

Research Article

Question Processing and Clustering in INDOC: A Biomedical Question Answering System

Parikshit Sondhi, Purushottam Raj, V. Vinod Kumar, and Ankush Mittal

Department of Electronics and Computer Engineering, Indian Institute of Technology Roorkee, Roorkee 247667, India

Received 12 April 2007; Accepted 22 September 2007

Recommended by Paola Sebastiani

The exponential growth in the volume of publications in the biomedical domain has made it impossible for an individual to keep pace with the advances. Even though evidence-based medicine has gained wide acceptance, the physicians are unable to access the relevant information in the required time, leaving most of the questions unanswered. This accentuates the need for fast and accurate biomedical question answering systems. In this paper we introduce INDOC—a biomedical question answering system based on novel ideas of indexing and extracting the answer to the questions posed. INDOC displays the results in clusters to help the user arrive at the most relevant set of documents quickly. Evaluation was done against the standard OHSUMED test collection. Our system achieves high accuracy and minimizes user effort.

Copyright © 2007 Parikshit Sondhi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. INTRODUCTION

An estimate of the around 14 million citations in PubMed [1] database of National Library of Medicine clearly indicates the exponential growth of published biomedical literature. It is thus impossible for any individual to keep pace with the advances. Thus, though evidence-based medicine has gained wide acceptance [2–5], the physicians are unable to access the relevant information in the required time, leaving most of the questions unanswered [6]. The problem is further compounded by the inadequacy of the current search engines to perform well with biomedical literature. In a study conducted with a test set of 100 medical questions collected from medical students in a specialized domain, a thorough search in Google was unable to obtain relevant documents within top five hits for 40% of the questions [7]. The current search engines fail to satisfy a user's need for primarily two reasons.

- (1) Focus is more on keyword matching rather than semantics or relations between keywords.
- (2) Lack of understanding of complex biomedical terminology and its inconsistent use [8].

Hence there is a need to develop fast and effective question answering systems for the biomedical domain [9–11].

A number of strategies have been proposed for answering biomedical questions such as, answering by role identification [5, 12, 13], and answering based on document structure [14]. A survey of recent works can be found in [15].

In this paper, we present the design and implementation of Internet Doctor (INDOC), a biomedical question answering system. The system involves modules to perform indexing, question processing, document ranking, clustering, and display.

The paper is organized into 4 sections. The architecture of the system is presented in Section 2. Section 3 presents the performance analysis of the system and Section 4 describes future work and the conclusions.

2. ARCHITECTURE OF INDOC

The architecture of INDOC is as shown in Figure 1. The entire document set is first indexed by the indexing module. A detailed explanation of the indexing method is given later. At runtime, the query from the user is processed by the question processing module recognizing the difference in significance of different parts of the query, and the ranking module ranks the documents by assigning weights on the basis of their relevance to the question. Finally, the display module displays the documents in a decreasing order of their weights. It also

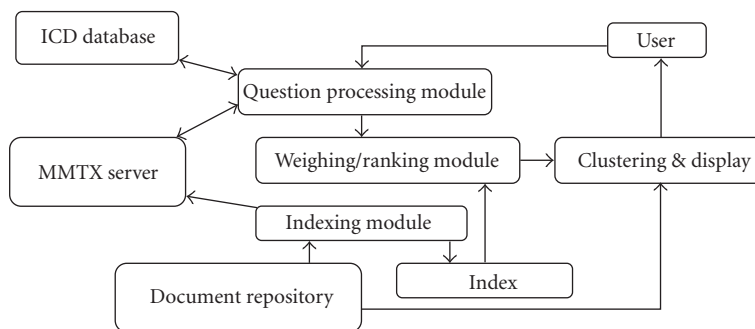


FIGURE 1: Complete architecture of the system.

You asked: Tell me about pathophysiology and treatment of disseminated intravascular coagulation

Disseminated intravascular coagulation.

