paralyzed laryngeal nerve?*

(8) *Can this lead to a pre-cancerous condition?*

(9) *I would like to ask about the serious complications associated with this disease.*

(10) *What complications may occur in the future?*

DIAGNOSIS questions are concerned with the method of diagnosing a particular disease. These questions can simply ask about the diagnostic process for a given disease (1-6), specific types of diagnostic tests (7-9) and even questions asking for a specific diagnosis[7] (10):

(1) *Is it true that Smith-Magenis syndrome become more difficult to diagnose through genetic testing if the child is older than 24 months?*

(2) *Can Polycystic Kidney Diseases be diagnosed in pregnancy?*

(3) *Can Bloom syndrome be detected before symptoms appear?*

(4) *Can you provide me with information regarding diagnosis?*

(5) *How is it diagnosed?*

(6) *I would like to know what test needs to be performed in order to be able to rule it out.*

(7) *Could you please let me know the name of the blood test used to diagnose a vitiligo carrier?*

(8) *Is genetic testing available for Asperger syndrome?*

(9) *How can I learn more about genetic testing for this syndrome?*

(10) *How can we know if it is hereditary amyloidosis or not?*

---

[7]Note that our QA system cannot provide a diagnosis, so for these types of questions we simply provide a disclaimer along with general diagnostic information.

INFORMATION questions are concerned with basic information about a disease. This includes specific requests for general information (1-5) and sub-type/super-type information (6 & 7), which commonly concerns cancer (8-10):

(1) *What is this condition?*

(2) *Can you please provide me with general information about hyper IgD syndrome?*

(3) *I have osteopoikilosis and would like more information on this condition.*

(4) *How can I find information specific to adults with this condition?*

(5) *I would like general information on mucopolysaccharidosis I (MPS I).*

(6) *Is Machado Joseph disease a form of Parkinson's disease?*

(7) *What are the different types of syringomyelia?*

(8) *Is TMT a cancerous classification?*

(9) *Is someone diagnosed with teratoma with malignant transformation (TMT) considered to have cancer?*

(10) *Is astroblastoma a brain cancer?*

MANAGEMENT questions are concerned with the management, treatment, and prevention of a disease. These questions can be about general disease management (1), specific types of management (2 & 3), treatment (4-7), cure (8 & 9), and prevention information (10):

(1) *What can we do to help him?*

(2) *Is there a diet guide that can help me plan a diet that is low in protein?*

(3) *She had scalded skin syndrome when she was a baby: will she need pain relief?*

(4) *Are there new therapies for treatment of pili torti?*

(5) *What treatments are available for this condition?*

(6) *Does this mean that if I were to get cancer in the future I would not be able to be treated with chemotherapy?*

(7) *How long does it usually take for this medication to take effect?*

(8) *Is there a cure for HTLV-1 in general?*

(9) *Can the condition be cured?*

(10) *Is there anyway to prevent LHON with certain vitamins?*

MANIFESTATION questions are concerned with the signs and symptoms of a disease, including asking what the symptoms of a disease are (1-6), whether specific issues are symptoms of the disease (7-9), and the typical magnitude of the symptoms (10):

(1) *What are symptoms of ADPKD?*

(2) *I'd like to learn more about this syndrome including its symptoms.*

(3) *Are there any physical characteristics associated with the disorder?*

(4) *What symptoms may people experience as their disease progresses?*

(5) *Can women have symptoms of glucose 6 phosphate dehydrogenase (G6PD) deficiency?*

(6) *Are there any things I should look out for?*

(7) *Is fatigue a common symptom of polyglucosan body disease?*

(8) *Are severe headaches symptoms of this condition?*

(9) *Could this indicate that I am developing scleroderma inside my organs?*

(10) *In autosomal dominant polycystic kidney disease (ADPKD), does the severity of symptoms of tend to be the same among affected family members?*

OTHEREFFECT questions are concerned with the effects of a disease that are not explicitly COMPLICATIONS or MANIFESTATIONS. This can include questions about a causal link between the FOCUS and another condition (1-5) or a general link (6 & 7), questions about specific effects of a disease (8 & 9), and unspecified effects that may alter normal life processes (10):

(1) *Does Klinefelter syndrome cause weight gain?*

(2) *Does excess growth hormone have an effect on the uterus?*

(3) *Is it normal for her to be hungry all the time?*

(4) *Can hyperphenylalaninemia caused by tetrahydrobiopterin (BH4) deficiency cause problems with eating solids in babies?*

(5) *Can tuberous sclerosis affect eye closure?*

(6) *Could this be connected to the septo-optic dysplasia?*

(7) *I was wondering if people with Pfeiffer syndrome have circulatory problems in their legs.*

(8) *Is hearing loss in people with Waardenburg syndrome type II likely to be progressive?*

(9) *How common is it for only one leg to have bowing?*

(10) *Is it safe for them to bear children?*

PERSONORG questions are concerned with finding individuals or organizations that specialize in a particular disease. This typically includes individual specialists (1-5), researchers (6 & 7), and support groups (8-10).

