On Iterative Intelligent Medical Search

Gang Luo   Chunqiang Tang

IBM T.J. Watson Research Center, 19 skyline drive, Hawthorne, NY 10532, USA

{luog, ctang}@us.ibm.com

ABSTRACT

Searching for medical information on the Web has become highly popular, but it remains a challenging task because searchers are often uncertain about their exact medical situations and unfamiliar with medical terminology. To address this challenge, we have built an intelligent medical Web search engine called iMed, which uses medical knowledge and an interactive questionnaire to help searchers form queries. This paper focuses on iMed's iterative search advisor, which integrates medical and linguistic knowledge to help searchers improve search results iteratively. Such an iterative process is common for general Web search, and especially crucial for medical Web search, because searchers often miss desired search results due to their limited medical knowledge and the task's inherent difficulty. iMed's iterative search advisor helps the searcher in several ways. First, relevant symptoms and signs are automatically suggested based on the searcher's description of his situation. Second, instead of taking for granted the searcher's answers to the questions, iMed ranks and recommends alternative answers according to their likelihoods of being the correct answers. Third, related MeSH medical phrases are suggested to help the searcher refine his situation description. We demonstrate the effectiveness of iMed's iterative search advisor by evaluating it using real medical case records and USMLE medical exam questions.

Categories and Subject Descriptors

H.3.3 [Information Search and Retrieval]: search process, selection process

General Terms: Algorithms, Experimentation, Human Factors

Keywords: medical knowledge, medical query, intelligent medical Web search engine, iterative search process, language model

1. INTRODUCTION

As the healthcare industry is moving toward a more consumer-centric focus, searching for medical information on the Web has become highly popular. On an average day, 6% of American Internet users perform medical search [14] to better prepare for doctors' appointments and to better digest information obtained from doctors afterwards. Due to the increasing lack of new doctors, the interaction time between doctors and patients keeps shrinking, and this trend is expected to last in the foreseeable near future.

Since October 2005, several medical Web search engines have been launched, including Healthline [6], Google Health, SearchMedica, and Medstory [10]. They use the traditional keyword query interface, which works well when the searcher clearly knows his medical situation. For instance, a searcher knows that he has

high cholesterol and wants to learn about appropriate diet for himself. However, in many cases, the medical information searcher is uncertain about the problem he is facing and unaware of the related medical terminology (e.g., panophthalmitis). As a result, it is often difficult for him to choose a few accurate medical phrases as a starting point for his search.

To address this problem, we have built a prototype *i*ntelligent *medi*cal search engine called iMed [7], which uses medical knowledge and an interactive questionnaire to help searchers form queries. Below we first give a brief overview of iMed, and then focus on iMed's *iterative search advisor*, which integrates medical and linguistic knowledge to help searchers improve search results through iterative search.

Iterative search is fundamental to medical search because of medical problems' inherent fuzziness, which often makes it difficult even for medical professionals to distinguish between right and wrong choices. As reported by González-González et al. [5], partly due to the difficulty of identifying appropriate keywords that can clearly describe the medical situation, it takes a physician on average 30 minutes to search for answers to a clinical question and the success rate is only about 75%. For ordinary Internet users with little medical knowledge, we expect their medical search performance to be even worse.

1.1 Overview of iMed

iMed leverages its built-in medical knowledge in the form of diagnostic decision trees written by medical professionals [3]. As shown in Figure 1, each diagnostic decision tree corresponds to either a subjective symptom (e.g., fatigue) or an objective sign (e.g., hypertension). Each non-leaf, non-root node of a diagnostic decision tree corresponds to an answer to a question that iMed can ask. Each medical phrase in the leaf node of a diagnostic decision tree (possibly in combination with the searcher's other keyword inputs) can become a query that iMed uses.
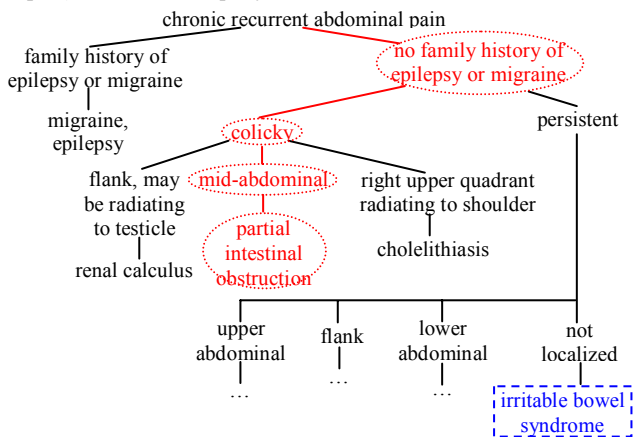

Figure 1. The diagnostic decision tree for the symptom "chronic recurrent abdominal pain."

iMed uses diagnostic decision trees to help the searcher form queries. The searcher first selects one or more symptoms and signs from a list of known symptoms and signs [3]. This list covers most chief complaints with which physicians are confronted. Then iMed asks questions related to these selected symptoms and signs. Based on the searcher's answers to the questions, iMed navigates the corresponding diagnostic decision trees and automatically forms multiple queries. Each query is used to retrieve some related Web pages. iMed's search results include the Web pages retrieved for all these queries.