Known variously as disseminated intravascular coagulation, defibrination consumption coagulopathy or, more simply, as defibrination, disseminated intravascular coagulation is a serious epiphenomenon that occurs most often as a complicating factor of an underlying disease process. Although frequently triggered by underlying disease such as infection or tumor, if not recognized and treated appropriately, disseminated intravascular coagulation alone may lead to the patient's death as a result of hemorrhage or thrombosis, or both, of vital organs. Frequently, it may only manifest itself as an abnormality of coagulation tests, causing no immediate problem for the patient, and potentially normalizing when the inciting cause is appropriately managed. The central process that marks disseminated intravascular coagulation is the generation of thrombin in the circulating blood by means of the activation of the coagulation mechanism, leading to the conversion of fibrinogen to fibrin, which, in turn, may lead to thrombosis mainly of the microcirculation. Because platelets and coagulation factors are consumed and fibrinolysis is enhanced during the coagulation process, hemorrhage may also ensue. Although disseminated intravascular coagulation is frequently encountered in medical and obstetric patients, the difficulty in diagnosis and controversy regarding optimal therapy are frustrating for both patient and physician. By understanding the pathophysiology of disseminated intravascular coagulation and combining clinical observation and laboratory data, one can arrive at the appropriate diagnosis. Therapy must be individualized, and assessment of the benefit versus risk ratio of intervention must be made. Early recognition of acute and life-threatening disseminated intravascular coagulation can be lifesaving with appropriate supportive measures.

Potentially catastrophic bleeding disorders. Approach to diagnosis and management.

Clinical and laboratory evaluation of severe bleeding can detect the presence of an intrinsic or acquired coagulation disorder. The three most common inherited coagulation disorders are factor VIII deficiency (hemophilia A), factor IX deficiency (hemophilia B), and von Willebrand's disease. Vitamin K deficiency, liver disease, and disseminated intravascular coagulation are the most common acquired disorders. A thorough clinical history is crucial to diagnosis. Screening tests that measure prothrombin time, partial thromboplastin time, thrombin time, and platelet count permit initial classification and guide selection of more specific tests. Results can then be used to determine appropriate therapy.

FIGURE 2: Screen-shot of the results.

In Cluster 1 of 46 documents and (rank 1) <ol style="list-style-type: none"> 1) Lupus anticoagulant, thrombosis and thrombocytopenia in systemic lupus erythematosus. 2) The prevalence of autoantibodies during third-trimester pregnancy complicated by hypertension or idiopathic fetal growth retardation. 3) Hepatic infarction in a patient with the lupus anticoagulant. 4) Genitourinary complications of Inflammatory bowel disease. 5) Tube drainage of lung abscesses. 6) Medical and surgical complications of renal transplantation: diagnosis and management. 7) Medical complications and medical management of bulimia. 8) Ureteral obstruction as a complication of the Burch 	In Cluster 2 of 9 documents and (rank 2) <ol style="list-style-type: none"> 1) Toxic shock syndrome as a complication of dermatologic surgery. 2) Squamous cell carcinoma as a late complication of intracerebroventricular epidermoid cyst. Case report. 3) Syringomyelia as a complication of Paget's disease. Case report. 4) Contralateral intracerebral hemorrhage as a complication of tumor biopsy. 5) Accidental hyperthermia as a complication of extracorporeal shock-wave lithotripsy under general anesthesia. 6) Double channel pylorus as a complication of carcinoma of the stomach.
In Cluster 3 of 42 documents and (rank 4) <ol style="list-style-type: none"> 1) Lupus anticoagulant. Implications for obstetric anaesthetists. 2) Vaso-occlusive retinopathy associated with antiphospholipid antibodies (lupus anticoagulant retinopathy) [see comments] 3) Autoantibodies and rheumatic disorders in a neurology inpatient population: a prospective study. 4) Lupus anticoagulants, anticardiolipin antibodies, and fetal loss. A case-control study. 5) The lupus anticoagulant and retinal vaso-occlusive disease. 6) Effect of lupus anticoagulants on the activated partial thromboplastin time. Results of the College of American Pathologists survey program. 7) Ischemic stroke in a girl with lupus anticoagulant. 	In Cluster 4 of 28 documents and (rank 3) <ol style="list-style-type: none"> 1) Thrombotic endocarditis and lupus anticoagulant. A pathogenetic possibility for 'Idiopathic 'rheumatic type' valvular heart disease. 2) Epidemiology and treatment of tinea capitis: ketoconazole vs. griseofulvin. 3) Primary care epidemiology: new uses of old tools. 4) Reye's syndrome, salicylates, epidemiology, and public health policy [editorial] 5) Nutritional epidemiology and national surveys. 6) The epidemiology of breast cancer. 7) Mechanisms of disease in Hantavirus infection: pathophysiology of hemorrhagic fever with renal syndrome.