(1) *How can I find a physician who is knowledgeable about this condition?*

(2) *Are there doctors who specialize in this condition?*

(3) *I would like information about which doctors could treat me.*

(4) *I have many questions, who can I talk to?*

(5) *Are there any doctors who specialize in this type of neurological dysfunction?*

(6) *Are there any new research studies enrolling people with carbamoyl phosphate synthase 1 deficiency?*

(7) *How can I be seen by researchers at the National Institutes of Health?*

(8) *Are there any support groups for adults with blepharophimosis type 1?*

(9) *How can I find other families that are in a similar situation?*

(10) *I have prolidase deficiency and would like contact other sufferers.*

PROGNOSIS questions are concerned with life expectancy and long-term quality of life. Specific types of questions include general prognosis (1 & 2), life expectancy (3), survival probability (4 & 5), lifestyle changes (6), long-term prospects of symptoms (7 & 8), and the long-term probability of symptoms emerging or returning (9 & 10):

(1) *What can I expect from this condition?*

(2) *What is the prognosis?*

(3) *I would like to know what the life expectancy is for people with this syndrome.*

(4) *What are her chances of survival if it is a cancerous tumor?*

(5) *Is Devic disease fatal?*

(6) *Can you do any other types of exercises if you have Wolff-Parkinson-White syndrome?*

(7) *Is there any hope for a return of smell for someone who has never smelled?*

(8) *Am I going to have this for the rest of my life?*

(9) *Is there any data that would suggest that kidney problems could show up later?*

(10) *As Burkitt is so rare, there is a lack of research whether the treatment is curative or palliative and I want to know what the chances are that it will return.*

SUSCEPTIBILITY questions are concerned with the spread and prevalence of a disease. Common types of questions include whether a disease is genetic (1 & 2), its inheritance patterns (3 & 4), whether it is considered rare (5 & 6), and its prevalence in certain populations (7-10):

(1) *Is this is a genetic disease?*

(2) *Is idiopathic thrombocytopenic purpura hereditary?*

(3) *Is there any evidence that hemorrhagic shock and encephalopathy syndrome is passed from parent to child?*

(4) *Will our baby inherit this condition from him?*

(5) *How rare is this disorder?*

(6) *Is Gilles de la Tourette syndrome a rare disease?*

(7) *Are people of certain religious backgrounds more likely to develop this syndrome?*

(8) *If so, what are the chances for an individual to have both conditions?*

(9) *Could you tell me which database I could look at to find the most recent data on neuroblastoma epidemiology; in particular the prevalence in Europe?*

(10) *How many cases of this condition have been identified throughout the world?*

OTHER questions are by definition outliers from the normal question types our QA system is concerned with, though they are about a specific disease. Some more common include questions about a disease's name or history (1-3) and financial considerations (4-6), though many more types of OTHER questions could be categorized (7-10):

(1) *How did Zellweger Syndrome get its name?*

(2) *Does the "M" in MOPD stand for microcephaly or Majewski?*

(3) *How will the recent discovery of the gene related to Kabuki syndrome affect her?*

(4) *Does Medicaid cover genetic testing?*

(5) *Are there any programs in New York to help with the cost of his dental work?*

(6) *Is there any funding out there to help me start some research on the effects of cooking naturally for children with this condition?*

(7) *Can I donate blood?*

(8) *Can any doctor cure Buschke Lowenstein tumor?*

(9) *Does this type of infection get reported to the department of health?*

(10) *Could my family, which has French Canadian heritage, be related to the Canadian family described by Renpenning in the 1960s?*

NOTDISEASE questions in the context of the GARD corpus are generally about genetic tests (1), treatments (2), anatomical objects (3), or gene functions (4 & 5). While there are question types related to many of these (e.g., DIAGNOSIS, MANAGEMENT, ANATOMY), these questions are not specifically about a disease and thus cannot be answered by a disease question answering system:

(1) *Are there factors that may contribute to a falsely positive result?*

(2) *However, being a dextrocardia patient, I wonder if a pacemaker can be put in my heart.*