For example, Figure 1 shows the diagnostic decision tree in Collins [3] for the symptom "chronic recurrent abdominal pain." If "chronic recurrent abdominal pain" is the only symptom chosen by the searcher, iMed's first question is "Is there a family history of epilepsy or migraine?" If the searcher answers "no" to this question, iMed's next question is "Is the pain colicky or persistent?" If the searcher answers "colicky" to the second question, iMed continues to ask "What is the location of the pain?" If the searcher's answer to the third question is "mid-abdominal," iMed forms multiple queries including partial intestinal obstruction. (Medical phrases in the non-selected leaf nodes of the diagnostic decision tree also form queries, but with lower weights. See Luo [7] for details.)

1.2 Iterative Search Advisor

We studied how users interact with an intelligent medical search engine that leverages built-in medical knowledge, and observed significant improvement over general Web search engines. However, we also observed that searchers still frequently miss desired information due to several reasons:

(1) In a typical medical case, the patient often has multiple symptoms and signs. The searcher frequently could not nail down the correct chief complaint (i.e., the most important symptom or sign), which is a task that usually requires rigorous medical training and lots of hands-on experience. As a result, the searcher started with some inappropriate symptoms and signs, whose corresponding diagnostic decision trees do not cover the patient's disease.

(2) Even if the searcher started with the right symptoms and signs, he sometimes provided improper answers to questions asked by the search engine due to misunderstanding of the questions and his medical situation. This leads to irrelevant search results.

(3) When the searcher could not find desired search results in a single pass, he usually resorted to iterative search. After reading the returned search results, the searcher might realize his inappropriate choices, correct them, and redo the search. However, a frequently occurring scenario is that the searcher could not realize the inappropriate choices he has made and was stuck there.

In response to these challenges, we recently developed an iterative search advisor (ISA) in iMed, which automatically suggests symptoms and signs and recommends alternative answers to the questions. Such functionalities are highly desired by both ordinary Internet users and medical professionals [13].

In iMed, whenever the searcher needs help, he inputs a description of his medical situation into the ISA. iMed then does three things in response. First, iMed sorts all the symptoms and signs in its knowledge base in descending order of their relevance to the searcher's situation description, and suggests the top ranked symptoms and signs to the searcher. Second, for all the questions asked by iMed, alternative answers that are not selected by the searcher are ranked according to their likelihoods of being the correct answers and recommended to the searcher. Third, considering that the searcher often cannot accurately describe his medical situation in a single pass, iMed suggests related medical phrases in the MeSH medical ontology [9] to help the searcher refine his situation description. All these medical phrases are ordered by their relevance to the searcher's situation. During an iterative search, the above three features help the searcher select different symptoms and signs, provide alternative answers to the questions asked by iMed, and revise his situation description. Whenever the searcher's situation description is modified, the suggestions provided by iMed's ISA may also change.

In essence, our iterative search method combines automatic text analysis with human intelligence to cope with the inherent difficulty of the medical search problem and the limitations of existing natural language processing techniques. iMed provides useful suggestions to the searcher while the suggestions can be imperfect. We ask the searcher to use his intelligence to filter out inappropriate suggestions. This suggestion process repeats until the searcher finds his desired information, and human intelligence is involved in every iteration of this process.

1.3 Technical Challenges and Our Solutions

The fundamental challenge in building iMed's ISA is to construct a statistical model for the iterative search task that can seamlessly integrate information from natural language text description with information from the diagnostic decision trees. We combine the language modeling method [12] with several novel techniques to address this challenge. More specifically, we address the following problems.

First, when suggesting symptoms/signs and recommending medical phrases related to the searcher's situation, there is a vocabulary mismatch between the medical phrases and the searcher's situation description in layman terms. To address this problem, we use the representative page technique [7, 8] to "translate" each such phrase into multiple representative Web pages. The searcher's situation description is matched with the representative Web pages instead of with the original medical phrases.

