A Framework of a Logic-based Question-Answering System for the Medical Domain (LOQAS-Med)

Sofia J. Athenikos
College of Information Science & Technology, Drexel University
Philadelphia, PA 19104
1-215-895-2474
sofia.j.athenikos@acm.org

Hyoil Han
College of Information Science & Technology, Drexel University
Philadelphia, PA 19104
1-215-895-0493
hyoil.han@acm.org

Ari D. Brooks
College of Medicine
Drexel University
Philadelphia, PA 19102
1-866-373-9352
Ari.Brooks@DrexelMed.edu

ABSTRACT

Question-answering systems that provide precise answers to questions, by combining techniques for information retrieval, information extraction, and natural language processing, are seen as the next-generation search engines. Due to the growth and real-world impact of biomedical information, the need for question-answering systems that can aid medical researchers and health care professionals in their information search is acutely felt. In order to provide users with accurate answers, such systems need to go beyond lexico-syntactic analysis to semantic analysis and processing of texts and knowledge resources. Moreover, question-answering systems equipped with reasoning capabilities can derive more adequate answers by using inference. Research on question answering in the medical and health care domain is still in its inception stage. While several recent approaches to medical question answering have explored use of semantic knowledge, few approaches have exploited the utility of logic formalisms and of inference mechanisms. In this paper, we present a framework for a logic-based question-answering system for the medical domain, which uses Description Logic as the formalism for knowledge representation and reasoning. As a first step toward building the proposed system, we present semantic analysis and classification of medical questions.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval; I.2.1 [**Artificial Intelligence**]: Applications and Expert Systems – *medicine and science*; I.2.3 [**Artificial Intelligence**]: Deduction and Theorem Proving – *answer/reason extraction, deduction, inference engines, logic programming*; I.2.4 [**Artificial Intelligence**]: Knowledge Representation Formalisms and Methods – *predicate logic, representation languages, semantic networks*; I.2.7 [**Artificial Intelligence**]: Natural Language Processing – *language parsing and understanding, text analysis*; J.3 [**Life and Medical Sciences**]: *medical information systems.*

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
SAC'09, March 8-12, 2009, Honolulu, Hawaii, U.S.A.
Copyright 2009 ACM 978-1-60558-166-8/09/03…$5.00.

General Terms

Algorithms, Design, Languages, Theory.

Keywords

Question answering, Biomedical informatics, Description logics, Natural language processing, Semantic text mining.

1. INTRODUCTION

Question Answering (QA) is a field of research which aims to provide inquirers with direct, precise answers to their natural language questions, instead of simply returning a list of relevant documents in response to keyword queries, as in traditional Information Retrieval (IR). QA thus utilizes techniques of Natural Language Processing (NLP) and Information Extraction (IE).

Most of the research developments in the area of QA, as fostered by TREC [28], have so far focused on open-domain QA. Recently, however, the field has witnessed a growing interest in restricted-domain QA [17]. Also, while the earlier TREC QA systems mainly relied on a surface-level lexico-syntactic analysis in locating answers, there has been a growing interest in developing techniques that use semantic knowledge.

The medical domain is one of the most information-critical domains in need of intelligent QA systems that can effectively aid medical researchers and health care professionals in their daily information search.

In order to provide medical information seekers with accurate answers, the QA systems need to go beyond lexico-syntactic level to utilize semantic analysis and processing. Moreover, QA systems equipped with reasoning capabilities can potentially derive implicit evidences by using inference.

In our previous work [2], we briefly presented semantic analysis and classification of medical questions for a logic-based medical QA system. In this paper, we extend the work in presenting a framework of LOQAS-Med (LOgic-based QA System for the Medical domain) which uses Description Logic (DL) [3] for knowledge representation and reasoning (KR&R).

The rest of the paper is organized as follows: First, we discuss the background concerning medical QA. Next, we present semantic analysis and classification of medical questions. Then we describe the proposed QA system. Next, we briefly discuss related work. The final section concludes the paper.

2. BACKGROUND

A dominant paradigm in the medical/clinical field is that of EBM (Evidenced-Based Medicine) [24], which refers to the use of the best evidences obtained from scientific research in making clinical decisions. Within the EBM framework, physicians are urged to ask questions to find the best available evidences.