FIGURE 3: Clustered display of the result-set.

clusters the result-set, marks the most relevant portions of each document, and thus reduces the user effort required in locating the answer. In order to tackle the problems with complex biomedical terminology and its inconsistent use, we have used the UMLS concepts [16] instead of keywords. The task of parsing the text and returning the relevant concepts is performed by MMTX [17], a programming implementation of MetaMap [18].

2.1. MMTX server

The MMTX program is used to map the free text into corresponding UMLS concepts. This operation of concept mapping is performed both while indexing the documents and while processing the query. However, as creating an MMTX object is expensive and takes a considerable amount of time, we implemented a server which instantiates an MMTX object once and waits for free text (which may be either a query string or a document) to be sent. It then returns back the mapping concepts.

2.2. Indexing

Unlike other indexing techniques, we do not just select the important keywords or concepts. Rather, the entire document is represented in the form of sections as shown in Section 2.2.1. Each section has a section heading and a number of sentences in it. The section heading consists of one or more UMLS concepts that represent the section. Further, only successive sentences can belong to a section and any individual sentence cannot be present in more than one section. At the time of document retrieval, a document may be considered useful if some or all of the question concepts are present in one of the section headings. In order to minimize the runtime overhead, we also store all the concepts present in a document.

2.2.1. Indexed representation

Sample document

“Lack of attenuation of a candidate dengue 1 vaccine (45AZ5) in human volunteers. A dengue type1, candidate live virus vaccine (45AZ5) was prepared by serial virus passage in fetal rhesus lung cells. Infected cells were treated with a mutagen, 5-azacytidine, to increase the likelihood of producing attenuated variants. The vaccine strain was selected by cloning virus that produced only small plaques in vitro and showed reduced replication at high temperatures (temperature sensitivity). Although other candidate live dengue virus vaccines, selected for similar growth characteristics, have been attenuated for humans, two recipients of the 45AZ5 virus developed unmodified acute dengue fever. Viremia was observed within 24 hr of inoculation and lasted 12 to 19 days. Virus isolates from the blood produced large plaques in cell culture and showed diminished temperature sensitivity. The 45AZ5 virus is unacceptable as a vaccine candidate. This experience points out the uncertain relationship between in vitro viral growth characteristics and virulence factors for humans.”

Corresponding indexed form

*Lacking (qualifier value) |attenuation |Dengue |Vaccines
|Human Volunteers |*
*Lacking (qualifier value) |attenuation |Dengue |Vaccines
|Human Volunteers|*
0 0
Cells|
1 2
Selection (Genetics) |Virus|
3 5
Virus|.

2.2.2. Algorithm

The algorithm to perform the task of indexing is shown in Algorithm 1.

The algorithm begins by first obtaining all the concepts in the title and storing them in the index file. This is done as the title is usually a good indicator of the content of the document.

The first phase involves formation of sections on the basis of concepts present in the sentences. It begins by adding S_1 , the first sentence of the document to the section X_1 and all its concepts SC_1 to XC_1 . We then add the next sentence S_2 into X_1 and update the concepts in section heading to $XC_1 = XC_1 \cap SC_2$. The section heading thus contains the concepts common to both the sentences. This process is carried out till we find a sentence S_j for which $XC_1 \cap SC_j$ is an empty set. However, the above steps done alone leave a problem unsolved.

Suppose a section X_i has m (m is large) sentences and the concept set XC_i has $n1$ concepts. Thus effectively m sentences are relevant to $n1$ concepts. Now if we try to add a new sentence S_j to the current section X_i such that $|XC_i \cap SC_j| = n2$ ($n2 < n1$), we miss out $n1 - n2$ concepts which are also used frequently in the section.

In order to avoid this, we define a constant M which is the minimum number of sentences to help us decide when to add a new sentence.

- (1) For $|X_i| < M$ —the sentence is added if it contains at least one of the concepts present in the section heading.
- (2) For $|X_i| > M$ —the sentence is added if it contains all the concepts present in the section heading, otherwise, we start constructing a new section X_{i+1} .

Once the formation of sections is complete, we need to perform the task of section merging. This step is necessary because of the following.