Second, when ranking alternative answers to the questions, there is another vocabulary mismatch between the alternative answers and the searcher's situation description in layman terms. This problem cannot be solved by directly using the representative page technique for the alternative answers. This is because many answers to the questions are about the absence of certain symptoms, such as "little or no sputum" and "no fever." Existing keyword matching techniques cannot retrieve appropriate representative Web pages for these answers. Instead, we exploit the semantics of diagnostic decision trees to obtain an indirect representation of the alternative answers to the questions. Recall that each answer A to a question Q corresponds to a non-leaf, non-root node N_A in a diagnostic decision tree T. All the diseases in T's leaf nodes that are N_A's descendants form an indirect representation of A. We use the indirect representations of the alternative answers, instead of the alternative answers themselves, to obtain representative Web pages. We then match these pages with the searcher's situation description.

We crawled a large number of medical Web pages from the Internet and evaluated the effectiveness of our techniques using both real medical case records and real medical exam questions. Our results show that iMed's ISA significantly improves user satisfaction. Many suggestions are useful and can facilitate searchers to quickly find their desired information.

The rest of the paper is organized as follows. Section 2 shows two medical case examples. Section 3 describes the user interface of

iMed's new ISA. Section 4 presents our techniques for making suggestions. Section 5 evaluates the effectiveness of our techniques.

2. MEDICAL CASE EXAMPLES

In this section, we present two example medical case records from the Family Medicine Online Database [4] to give the reader a feeling of real medical cases. Some searchers failed to use iMed to find desired information for these two medical cases.

Example 1

Figure 2 shows the first example medical case record, whose correct diagnosis is irritable bowel syndrome (IBS). As can be seen in Figure 1, the correct diagnosis IBS can only be found by answering "persistent" to the question Q, "Is the pain colicky or persistent?". When using iMed, some users did not know the word "colicky" and consulted the explanation that iMed provided for Q — a colicky pain is a gradual onset of pain that increases in a crescendo fashion until it reaches a peak of severity and then slowly subsides. According to the medical record, the patient's pain is intermittent. Some users treated intermittent as colicky rather than persistent, and hence answered "colicky" to Q. Consequently, they found the disease "partial intestinal obstruction" but missed the correct diagnosis IBS, which is ranked rather low in the returned search results.

A 45-year-old white female presents to your outpatient office with complaints of abdominal pain. Her pain is intermittent and accompanied by bouts of diarrhea. Between bouts of pain with diarrhea (which predominate) she is constipated with small hard stools. She describes the pain as crampy without localization. She has not lost or gained weight but constantly feels bloated. Dairy products and caffeine exacerbate her diarrheal symptoms. Her diet includes limited fiber intake.

Figure 2. A first example medical case record (www.hmc.psu.edu/ume/fcmonline/case20/index.htm).

Example 2

Figure 3 shows the second example medical case record, whose correct diagnosis is endometriosis. According to the medical record, the patient's chief complaint is vaginal bleeding. When using iMed, some users chose the symptom "metrorrhagia," which means uterine bleeding at irregular intervals, particularly between the expected menstrual periods. However, endometriosis is not in the diagnostic decision tree for the symptom "metrorrhagia" [3]. Therefore, those users could not find this correct diagnosis. In fact, endometriosis is in the diagnostic decision tree for the symptom "dyspareunia," which means having difficulty during sexual intercourse. Dyspareunia matches with the patient's situation of having lower abdominal cramping and pain with deep pelvic thrusts during intercourse. If a user chooses this symptom, he is likely to find the correct diagnosis.

Ms. Landry is a 30-year-old Caucasian female. She complains of a three-month history of intermittent vaginal bleeding throughout her menstrual cycle that is increased during menstrual periods. During this same period, she notes lower abdominal cramping and pain with deep pelvic thrusts during intercourse. She admits to having occasional post-coital bleeding. She denies any abnormal vaginal discharges, nausea, appetite changes, or breast tenderness. She also denies increased frequency, urgency, or changes in bowel habits. She denies weight loss, excessive exercise, or increased stress. Her menstrual periods typically last four days but recently they have persisted for as long as two weeks. She notes mild to moderate dysmenorrhea although not every cycle. She has had no known pregnancies but has used various methods of contraception in the past.

Figure 3. A second example medical case record (www.hmc.psu.edu/ume/fcmonline/case21/index.htm).

3. USER INTERFACE

iMed's novel user interface is an important contribution of the overall work. In iMed, the medical information searcher performs search through a questionnaire-based interface. He selects symptoms and signs, answers questions asked by iMed, and then obtains search result pages. Whenever the searcher cannot find desired information and needs help, he can click a button appearing on each page provided by iMed and invokes iMed's ISA. The searcher can invoke the advisor at the very beginning of the search process if he does not even know which symptoms and signs to begin with.