While several studies investigating the effectiveness of online biomedical resources in answering medical questions have validated the utility of those resources to a certain extent, they have also revealed serious problems in the medical QA process. For example, Ely et al. [8] have found that physicians spend on average two minutes or less in seeking an answer, whereas Hersh et al. [14] have found that it takes more than 30 minutes on average to search for an answer. As a result, many questions go unanswered. Studies investigating obstacles to finding answers to clinical questions [10,11] have identified physicians' doubts about the existence of answers, difficulty of formulating answer-able questions, uncertainty about an optimal search strategy, excessive time required for searching, and failure of the selected resource to provide synthesized answers, as the main factors.

The EBM framework recommends a specific frame for formulating a clinical question, namely, PICO (Problem or Population, Intervention, Comparison, Outcome) [1, 23].

However, Huang et al.'s study [15] has shown that the suitability of PICO as a framework for representing clinical questions is dependent on the type of clinical tasks involved in the questions. In particular, the study has found that the PICO frame is best suited for representing questions about therapy, while less suited for questions concerning diagnosis, etiology, and prognosis.

Within the EBM framework, medical domain-specific question taxonomies have also been developed by several researchers. Bergus et al. [4], for instance, have developed a taxonomy of medical questions according to the PICO elements of questions and the categories of clinical tasks involved in the questions. Most notably, Ely et al. [8-10] have developed a generic taxonomy of common clinical question types and an "evidence taxonomy" of clinical questions, from their empirical studies with mostly primary care doctors. On the top level of the evidence taxonomy, questions are classified into clinical vs. non-clinical. The clinical questions are divided into general vs. (patient-)specific. The general questions are divided into questions involving evidence vs. no evidence. The evidence questions are then classified into those involving intervention vs. no intervention. Ely et al. have concluded that only the evidence type questions are potentially answerable. The generic taxonomy of clinical questions developed by Ely et al. classifies questions using four hierarchical levels of specificity. The first level represents a classification among diagnosis, treatment, management, epidemiology, and non-clinical questions. The secondary, tertiary, and quaternary levels further classify the generic questions. Table 1 shows 10 most common generic clinical questions, ranked in the order of frequency, as identified by Ely et al. [9] from their analysis of 1396 questions collected from primary care doctors in Iowa and Oregon.

Table 1. 10 most common clinical question types [9].

Rank	Question Type
1	What is the drug of choice for condition x?
2	What is the cause of symptom x?
3	What test is indicated in situation x?
4	What is the dose of drug x?
5	How should I treat condition x (not limited to drug treatment)?
6	How should I manage condition x (not specifying diagnostic or therapeutic)?
7	What is the cause of physical finding x?
8	What is the cause of test finding x?
9	Can drug x cause (adverse) finding y?
10	Could this patient have condition x?

3. SEMANTIC CLASSIFICATION AND ANALYSIS OF MEDICAL QUESTIONS

Our ultimate goal is to develop a medical domain-specific QA system that incorporates semantics-based IE and logic-based KR&R for intelligent answer/evidence discovery. A first step in implementing a prototype system toward that goal consists in defining the scope and types of questions to be handled by the system. For this reason, we have conducted semantic analysis and classification of medical question types, based on Ely et al.'s taxonomy of generic clinical questions.

3.1 Classification of Medical Questions

As mentioned, the primary level of classification in Ely et al.'s taxonomy of generic clinical questions concerns broad categories of clinical tasks, namely, diagnosis, treatment, management, and epidemiology. According to this categorization, questions #1, 4, 5, and 9 in Table 1 belong to the treatment category, questions #2, 3, 7, 8, 10 belong to diagnosis, whereas question #6 belongs to management. However, we note that a different kind of semantic classification can be applied to these questions, based on the type of question focus. For example, questions #2, 7, 8, 9 all concern the cause-effect relation. In this regard, we observe that the clinical task-oriented scheme of classification in Ely et al.'s taxonomy tends to rather obscure the semantic relations between the questions classified. We have also noted some inconsistencies in the way subcategories are organized. For example, the second-level category that corresponds to questions #2, 7, 8 is "cause/interpretation of clinical finding", whereas the one that corresponds to question #9 is "drug prescribing". Moreover, while Table 1 shows one representative question form for each generic question type (as identified by Ely et al.), Ely et al.'s taxonomy lists several variations for each category, which include semantically distinct questions.