- (1) The size of some sections may become too small. In the extreme case, we might end up with just a single sentence in a section. To handle this we define L , the minimum number of sentences to be present in a section. If for a section X_i , $|X_i| < L$ then we merge it with the previous section X_{i-1} . Since $|X_i|$ is very small, the concepts in the set XC_i are not of much importance and hence can be discarded.

```

1. Obtain the concepts of the title and store them.
2. Initialize  $i = 1$  and  $j = 1$ , and set all  $X_i, SC_j, XC_i$ 
   to be empty where
    $S_j$ :  $j$ th sentence in the document
    $X_i$ :  $i$ th section
    $SC_j$ : set of concepts in  $j$ th sentence (concepts in
   an individual sentence)
    $XC_i$ : set of concepts in  $i$ th section
    $L$ : min number of sentences necessary in a
   section
    $M$ : minimum number of sentences in a section
   so that merging is not necessary
3. Formation of sections
   Set  $XC_i$  to concepts in the first sentence.
   Define  $|S|$  as the number of elements in set  $S$ .
   For each sentence  $S_j$  left in the document to
   process
   {
       If ( $|X_i| == 0$ ) {
           Add  $S_j$  to  $X_i$ 
           Add  $SC_j$  to  $XC_i$ 
       }
       else {
           if ( $(|X_i| < M \& \& |XC_i \cap SC_j| > 0) \& \&
           XC_i == SC_j$ )
           {
               Add  $S_j$  to  $X_i$ 
               Set  $XC_i = XC_i \cup SC_j$ 
           }
           else {
                $i = i + 1$ 
               Add  $S_j$  to the new section  $X_i$ 
               Add  $SC_j$  to  $XC_i$ 
           }
       }
   }
4. Final section merging step
   for each section  $X_i$ 
   {
       If ( $i > 1 \& \& |X_i| < L \& \& XC_i$  is a subset
       of  $XC_{i-1}$ ) {
           Merge  $X_i$  with  $X_{i-1}$ 
       }
   }

```

ALGORITHM 1

- (2) There may be cases where XC_i is a subset of XC_{i-1} . In such scenarios, X_i will be merged with X_{i-1} .

In either case, the set XC_{i-1} is left unchanged.

For scaling the algorithm to a large document set, we need to maintain a Concepts X Document matrix containing the section-heading concepts and the corresponding documents in which they are present. This would save us the expense of performing large file operations on indexed files of all documents that need to be done while answering the question. For the evaluation performed by us, since the document set was not excessively large, we could get equally good performance even without such matrix.

2.3. Question processing

The query input by the user is sent to the MMTX server which returns back the UMLS concepts present in it. For example,

Question

Tell me about pathophysiology and treatment of disseminated intravascular coagulation.

Concepts

Disseminated Intravascular Coagulation, Therapeutic procedure, physiopathological, therapeutic aspects.

However, all the key-concepts are not equally important. In the above example, the concept “disseminated intravascular coagulation” is of higher importance as compared to the rest. Therefore, different concepts need to be assigned different weights based on their relative importance, which is decided from their semantic type [19, 20]. In order to identify the relative importance of the semantic types, we analyzed 106 biomedical questions from the OHSUMED test collection [21]. The results are as shown in Table 1, where frequency of various semantic groups in the questions is presented.

From this analysis, it is quite clear that most questions are centered on concepts & ideas (CONC), disorders (DISO), and procedures (PROC); and therefore these semantic types are given higher weights.

In general, the mapped concepts from MMTx alone do not capture all the related senses of a key-concept. For example, back pain and lower back pain are mapped differently, thus a query for lower back pain will not look for back pain and vice versa. We have used the disease classification from the ICD-9-CM to deal with this problem.

2.4. ICD database of related terms

The query concepts with the highest weights are sent to the ICD-9-CM database to obtain a set of related concepts. The search for relevant documents is done on the basis of all these concepts along with the original concepts in the query.

ICD-9-CM stands for International Classification of Diseases, Ninth Revision, Clinical Modification. It is based on the World Health Organization’s Ninth Revision, International Classification of Diseases (ICD-9). It is the official system of assigning codes to diagnoses and procedures associated with hospital utilization in the United States [22].