Figure 4. The output interface of iMed's iterative search advisor.

iMed's ISA has an input interface and an output interface. The input interface is a multi-line text area. The searcher inputs a description of his medical situation into this text area and clicks the submit button. iMed then makes suggestions on three aspects: (1) symptoms and signs, (2) alternative answers to the questions, and (3) related medical phrases for refining this situation description. Accordingly, the output interface shown in Figure 4 is divided into three segments, one for each aspect. As in Medstory [10], for each suggestion g iMed makes in a segment s, there is a corresponding horizontal bar reflecting g's strength. The length of the bar is proportional to $\max\{c_s + \ln p(g), 0\}$, where c_s is a constant specific

to s, and $p(g)$ is the estimated probability that g is related to the searcher's situation. The computation formulas of $p(g)$'s are described in Section 4. The searcher can invoke iMed's ISA repeatedly. In each iteration, he works in only one of the three segments and clicks the corresponding submit button there, while any changes made in one segment can affect the contents in the other two segments.

3.1 Symptom and Sign Segment

The first segment of the output interface is dedicated to symptoms and signs. Those symptoms and signs chosen by the searcher are shown at the beginning of this segment. All the other symptoms and signs covered in the questionnaire are sorted in descending order of their relevance to the searcher's situation description. The top $N_s=30$ (the subscript s stands for symptom/sign) symptoms and signs are listed in the first segment under the title "alternative symptoms and signs." When the searcher moves the mouse to a symptom or sign S whose name is unfamiliar to ordinary users, the annotation of S in layman terms is displayed automatically. For example, the symptom "hemoptysis" is annotated with "coughing up blood." This helps the searcher understand these suggested symptoms and signs. The searcher can drop the previously chosen symptoms and signs, and/or add the symptoms and signs suggested by iMed. After the searcher clicks the submit button in the first segment, iMed will redo the search by asking questions and returning result pages. In a subsequent run, if a question Q has been asked before, iMed will reuse the searcher's previous answer to Q without asking Q again.

3.2 Question Answer Segment

The second segment of the output interface is devoted to answers to the questions asked by iMed. In general, the search process is performed iteratively. In each iteration, the searcher can revise his answers to some old questions that iMed asked in previous iterations. In addition, the searcher can answer new questions that iMed has not asked before. Essentially, iMed guides the searcher to navigate the diagnostic decision trees. The more iterations the search process has been performed, the more branches the searcher has visited in the diagnostic decision trees. It is desirable to show the searcher succinctly both the branches that he has visited and the branches that he has not visited. This helps the searcher avoid either unnecessarily visiting the same branch multiple times or forgetting visiting some important branches. On the other hand, it is undesirable to overwhelm the searcher with too much information. For example, after a few iterations of the search process, the searcher has provided many answers to the questions asked by iMed. In this case, iMed should not display these (many) answers to the searcher literally, e.g., one answer in a separate line. To strike a proper balance, we design a novel user interface called *question-answer trees*.

One naive user interface is to display the selected diagnostic decision trees and to mark the searcher's answers to the questions directly on them. However, the entire diagnostic decision tree T_S for a symptom or sign S can be too complex and contain much information that is unnecessary to the searcher. Instead, we use a question-answer tree B_S that contains only the most essential portion of a modified version of T_S, i.e., the searcher's answers to the questions related to S and the corresponding alternative answers. More specifically, for each symptom or sign S chosen by the searcher, a corresponding question-answer tree B_S is displayed at the beginning of the second segment of the output interface. S is shown

at the top of B_S in italics so that the searcher can easily see which question-answer tree corresponds to which symptom or sign. For all the symptoms and signs chosen by the searcher, their corresponding question-answer trees are displayed shoulder-to-shoulder.

In a question-answer tree B_S, the non-leaf nodes correspond to all the questions C_q that are related to S and iMed has asked in the past. The solid-line edges correspond to all the answers that the searcher has provided to the questions in C_q in the past. For those edges corresponding to the answers that the searcher selected in the latest iteration, i.e., the searcher's current selections, their solid lines are thicker than the solid lines for the other edges. The dotted-line edges correspond to all the alternative answers that the searcher has not selected for the questions in C_q. In general, a question can have more than two candidate answers and hence multiple alternative answers can exist for the same question. In the case that the searcher chose to provide no answer to a question Q in C_q, all the candidate answers to Q are treated as alternative answers. Each dotted-line edge is connected to a leaf node, which is shown as a dotted circle with an embedded number. This number is the sequence number that the corresponding alternative answer A is ranked among all the alternative answers according to their likelihoods of being the correct answers. As shown in Figure 4, A is listed under the title "alternative answers to the questions" with an accompanying horizontal bar, whose length reflects the strength of suggesting A. The searcher can use this sequence number to find the corresponding information easily.