(3) *Is complex III in the mitochondria itself or is it part of the body?*

(4) *Is there any information about what those genes do?*

(5) *Is there any information about what genes are on chromosome 20q12?*

RESEARCH questions are concerned with the most recent scientific information, and thus the answers likely wouldn't be found in a medical encyclopedia, but in medical publications or clinical trial databases. As the examples below illustrate, RESEARCH questions may come from almost any question type, including CAUSE (1), DIAGNOSIS (2), INFORMATION (3 & 4), MANAGEMENT (5 & 6), OTHEREFFECT (7), PERSONORG (8), PROGNOSIS (9), and SUSCEPTIBILITY (10):

(1) *Has any research been done to check whether there is any viral involvement?*

(2) *If cystic fibrosis testing was done by amniocentesis 6 years ago, have there been any new findings since then?*

(3) *Are there any clinical studies currently being performed?*

(4) *Could you please provide some of the latest information on this disease?*

(5) *Are there any new drugs (even under trial) to treat this condition.*

(6) *I would like to know whether current research is finding ways to cure this disease.*

(7) *Is there any research being done regarding a possible association between a history of diabetes and a diagnosis of Ledderhose disease?*

(8) *Is there any research studies enrolling women with this tumor?*

(9) *Where can I find studies which discuss these risks?*

(10) *Could you tell me which scientific reference I could look at to find the most recent data on neuroblastoma epidemiology; in particular the prevalence in Europe?*

6. Corpus Description

The annotated corpus contains 1,467 requests, with a total of 2,937 decomposed questions. The frequencies of the question types in this corpus is shown in Table 2. The most

Question Type	# Questions	# Requests with Type
ANATOMY	12	7
CAUSE	128	106
COMPLICATION	38	34
DIAGNOSIS	240	181
INFORMATION	531	445
MANAGEMENT	683	497
MANIFESTATION	104	85
OTHEREFFECT	277	199
PERSONORG	126	107
PROGNOSIS	311	244
SUSCEPTIBILITY	422	324
OTHER	52	40
NOTDISEASE	13	9

Table 2: Frequencies of question types in GARD data as well as the frequencies of requests with at least one question of a given type.

common question types are MANAGEMENT and INFORMATION, with almost one-third of all requests containing a question of this type. The next most common types are SUSCEPTIBILITY, OTHEREFFECT, PROGNOSIS, and DIAGNOSIS.

Table 3 shows the top ten words for each question type as ranked by the Fisher Exact Test. For classes with a significant number of annotations (e.g., MANAGEMENT, DIAGNOSIS), these words largely conform to our intuitions. Many of these words were used as trigger words by our previous, rule-based classification system. The fact that they are highly associated in the corpus means that a machine learning system trained on this data should be able to accurately identify questions of these types.

The annotated data set is available on our project website.[8]

7. Conclusion

We have presented a question classification scheme and an annotated corpus based on 2,937 consumer health questions. An automated system trained on this data should be able to aid a consumer question answering system find appropriate strategies for retrieving health information. In future work, we plan to experiment with machine learning approaches to this task as well as determine how well it generalizes to other, more grammatically challenging datasets.

8. References

Somnath Banerjee and Sivaji Bandyopadhyay. 2012. Question Classification and Answering from Procedural Text in English. In *COLING Workshop on Question Answering for Complex Domains*, pages 11–26.

[8]http://lhncbc.nlm.nih.gov/project/consumer-health-question-answering

Question Type	Top 10 Words
ANATOMY	*areas, inside, ip, body, affect, attached, felt, metachondromatosis, internal, researching*
CAUSE	*causes, cause, caused, taking, exposure, factors, trisomy, genes, chemical, increases*
COMPLICATION	*risk, complications, pre, individual, result, compared, country, greater, malaria, paralyzed*
DIAGNOSIS	*testing, diagnosed, diagnosis, test, tests, genetic, available, tested, how, performed*
INFORMATION	*information, about, provide, please, like, would, send, could, how, treatment*
MANAGEMENT	*treatment, treated, treatments, cure, treat, therapy, options, diet, effective, prevent*
MANIFESTATION	*symptoms, adpkd, among, dominant, polycystic, tend, autosomal, members, affected, kidney*
OTHEREFFECT	*between, problems, associated, syndrome, information, normal, issues, about, how, cause*
PERSONORG	*support, find, specialize, doctors, centers, contact, talk, area, groups, treating*
PROGNOSIS	*prognosis, life, expectancy, long, term, expect, affect, person, back, ability*
SUSCEPTIBILITY	*chances, genetic, inherited, rare, chance, information, child, passed, about, prevalence*

Table 3: Most associated words for each question type.

Fan Bu, Xingwei Zhu, Yu Hao, and Xiaoyan Zhu. 2010. Function-based question classification for general QA. In *Proceedings of EMNLP*.

Yllias Chali and Sadid A. Hasan. 2012. Simple or Complex? Classifying the Question by the Answer Complexity. In *COLING Workshop on Question Answering for Complex Domains*, pages 1–10.