The ICD-9-CM consists of:

- A numerical list of the disease code numbers in tabular form;
- An alphabetical index to the disease entries; and
- A classification system for surgical, diagnostic, and therapeutic procedures (alphabetic index and tabular list).

TABLE 1: Analysis of questions.

Abbreviation	Semantic group	Frequency
ACTI	Activities & behaviors	27
ANAT	Anatomy	13
CHEM	Chemicals & drugs	58
CONC	Concepts & ideas	137
DEVI	Devices	1
DISO	Disorders	144
GENE	Genes & molecular sequences	0
GEOG	Geographic areas	0
LIVB	Living beings	9
OBJC	Objects	2
ORGA	Organizations	0
OCCU	Occupations	2
PHEN	Phenomena	3
PHYS	Physiology	9
PROC	Procedures	89

All terms in the same parental three-digit code are related and a search can be made for all of these terms whenever a search for any disease in a group is made. For example, Cholera is given code 001 with the following subclassifications.

- (i) 001 cholerae
- (ii) 001.0 Due to *Vibrio cholerae*
- (iii) 001.1 Due to *Vibrio cholerae el tor*
- (iv) 001.9 Cholera, unspecified.

Using ICD database the focus terms (Disseminated Intravascular Coagulation, Therapeutic procedure, physiopathological, therapeutic aspects) of the question mentioned in the previous section are expanded into the following set.

“Disseminated Intravascular Coagulation, Therapeutic procedure, physiopathological, therapeutic aspects, Acquired coagulation factor deficiency NOS (disorder), Afibrinogenemia, Antithromboplastinogenemia, Blood Coagulation Disorders, Blood Coagulation Factor, Blood coagulation pathway observation, Blood coagulation tests, Circulating anticoagulants, Coagulation Therapy, Coagulation factor deficiencies, Coagulation procedure, Congenital deficiency (morphologic abnormality), coagulation, Disseminated Intravascular Coagulation, Dysfibrinogenemia (disorder), Fibrinogen, Hemorrhagic Disorders, Hemorrhagic disorder due to antithrombinemia (disorder), Hemostasis procedure, Pathologic fibrinolysis, Thrombolytic Therapy, Thromboplastin, Unfractionated heparin (substance).”

After the question processing is performed with the help of this diseases classification, we proceed to the document retrieval and their subsequent ranking.

2.5. Document ranking

This step involves assigning the documents a weight on the basis of their relevance to the question. For each document, we search the index file to see which section headings match

the question concepts. We are interested in sections whose headings have at least one of the question concepts. The corresponding sentences are checked to see if they contain any more of the question concepts, which are not present in the heading. Thus, the score of each section is the sum of weights of question concepts present in it. If matches are found in two consecutive sections then they can be combined to form a bigger section, so as to highlight them together while providing the answer. Further, we can also include the neighboring sections of a selected section in order to ensure that no irrelevant sentences are skipped.

Weight of the document W_d is given by the (1):

$$W_d = N_d + \log_{10}(N_l), \quad (1)$$

where N_d = sum of weights of all the matched concepts in the best section and N_l = number of lines in the best section.

Here, by best section, we refer to the section that has the maximum total weight of question concepts.

We justify the importance of N_l as it gives a measure of the relevant information in the current document. Between two documents with same number of concept matches, the document with higher value of N_l contains more information.

Logarithm of N_l is taken because N_d , the total weight of all concept matches, is of higher significance. Since the document weight (W_d) is calculated on the basis of concepts present in the best section and not in the entire document, we are sure that the concepts appear in proximity, and are not just arbitrarily present.

2.6. Clustering

We clustered the final document set so as to make it easier for the user to arrive at the most relevant set of documents, not just one best document.

For clustering the documents, we employed k -means clustering. The algorithm steps [23] are as follows.

- (i) Choose the number of clusters, k .
- (ii) Randomly generate k clusters and determine the cluster centers, or directly generate k random points as cluster centers.
- (iii) Assign each point to the nearest cluster center.
- (iv) Recompute the new cluster centers.
- (v) Repeat the two previous steps, stopping when the assignment does not change anymore.

The maximum number of clusters to be formed can either be fixed beforehand or specified separately for each query by the user. For our analysis, we fixed the number of clusters to four.

The distance measure used for clustering is Euclidean, based on the occurrence of key-concepts present in the question. Each document is represented in terms of a vector of weights that are decided according to the respective semantic types.