For example, consider the symptom "chronic recurrent abdominal pain" and its associated diagnostic decision tree in Figure 1. Suppose in the first iteration of the search process, the searcher provides the question answers that are mentioned in the paragraph before Figure 1. Then the corresponding question-answer tree is shown in Figure 4. In the second iteration of the search process, suppose the searcher answers "persistent" to the question "Is the pain colicky or persistent?" and answers "not localized" to the question "What is the location of the pain?". Then the corresponding question-answer tree is shown in Figure 5.
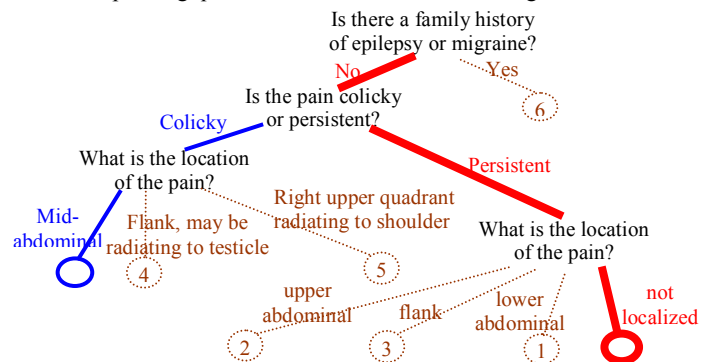

Figure 5. The question-answer tree displayed after the second iteration of the search process.

For all the alternative answers, iMed ranks them according to their likelihoods of being the correct answers and lists them in the second segment of the output interface under the title "alternative answers to the questions." The alternative answers for the same symptom or sign are shown in the same color, while the alternative answers for various symptoms and signs are shown in different colors. For each symptom or sign, the searcher can select at most one appropriate alternative answer. (The searcher can also choose answers by clicking the corresponding circles in the question-answer trees.) After the searcher clicks the submit button in the

second segment of the output interface, iMed will revise the searcher's answers to the questions accordingly and redo the search. For example, consider the question Q in Figure 5, "Is the pain colicky or persistent?". The searcher answered "colicky" to Q in the first iteration of the search process and answered "persistent" to Q in the second iteration. If the searcher selects the fifth alternative answer in Figure 5 in the third iteration, iMed will automatically think that the searcher answers "colicky" to Q and answers "right upper quadrant radiating to shoulder" to the question "What is the location of the pain?".

In general, for any symptom or sign S, a dominance relationship can exist among its alternative answers. Specifically, in the question-answer tree B_S for S, consider two alternative answers A_1 and A_2 whose corresponding questions reside in the non-leaf nodes N_1 and N_2, respectively. If N_1 is a descendant of N_2, then A_1 is dominated by A_2. Note that the answer to a question Q can affect the subsequent navigation path in the corresponding diagnostic decision tree and all the related questions that iMed will ask after Q. Consequently, in the second segment of the output interface, if the searcher feels that he has made inappropriate choices in answering multiple questions related to S and the corresponding several alternative answers D_A all look reasonable to him, he is advised to choose an appropriate one in D_A that is not dominated by any other one in D_A. For example, consider the six alternative answers in Figure 5. The first five alternative answers are dominated by the sixth alternative answer. Thus, if the searcher feels that both the second and the sixth alternative answers look reasonable to him, he should choose the sixth alternative answer.

3.3 Medical Phrase Segment

The third segment of the output interface is devoted to related medical phrases. To help the searcher refine his situation description, iMed uses the Medical Subject Headings (MeSH) ontology [9], a standard vocabulary edited by the National Library of Medicine and widely used for indexing and cataloging biomedical and health-related documents. The MeSH ontology is organized into a tree structure, whose branches correspond to different categories of medical phrases. iMed's ISA uses the medical phrases in the branches of the MeSH tree that correspond to categories C23 (pathological conditions, signs, and symptoms) and F01.145.126 (behavioral symptoms), as those phrases are most useful for refining the searcher's situation description. Moreover, those medical phrases include relatively minor conditions and are more comprehensive than the symptoms and signs covered in the questionnaire. Since each condition usually accompanies with some chief complaint that is one of the symptoms and signs covered in the questionnaire, no separate diagnostic decision trees are needed for these conditions, while these conditions are still useful for describing the searcher's situation.

The searcher's situation description is displayed at the beginning of the third segment of the output interface (see Figure 4). Following that, under the title "related medical phrases" are $N_p=50$ (the subscript p stands for phrase) medical phrases from the MeSH ontology. None of these medical phrases has appeared in the searcher's latest situation description. All these medical phrases are ordered by their relevance to the searcher's situation. When the searcher moves the mouse to a medical phrase M, the explanation of M that comes from the annotation field in the MeSH ontology is automatically displayed. This helps the searcher understand these suggested medical phrases. The searcher can use these phrases to refine his situation description displayed at the beginning of the

third segment. After rewriting his situation description, the searcher can click the submit button in the third segment and re-invoke iMed's ISA. With an improved situation description, it is likely that iMed will make more accurate suggestions.