Based on the above observations (and based on the manual inspection of 1095 Iowa physicians' questions from Ely et al.'s study [8], accessed via the NLM Clinical Questions Collection [21]), we have devised our own semantic categorization scheme to re-classify the most common generic clinical questions (except #10 which concerns a strictly patient-specific inquiry) and some of their variations.

Table 2 shows the first two levels of categorization according to our scheme. (Note: The numbers in parentheses correspond to the question (category) frequency ranks in Table 1.) As shown in Table 2, the first level of classification is based on the distinction of types of semantic relations involved in questions (e.g., cause-effect), whereas the second level is based on the types of semantic classes involved (e.g., drug). The third level further classifies questions depending on whether or not they ask about specific targets (any vs. specific), while the fourth level classifies them depending on whether or not they concern particular situations (general vs. contextual).

Table 2. Classification scheme for medical questions.

Category	Division
Cause-Effect	Symptom (Cause of) (#2)
	Physical Sign (Cause of) (#7)
	Test Finding (Cause of) (#8)
	Drug (Adverse Effect of) (#9)
Method	Drug (#1)
	Diagnosis (#3)
	Evaluation/Test (#3)
	Treatment (#5)
	Management (#6)
Indication	Drug (#1)
	Evaluation/Test (#3)
	Treatment (#5)
Efficacy	Drug (#1)
	Treatment (#5)
Quantity	Drug (#4)
Discrimination	Condition (#3)
Significance	Test Finding (#8)

Table 3 shows generic question types that correspond to the category cause-effect, in particular, those that belong to the cause of symptom category.

Table 3. Questions concerning cause of symptom.

Cause-Effect → Symptom → Any → General: What is the cause of symptom X?
Cause-Effect → Symptom → Any → Contextual: What is the cause of symptom X in situation Z?
Cause-Effect → Symptom → Specific → General: Can condition Y cause symptom X?
Cause-Effect → Symptom → Specific → Contextual: Can condition Y cause symptom X in situation Z?

3.2 Analysis of Medical Questions

The advantage of our classification scheme for QA processing consists in the fact that we can classify questions based on the semantic relations represented by the predicates first and further classify them in terms of the semantic types corresponding to the arguments. Based on the scheme, therefore, we have semantically analyzed each question category. More specifically, we have constructed question and answer patterns as semantic triples in the form of subject-predicate-object, based on the identification of the UMLS (Unified Medical Language System) Semantic Network [27] semantic types and semantic relations that correspond to the semantics of the arguments and predicates represented by each question type.

We illustrate the construction of question and answer patterns using the cause of symptom category. We here restrict ourselves to the "general" questions, i.e., the first and third types of questions in Table 3.

First, in the case of the generic question of the form "What is the cause of symptom X?", the semantic relation between X and the target can be represented either by "Target <Causes> X" or by "X <Result_of> Target". Also, the semantic type of X is <Sign or Symptom>, while the semantic type of the target (i.e., the cause of X) should be <Disease or Syndrome>. The resulting question and answer patterns are shown in Table 4. Note that the patterns include alternative semantic relations that may be considered in the processing of the question and candidate answers.

Table 4. Q/A patterns involving unspecified cause of symptom.

<table>
<tr><td>Q Pattern</td><td>Subj: What<Disease or Syndrome>
Pred: <Causes>(||<Produces>||<Complicates>
||<Precedes>||<Diagnoses>)
Obj: X<Sign or Symptom></td></tr>
<tr><td>A Pattern 1</td><td>Subj: <Disease or Syndrome>
Pred: <Causes> (||<Produces>||<Complicates>
||<Precedes>||<Diagnoses>)
Obj: X<Sign or Symptom></td></tr>
<tr><td>A Pattern 2</td><td>Subj: X<Sign or Symptom>
Pred: <Result_of>(||<Manifestation_of>
||<Indicates>||<Co-occurs_with>)
Obj: <Disease or Syndrome></td></tr>
</table>

In the case of the generic question of the form "Can condition Y cause symptom X?", the semantic relation between X and Y can be represented either by "Y <Causes> X" or by "X <Result_of> Y". The semantic type corresponding to X is <Sign or Symptom>, while that corresponding to Y is <Disease or Syndrome>, and both X and Y must be specifically matched in the answers. Note that, as shown in Table 5 below, question patterns are identical to answer patterns for this type of question. This is due to the fact that both arguments are specified in the question. In other words, the QA process for this type of generic question can be thought of as a form of hypothesis testing.