Further, while determining the centers initially in the second step of k -means algorithm, we biased centers, so that first one-fourth documents in the ranked list go into the first cluster, the next one-fourth in the second, and so on.

The cluster that contains the top-ranked document is suggested to the user as the cluster most relevant to the query.

2.7. Displaying the results

The documents are finally displayed in descending order of weights. The most relevant sentences are highlighted. Thus the user effort required to locate the answer is minimized.

3. EVALUATION

For the sake of evaluating our system, we used the standard OHSUMED collection which is used extensively in information retrieval research.

3.1. About OHSUMED collection

The OHSUMED test collection [21] was created to assist information retrieval research. It is a clinically-oriented Medline subset, consisting of 348,566 references (out of a total of over 7 million), covering all references from 270 medical journals over a five-year period (1987–1991). The collection includes 106 queries generated using Medline by novice physicians. It also includes 12,565 unique query-reference pairs obtained after judgment for relevance. We used a subset of around 7000 documents from this collection as the document repository and the 101 queries as the questions for INDOC. Five queries were left out as our subset of documents did not contain an answer for them.

3.2. Performance evaluation and results

To evaluate our system, we compare the results returned by our system with the query-document pairs that have been judged for relevance. The OHSUMED collection includes the file drel.i that contains the query-document pairs rated as definitely relevant, with documents listed by sequential number in the format (<query><tab><document-i>). Corresponding to each query, we select the set of documents judged as definitely relevant as the set of correct documents and evaluate our results against this set. We illustrate the results in Table 2.

We observed that 58.4% of the questions posed were answered correctly by the first document itself. We also noted that the top 5 ranked documents have answers to 76.23% of all the queries.

Table 2 illustrates cumulative percentage of the queries answered, against the rank of documents.

For example for 81.18% of the queries, the first relevant result was obtained within top 10 results.

In total, we used 6637 documents and the system was able to answer 93.07% of the queries posed. No answer could be retrieved for 7 questions.

On an average, 54.79% of relevant documents were correctly identified by the system (Recall).

TABLE 2: Experimental results of our system on OHSUMED dataset.

Rank of first answer	Number of queries	% answered correctly
1	59	58.4
2	70	69.3
3	75	74.2
4	76	75.24
5	77	76.23
10	82	81.18
50	84	83.17

4. CONCLUSIONS AND FUTURE WORK

In this paper, we presented an effective implementation of a biomedical question answering system. We devised methods for query processing, document indexing and procedures for extracting the answer to the questions posed. The system was evaluated against the standard OHSUMED test collection and high performance (93.07% correctly answered, out of which 76.23% were answered within the top 5 documents) was obtained. We minimized the user effort by clustering the result set, identifying the most relevant sentences, and highlighting them. The technique and system presented in this paper can be useful in designing a new generation efficient framework for biomedical question answering system.

Apart from the ideas presented in this paper, there are some improvements possible on the present system. First the question's taxonomy as given in [24] can be implemented. Questions about patient care can be organized into a limited number of generic types, which could help guide the efforts of knowledge base developers. These generic types can be used in finding excerpts from the documents as short answers to the questions posed.

Secondly, the system relies on effective generation of heading concepts for each subsection as described in the proposed algorithm. From the algorithm, it is clear that any anaphora in sentences referring to potential heading concepts are not taken care of and they have to be dealt with to ensure effective indexing. As such, anaphora resolution is by large an unsolved problem. Addressing the problem of resolving Anaphora problem can be a potential area for future work.

REFERENCES

- [1] <http://www.ncbi.nlm.nih.gov/>.
- [2] P. Gorman, J. Ash, and L. Wykoff, "Can primary care physicians' questions be answered using the medical journal literature?" *Bulletin of the Medical Library Association*, vol. 82, no. 2, pp. 140–146, 1994.
- [3] S. E. Straus and D. L. Sackett, "Bringing evidence to the point of care," *Journal of the American Medical Association*, vol. 281, pp. 1171–1172, 1999.
- [4] G. H. Guyatt, M. O. Meade, R. Z. Jaeschke, D. J. Cook, and R. B. Haynes, "Practitioners of evidence based care," *British Medical Journal*, vol. 320, no. 7240, pp. 954–955, 2000.
- [5] D. L. Sackett, S. E. Straus, W. S. Richardson, W. Rosenberg, and R. B. Haynes, *Evidence-Based Medicine: How to Practice*