4. ALGORITHMS FOR MAKING SUGGESTIONS

Given a medical situation description, iMed makes suggestions to the searcher in three ways:
(1) Suggesting relevant symptoms and signs.
(2) Ranking alternative answers to the questions.
(3) Recommending related MeSH medical phrases.
Below, we first provide some related background, and then present the details of our algorithm for each of the three functions above.

4.1 Background on Language Modeling

Language modeling [12] with Dirichlet smoothing [18] is a state-of-the-art method for ranking documents. Due to its superior performance and solid mathematical foundation, this method attracts much attention in recent years. We extend it to make suggestions in iMed's ISA.

Assuming that all the documents in a collection C have the same prior probability of being relevant to a query Q, the language modeling method with Dirichlet smoothing uses the following formulas to compute the conditional probability of a document $R \in C$ given Q:

$$p(R \mid Q) = p(Q \mid R)p(R)/p(Q) \propto p(Q \mid R) = \prod_{q \in Q} p(q \mid R),$$

$$p(q \mid R) = [c(q, R) + u \times c(q, C)/|C|]/(|R| + u).$$

Here, $c(q, R)$ is query term q's frequency in R, $c(q, C)$ is q's frequency in C, $|R|$ is the length of R in the number of terms, and $|C|$ is the length of C in the number of terms. u is a predetermined constant. Typically, as suggested in Zhai and Lafferty [18], $1000 \le u \le 10000$. The first formula uses Bayes' rule and assumes that all the query terms are independent of each other given R. The second formula uses a Dirichlet prior to avoid having zero probabilities. In iMed, we use $p(Q \mid R)$ as the building block for make suggestions.

4.2 Suggesting Symptoms and Signs

iMed's questionnaire currently covers 267 symptoms and signs. For all the symptoms and signs that are covered in the questionnaire but not selected by the searcher, iMed sorts them in descending order of their relevance to the searcher's situation description D_e and suggests the top $N_s=30$ ones to the searcher. The key step here is to compute $p(S \mid D_e)$, the conditional probability that a symptom or sign S is relevant to the searcher's situation given D_e. According to Bayes' rule:

$$p(S \mid D_e) = p(D_e \mid S)p(S)/p(D_e) \propto p(D_e \mid S)p(S).$$

We first discuss how to compute the conditional probability $p(D_e \mid S)$, i.e., if a searcher has symptom or sign S, how likely he will write D_e as his situation description. We cannot directly compute $p(D_e \mid S)$ due to the vocabulary mismatch between the symptoms/signs in medical terminology and the searcher's situation description D_e in layman terms. Even if some symptom or sign S matches with D_e semantically, it is not uncommon that S does not appear in D_e. To address this problem, we use the representative page technique in [7, 8] to "translate" each

symptom or sign into multiple representative Web pages. D_e is matched with these representative Web pages instead of directly with the symptoms and signs. Specifically, we have a predetermined constant r. For each symptom or sign S, we retrieve the top-ranked r Web pages R_i ($1 \leq i \leq r$) in the collection C of crawled Web pages and use them as the representative Web pages of S. To reduce noise, we keep in each R_i ($1 \leq i \leq r$) only those terms that appear close to S in the document text and have the largest tf×idf values. The representative Web pages for each symptom or sign are pre-processed offline and can be retrieved easily during a search. The conditional probability $p(D_e \mid S)$ is estimated using a weighted geometric mean of the conditional probabilities $p(D_e \mid R_i)$ ($1 \leq i \leq r$):

$$p(D_e \mid S) \leftarrow [\prod_{i=1}^{r} p(D_e \mid R_i)^{w_i}]^{1/\sum_{i=1}^{r} w_i}. \qquad (1)$$

Here, we use the symbol "$\leftarrow$" in (1) to emphasize that our estimation is an approximation under certain assumptions. We do not use the symbol "=" because $p(D_e \mid S)$ cannot be proven to be equal to the right side of (1). The weight for R_i is $w_i = 1/i$ ($1 \leq i \leq r$). It is a decreasing function of i and reflects the fact that higher-ranked Web pages are generally more important than lower-ranked ones. $p(D_e \mid R_i)$ is computed using the language modeling method described in Section 4.1.