Table 5. Q/A patterns involving specified cause of symptom.

<table>
<tr><td>Q Pattern 1</td><td>Subj: Y<Disease or Syndrome>
Pred: <Causes>(||<Produces>||<Complicates>
||<Precedes>||<Diagnoses>)
Obj: X<Sign or Symptom></td></tr>
<tr><td>Q Pattern 2</td><td>Subj: X<Sign or Symptom>
Pred: <Result_of>(||<Manifestation_of>
||<Indicates>||<Co-occurs_with>)
Obj: Y<Disease or Syndrome></td></tr>
<tr><td>A Pattern 1</td><td>Subj: Y<Disease or Syndrome>
Pred: <Causes> (||<Produces>||<Complicates>
||<Precedes>||<Diagnoses>)
Obj: X<Sign or Symptom></td></tr>
<tr><td>A Pattern 2</td><td>Subj: X<Sign or Symptom>
Pred: <Result_of>(||<Manifestation_of>
||<Indicates>||<Co-occurs_with>)
Obj: Y<Disease or Syndrome></td></tr>
</table>

4. ARCHITECTURE OF LOQAS-MED

Our choice of Description Logic (DL) [3] as the KR&R formalism for LOQAS-Med is based on the fact that DL provides a flexible way of representing the semantics of the predicate-argument structure and that it is particularly well-suited for subsumption reasoning. In this regard, we also note that some of the prominent semantic knowledge resources in the biomedical domain are encoded in DL (e.g., [13, 18]). We have decided to use OWL DL [22] in particular, considering the availability of compatible reasoners and interoperability with other existing and forthcoming OWL-based resources.

The proposed QA system is intended to provide answers to medical questions based on explicitly-stated facts as well as to derive hypothesis-confirming or hypothesis-denying evidences by utilizing inference. The system will thus have two modes of operation: strict QA mode and research mode. Figure 1 shows the architecture of the QA system (in the QA mode).

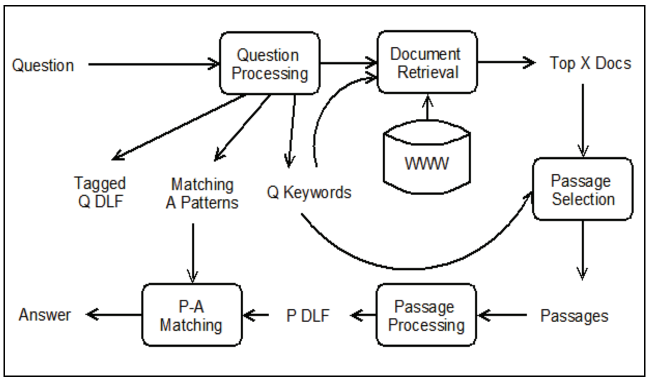

Figure 1. Architecture of LOQAS-Med.

As shown in Figure 1, the system first processes the user-input question, generating a semantically tagged question DL form (DLF), corresponding answer DLF patterns, and semantically expanded question keywords. In the document retrieval stage, the question keywords are fed into a Web search engine, such as Google or PubMed, to retrieve and rank relevant documents. From the ranked list, the top *X* documents are selected. Next, in the passage selection stage, the question keywords are again used to extract question-relevant passages from the selected documents. The extracted passages go through the passage processing phase during which a semantically-tagged passage DLF is generated for each candidate passage. The system matches each passage DLF against the target answer DLF patterns identified in the question processing phase in order to select answer(s). Finally, the system presents the selected answer(s) to the user.

Figure 2 shows the structure of question processing in the system. The question is first parsed to generate a syntactic dependency tree (DT) structure. Based on a preconstructed set of transformation rules, the system derives Q DLF from the dependency tree. Next, the Q DLF is tagged with semantic information obtained from the UMLS Semantic Network, by mapping the arguments and predicate(s) to the semantic types and semantic relation(s), respectively. The tagged Q DLF is then matched against a preconstructed set of Q patterns (see Tables 4-5). If a matching pattern is found, the Q pattern is indexed into a preconstructed set of answer patterns in order to identify matching answer patterns.