Wen Chan, Weidong Yang, Jinhui Tang, Jintao Du, Xiangdong Zhou, and Wei Wang. 2013. Community question topic categorization via hierarchical kernelized classification. In *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management*.

Dina Demner-Fushman and Jimmy Lin. 2007. Answering Clinical Questions with Knowledge-Based and Statistical Techniques. *Computational Linguistics*, 33(1).

Huizhong Duan, Yunbo Cao, Chin-Yew Lin, and Yong Yu. 2008. Searching Questions by Identifying Question Topic and Question Focus. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics*, pages 156–164.

John W. Ely, Jerome A. Osheroff, Mark H. Ebell, George R. Bergus, Barcey T. Levy, M. Lee Chambliss, and Eric R. Evans. 1999. Analysis of questions asked by family doctors regarding patient care. *BMJ*, 319(7206):358–361.

John W. Ely, Jerome A. Osheroff, Paul M. Gorman, Mark H. Ebell, M. Lee Chambliss, Eric A. Pifer, and P. Zoe Stavri. 2000. A taxonomy of generic clinical questions: classification study. *BMJ*, 321:429–432.

John W. Ely, Jerome A. Osheroff, M. Lee Chambliss, Mark H. Ebell, and Marcy E. Rosenbaum. 2005. Answering Physicians' Clinical Questions: Obstacles and Potential Solutions. *J Am Med Inform Assoc*, 12(2).

Ulf Hermjakob. 2001. Parsing and Question Classification for Question Answering. In *Proceedings of the ACL Workshop on Open-Domain Question Answering*.

Andrew Hickl, Kirk Roberts, Bryan Rink, Jeremy Bensley, Tobias Jungen, Ying Shi, and John Williams. 2007. Question Answering with LCC's CHAUCER-2 at TREC 2007. In *Proceedings of TREC-2007*.

Halil Kilicoglu, Marcelo Fiszman, and Dina Demner-Fushman. 2013. Interpreting Consumer Health Questions: The Role of Anaphora and Ellipsis. In *Proceedings of the 2013 BioNLP Workshop*, pages 54–62.

Tetsuya Kobayashi and Chi-Ren Shyu. 2006. Representing Clinical Questions by Semantic Type for Better Classification. In *Proceedings of the AMIA Annual Symposium*.

Vijay Krishnan, Sujatha Das, and Soumen Chakrabarti. 2005. Enhanced Answer Type Inference from Questions using Sequential Models. In *Proceedings of EMNLP*.

X. Li and Dan Roth. 2002. Learning question classifiers. In *Proceedings of COLING*.

Feifan Liu, Lamont D. Antieau, and Hong Yu. 2011. Toward automated consumer question answering: Automatically separating consumer questions from professional questions in the healthcare domain. *Journal of Biomedical Informatics*, 44(6).

Donald Metzler and W. Bruce Croft. 2005. Analysis of Statistical Question Classification for Fact-Based Questions. *Information Retrieval*, 8:481–504.

John M. Prager. 2006. Open-Domain Question Answering. *Foundations and Trends in Information Retrieval*, 1(2):91–231.

Kirk Roberts and Sanda M. Harabagiu. 2012. Locational Relativity and Domain Constraints in Spatial Questions. In *Proceedings of the ACM International Conference on Advances in Geographic Information Systems*.

Kirk Roberts and Andrew Hickl. 2008. Scaling Answer Type Detection to Large Hierarchies. In *Proceedings of LREC*.

Kirk Roberts, Kate Masterton, Marcelo Fiszman, Halil Kilicoglu, and Dina Demner-Fushman. 2014. Annotating Question Decomposition on Complex Medical Questions. In *Proceedings of LREC*.

Connie Schardt, Martha B. Adams, Thomas Owens, Sheri Keitz, and Paul Fontelo. 2007. Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC Med Inform Decis Mak*, 7(16).

Jessi Van Der Volgen, Bethany R. Harris, and Dina Demner-Fushman. 2013. Analysis of Consumer Health Questions for Development of Question-Answering Technology. In *Proceedings of the 2013 Annual Meeting and Exhibition of the Medical Library Association*.

Hong Yu and YongGang Cao. 2008. Automatically Extracting Information Needs from Ad Hoc Clinical Questions. In *Proceedings of the AMIA Annual Symposium*.

Hong Yu and Carl Sable. 2005. Being *Erlang Shen*: Identifying Answerable Questions. In *IJCAI Workshop on Knowledge and Reasoning for Answering Questions*.

Yan Zhang. 2010. Contextualizing Consumer Health Information Searching: An Analysis of Questions in a Social Q&A Community. In *Proceedings of the 1st ACM International Health Informatics Symposium*.