Next, we describe how to compute the prior probability $p(S)$ that a symptom or sign S is relevant to the searcher's situation. In general, for each S, the following two kinds of information are useful for estimating $p(S)$. First, from iMed's log of user search sessions, we can compute the probability $p_1(S)$ that a searcher will select S. Second, $p(S)$ is related to the diseases that have S. Recall that S has a corresponding diagnostic decision tree T. All the diseases E_S mentioned in the leaf nodes of T have S. (Note that the leaf nodes of T can contain other medical phrases such as tests [3].) For each disease $d \in E_S$, its incidence rate $r(d)$ is the number of new cases per 1,000 people per year and reflects the probability of developing d [2]. Information about disease incidence rates is available from many sources, e.g., the US Centers for Disease Control and Prevention [1] and the WrongDiagnosis Web site [2]. We can compute the probability $p_2(S)$ that a person will have S using the sum of the incidence rates of all the diseases having S:

$$p_2(S) \propto \sum_{d \in E_S} r(d).$$

In iMed, we estimate the probability $p(S)$ using a weighted sum of $p_1(S)$ and $p_2(S)$: $p(S) \leftarrow v \times p_1(S) + (1-v) \times p_2(S)$, (2) where v is a predetermined constant. Here, we use the symbol "$\leftarrow$" in (2) to emphasize that our estimation is an approximation under certain assumptions. Neither $p_1(S)$ nor $p_2(S)$ is equal to $p(S)$:

(1) $p_1(S)$ is the probability that a searcher will select S. The searcher may not select S correctly.
(2) $p_2(S)$ is the probability that a person will have S. Not all the people with S will use iMed. Moreover, the proportion of people that have a symptom or sign and will use iMed varies for different symptoms and signs.

4.3 Ranking Alternative Answers

Consider the questions asked by iMed. For all the alternative answers to these questions that were not previously selected by the searcher, iMed ranks and suggests them according to their likelihoods of being the correct answers. The challenge is to compute $p(A \mid D_e)$, the conditional probability that an alternative answer A is the correct answer given the searcher's situation description D_e.

At a first thought, one might want to compute $p(A \mid D_e)$ using a method similar to that in Section 4.2. The procedure is as follows. According to Bayes' rule:

$$p(A \mid D_e) = p(D_e \mid A)p(A) / p(D_e) \propto p(D_e \mid A)p(A).$$

One first attempts to compute the conditional probability $p(D_e \mid A)$, i.e., if alternative answer A is the correct answer, how likely the searcher will write D_e as his situation description. Again, $p(D_e \mid A)$ cannot be directly computed due to the vocabulary mismatch between A and the searcher's situation description D_e in layman terms. Moreover, the representative Web pages of A often cannot be obtained directly. For example, many alternative answers are about the absence of certain symptoms. A Web page describing a disease that does not have symptom S can either mention the absence of S in many different ways or do not mention S at all. Consequently, existing keyword matching techniques cannot retrieve appropriate representative Web pages for these answers. Without representative pages, it is rather difficult to compute $p(D_e \mid A)$. Hence, one cannot easily compute $p(A \mid D_e)$ using $p(D_e \mid A)$.

To address this problem, we exploit the semantics of diagnostic decision trees to obtain an indirect representation of each alternative answer A and then use it to compute the conditional probability $p(A \mid D_e)$. Essentially, we build a statistical model that seamlessly integrates the information from natural language text description with the information from the diagnostic decision trees. A similar problem is frequently encountered in ontology-based information retrieval while no good solution exists.
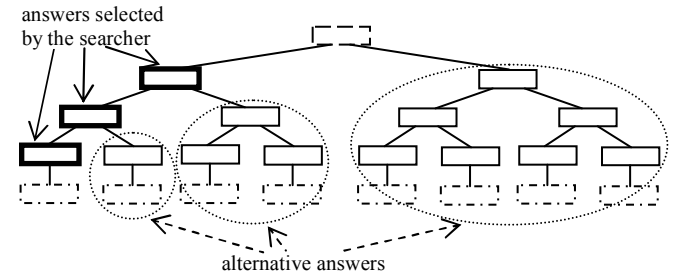
Figure 6. An illustration of the searcher's selected answers vs. the alternative answers in a diagnostic decision tree.