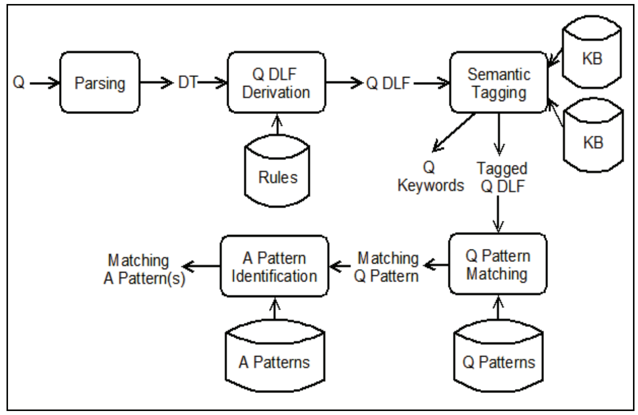

Figure 2. Question processing in LOQAS-Med.

5. RELATED WORK

Niu et al. [19, 20] and Demner-Fushman et al.[6, 7] have presented semantics-based approaches to medical QA. Their approaches are mainly based on the identification and extraction of PICO elements (in particular, Outcome) from medical texts. Jacquemart, Zweigenbaum, and Delbecque [5, 16] have investigated semantic modeling of medical questions in the form of concept-relation-concept triples and extraction of subject-verb-complement clauses from medical texts. Their approaches, however, are not based on the classification of medical questions in terms of the type of semantic relations covered. The task of extracting semantic predications from medical texts has been studied by Fiszman et al. [12]. Terol et al. [25, 26] have presented a logic-based medical QA system that is intended to handle the 10 common question types identified by Ely et al. (see Table 1). The Q-A pattern matching used by their system, however, is mainly based on the matching of the number of medical entities of corresponding semantic types.

6. CONCLUSION

In this paper, we have presented a framework of semantic analysis and classification of medical questions as a first step in building a logic-based medical domain-specific QA system. The proposed QA system uses Description Logic (DL), OWL DL, in particular, as the basis for knowledge representation and reasoning. The system is intended to function both as a direct answer provider and as an evidence prospecting indicator, addressing the critical information- and evidence-seeking needs in the medical domain. It will extract explicitly-stated answers and derive hypothesis-supporting/denying evidences for a given natural language question/hypothesis by applying NLP techniques and KR&R techniques to both full-text primary medical research articles and domain-specific terminological resources. For an initial prototype QA system we are implementing, we have decided to focus on the questions involving the cause-effect relation in order to provide QA and evidence discovery functionalities for this important category of questions, which has been shown by some empirical studies to constitute the most common type of medical questions asked/answered by health care consumers/providers. We believe that the proposed research can make a significant impact both as an empirical contribution to the field of artificial intelligence and as a real-world application in the critical domain of medical and health care informatics.

7. ACKNOWLEDGMENTS

We thank anonymous reviewers for their helpful comments.

8. REFERENCES

[1] Armstrong, E.C. The well-built clinical question: the key to finding the best evidence efficiently. *WMJ*, 98 (1999), 25-28.

[2] Athenikos, S.J., Han, H., and Brooks, A.D. Semantic analysis and classification of medical questions for a logic-based medical question-answering system. In *Proceedings of the International Workshop on Biomedical and Health Informatics (BHI 2008) in conjunction with 2008 IEEE Conference on Bioinformatics and Biomedicine (IEEE BIBM 2008)* (Philadelphia, PA, Nov. 3-5, 2008), 111-112.

[3] Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., and Patel-Schneider, P.F. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, West Nyack, NY, 2003.

[4] Bergus, G.R., Randall, C.S., Sinift, S.D., and Rosenthal, D.M. Does the structure of clinical questions affect the outcome of curbside consultations with specialty colleagues? *Arch Fam Med*, 9 (2000), 541-547.

[5] Delbecque, T., Jacquemart, P., and Zweigenbaum, P. Indexing UMLS semantic types for medical question-answering. In *Connecting Medical Informatics and Bio-Informatics*. ENMI, 2005, 805-810.

[6] Demner-Fushman, D., Few, B., Hauser, S.E., and Thoma, G. Automatically identifying health outcome information in MEDLINE records. *J Am Med Inform Assoc*, 13 (2006), 52-60.

[7] Demner-Fushman, D., and Lin, J. Answering clinical questions with knowledge-based and statistical techniques. *Computational Linguistics*, 33 (2007), 63-103.