- and Teach ENB, Churchill Livingstone, New York, NY, USA, 1997.
- [6] P. N. Gorman and M. Helfand, "Information seeking in primary care: how physicians choose which clinical questions to pursue and which to leave unanswered," *Medical Decision Making*, vol. 15, no. 2, pp. 113–119, 1995.
 - [7] P. Jacquemart and P. Zweigenbaum, "Towards a medical question-answering system: a feasibility study," in *Proceedings of Medical Informatics Europe (MIE '03)*, P. L. Beux and R. Baud, Eds., vol. 95 of *Studies in Health Technology and Informatics*, pp. 463–468, IOS Press, San Palo, Calif, USA, 2003.
 - [8] S. Schultz, M. Honeck, and H. Hahn, "Biomedical text retrieval in languages with complex morphology," in *Proceedings of the Workshop on Natural Language Processing in the Biomedical domain*, pp. 61–68, Philadelphia, Pa, USA, July 2002.
 - [9] J. Ely, J. A. Osherooff, and M. H. Ebell, "Analysis of questions asked by family doctors regarding patient care," *British Medical Journal*, vol. 319, no. 7206, pp. 358–361, 1999.
 - [10] J. W. Ely, J. A. Osherooff, M. H. Ebell, et al., "Obstacles to answering doctors' questions about patient care with evidence: qualitative study," *British Medical Journal*, vol. 324, no. 7339, pp. 710–713, 2002.
 - [11] G. R. Bergus, C. S. Randall, S. D. Sinift, and D. M. Rosenthal, "Does the structure of clinical questions affect the outcome of curbside consultations with specialty colleagues?" *Archives of Family Medicine*, vol. 9, no. 6, pp. 541–547, 2000.
 - [12] Y. Niu and G. Hirst, "Analysis of semantic classes in medical text for question answering," in *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics, Workshop on Question Answering in Restricted Domains*, pp. 54–61, Barcelona, Spain, July 2004.
 - [13] Y. Niu, G. Hirst, G. McArthur, and P. Rodriguez-Gianolli, "Answering clinical questions with role identification," in *Proceedings of 41st Annual Meeting of the Association for Computational Linguistics, Workshop on Natural Language Processing in Biomedicine*, pp. 73–80, Sapporo, Japan, July 2003.
 - [14] E. T. K. Sang, G. Bouma, and M. De Rijke, "Developing offline strategies for answering medical questions," in *Proceedings of the AAAI-05 Workshop on Question Answering in Restricted Domains*, vol. WS-05-10, pp. 41–45, Pittsburgh, Pa, USA, 2005.
 - [15] A. M. Cohen and W. R. Hersh, "A survey of current work in biomedical text mining," *Briefings in Bioinformatics*, vol. 6, no. 1, pp. 57–71, 2005.
 - [16] <http://www.nlm.nih.gov/research/umls/>.
 - [17] <http://mmtx.nlm.nih.gov/>.
 - [18] A. R. Aronson, "Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program," in *Proceedings of the AMIA Symposium*, pp. 17–21, 2001.
 - [19] A. T. McCray, A. Burgun, and O. Bodenreider, "Aggregating UMLS semantic types for reducing conceptual complexity," *Medinfo*, vol. 10, part 1, pp. 216–220, 2001.
 - [20] O. Bodenreider and A. T. McCray, "Exploring semantic groups through visual approaches," *Journal of Biomedical Informatics*, vol. 36, no. 6, pp. 414–432, 2003.
 - [21] W. R. Hersh, "OHSUMED: an interactive retrieval evaluation and new large test collection for research," in *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '94)*, pp. 192–201, Springer, Dublin, Ireland, July 1994.
 - [22] <http://www.cdc.gov/nchs/about/otheract/icd9/abtcd9.htm>.
 - [23] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of 5th the Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281–297, University of California Press, Berkeley, Calif, USA, June-July 1967.
 - [24] J. W. Ely, J. A. Osherooff, P. N. Gorman, et al., "A taxonomy of generic clinical questions: classification study," *British Medical Journal*, vol. 321, no. 7258, pp. 429–432, 2000.