[8] Ely, J.W., Osheroff, J.A., Ebell, M.H., Bergus, G.R., Levy, B.T., Chamliss, M.L., and Evans, E.R. Analysis of questions asked by family doctors regarding patient care. *BMJ*, 319 (1999), 358-361.

[9] Ely, J.W., Osheroff, J.A., Gorman, P.N., Ebell, M.H., Chambliss, M.L., Pifer, E.A., and Stavri, P.Z. A taxonomy of generic clinical questions: classification study. *BMJ*, 321 (2000), 429-432.

[10] Ely, J.W., Osheroff, J.A., Ebell, M.H., Chambliss, M.L., Vinson, D.C., Stevermer, J.J., amd Pifer, E.A. Obstacles to answering doctors' questions about patient care with evidence: qualitative study. *BMJ*, 324 (2002), 710-716.

[11] Ely, J.W., Osheroff, J.A., Chambliss, M.L., Ebell, M.H., and Rosenbaum, M.E. Answering physicians' clinical questions: obstacles and potential solutions. *J Am Med Inform Assoc*, 12 (2005), 217-224.

[12] Fiszman, M., Rindflesch, T.C., and Kilicoglu, H. Integrating a hypernymic proposition interpreter into a semantic processor for biomedical texts. In *Proceedings of the 2003 AMIA Annual Symposium (AMIA 2003)* (Washington, D.C., Nov. 8-12, 2003). AMIA, 2003, 239-243.

[13] GeneOntology. http://www.geneontology.org/.

[14] Hersh, W.R., Crabtree, M.K., Hickman, D.H., Sacherek, L., Friedman, C.P., Tidmarsh, P., Mosbaek, C., and Kraemer, D. Factors associated with success in searching MEDLINE and applying evidence to answer clinical questions. *J Am Med Inform Assoc*, 9 (2002), 283-293.

[15] Huang, X., Lin, J., and Demner-Fushman, D. Evaluation of PICO as a knowledge representation for clinical questions. In *Proceedings of the 2006 AMIA Annual Symposium (AMIA 2006)* (Washington, D.C., Nov. 11-16, 2006). AMIA, 2006, 359-363.

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

[17] Mollá, D., and Vicedo, J.L. Question answering in restricted domains: an overview. *Computational Linguistics*, 33 (2007), 41-61.

[18] NCI (National Cancer Institute) Thesaurus. www.nci.nih.gov/cancerinfo/terminologyresources/.

[19] Niu, Y., and Hirst, G. Analysis of semantic classes in medical text for question answering. In *Proceedings of the ACL 2004 Workshop on Question Answering in Restricted Domains* (Barcelona, Spain, Jul 25, 2004). ACL, 2004.

[20] Niu, Y., Zhu, X., and Hirst, G. Using outcome polarity in sentence extraction for medical question-answering. In *Proceedings of the 2006 AMIA Annual Symposium (AMIA 2006)* (Washington, D.C., Nov. 11-16, 2006). AMIA, 2006, 599-603.

[21] NLM Clinical Questions Collection. http://clinques.nlm.nih.gov/.

[22] OWL Web Ontology Language Guide. http://www.w3.org/TR/owl-guide/.

[23] Richardson, W.S., Wilson, M.C., Nishikawa, J., and Hayward, R.S. The well-built clinical question: a key to evidence-based decisions. *ACP J Club*, 123 (1995), A12-13.

[24] Sackett, D.L., Strauss, S., Richardson, W., Rosenberg, W., and Haynes, R. *Evidence-Based Medicine: How to Practice and Teach EBM* (2nd ed.). Churchill Livingstone, Edinburgh, UK, 2000.

[25] Terol, R.M., Martínez -Barco, P., and Palomar, M. Applying NLP Techniques and Biomedical Resources to Medical Questions in QA Performance. In *Proceedings of the Fifth Mexican International Conference on Artificial Intelligence (MICAI 2006)* (Apizaco, Mexico, Nov. 13-17, 2006). Springer-Verlag, Berlin; Heidelberg, 2006, 996-1006.

[26] Terol, R.M., Martínez-Barco, P., and Palomar, M. A knowledge based method for the medical question answering problem. *Computers in Biology and Medicine*, 27 (2007), 1511-1521.

[27] UMLS (Unified Medical Language System). www.nlm.nih.gov/research/umls/.

[28] Vorhees, E.M. The TREC question answering track. *Natural Language Engineering*, 7, 4 (2001), 361-378.