Our concrete method is as follows. According to the definition of conditional probability,

$$p(A \mid D_e) = p(A, D_e) / p(D_e) \propto p(A, D_e).$$

As illustrated in Figure 6, each alternative answer A to a question Q corresponds to a non-leaf, non-root node N_A in a diagnostic decision tree T. All the diseases E_A in T's leaf nodes that are N_A's descendants form an indirect representation of A. In the case that A is the correct answer and the information in T is complete, the searcher can only have diseases in the set E_A. Suppose the searcher does not have multiple diseases at the same time. Then event A is the same as the union of all the events d, $d \in E_A$. That is, $A = \bigcup_{d \in E_A} d$. Note that A is interpreted as both an event and an

alternative answer. Similarly, d is interpreted as both an event and a disease. Consequently,

$$p(A, D_e) = \sum_{d \in E_A} p(d, D_e) = \sum_{d \in E_A} p(D_e \mid d)p(d).$$

The conditional probability $p(D_e \mid d)$ can be computed using the representative page technique described in Section 4.2, as disease d is a medical phrase and its representative Web pages can be obtained easily [8]. The remaining problem is to compute the probability $p(d)$ that a searcher who correctly selects the symptom or sign associated with T will have disease d.

Not all the people with disease d will use iMed. Also, the proportion of people that have a disease and will use iMed varies for different diseases. Hence, $p(d)$ is generally not proportional to the incidence rate $r(d)$ of d. We exploit the fact that d is contained in a leaf node N_d of the diagnostic decision tree T and estimate $p(d)$ in two steps. First, from iMed's log of user search sessions, we can compute the probability $p(N_d)$ that a searcher will select N_d. Second, N_d generally contains multiple diseases $d_1, d_2, \ldots,$ and d_k, where d is one of them. We assume that each disease d_i $(1 \le i \le k)$ obtains its share of $p(N_d)$ in proportion to its incidence rate $r(d_i)$. Then we estimate $p(d)$ as

$$p(d) \leftarrow p(N_d) \cdot r(d) / \sum_{i=1}^{k} r(d_i). \qquad (3)$$

Here, we use the symbol "$\leftarrow$" in (3) to emphasize that our estimation is an approximation under certain assumptions.

4.4 Recommending Related Medical Phrases

iMed's ISA only recommends the medical phrases in the branches of the MeSH tree that correspond to categories C23 and F01.145.126. As mentioned in Section 3.3, these MeSH phrases are most useful for refining the searcher's situation description. For all those MeSH phrases that are not in the searcher's situation description D_e, iMed sorts them in descending order of their relevance to the searcher's situation and recommends the top $N_p=50$ ones to the searcher. The key step here is to compute $p(M \mid D_e)$, the conditional probability that a MeSH phrase M is related to the searcher's situation given D_e. Assuming that all MeSH phrases have the same prior probability of being relevant to the searcher's situation, Bayes' rule shows that:

$$p(M \mid D_e) = p(D_e \mid M)p(M) / p(D_e) \propto p(D_e \mid M).$$

The conditional probability $p(D_e \mid M)$ can be computed using the representative page technique described in Section 4.2, as representative Web pages can be easily retrieved for the medical phrase M [8].

5. EXPERIMENTAL RESULTS

We conducted experiments with various medical cases to demonstrate the effectiveness of iMed's ISA.

5.1 Setup

iMed is a vertical search engine that crawls Web pages from a few selected, high-quality medical Web sites instead of all the Web sites. In our experiments, we crawled 22GB of Web pages from WebMD [17], Healthline [6], and Merck [11], three of the most popular medical Web sites. We used both real medical case records from the Family Medicine Online Database (FMOD) [4] and USMLE medical exam questions [16]. Correct diagnoses are available for both of them and serve as the ground truth for our evaluation. USMLE stands for the United States Medical Licensing Examination, whose exam question format is similar to the format of actual, well-documented medical case records. Physicians have to pass this exam to obtain their licenses for practicing medicine. In our tests, each exam question is treated as a medical case. FMOD was developed by the College of Medicine of the Pennsylvania State University for educating medical students. The FMOD records document patients' medical situations in great detail using mostly layman terms and can be easily understood by ordinary people.

Ten colleagues served as assessors and independently evaluated iMed. None of them has formal medical training. They performed search based on medical cases randomly selected from the FMOD records and the USMLE exam questions. For each user, we use 30 random medical cases that he cannot find the correct answers in a single pass. (We exclude from our evaluation those medical cases that the user can find the correct answers in a single pass, because he never invokes iMed's ISA for those cases.) Two such medical cases were shown earlier in Figures 2 and 3 in Section 2. Since both USMLE and FMOD cover almost every aspect of medical practice, our random samples have a broad coverage of medical topics. For each medical case, we randomly divided all ten users into two groups of the same size. When performing search, one group used iMed with the ISA while the other group used iMed without the ISA.

Currently, iMed is a research prototype. We have limited logs of user search sessions and are still in the process of collecting disease incidence rates from various sources [1, 2], merging them, and matching them with the diseases in the diagnostic decision trees [3]. (Note that different sources can use various names for the same disease, and a general disease name can cover several more specific diseases.) Due to these limitations, the current iMed prototype uses uniform prior probabilities at several places. iMed is still effective with these priors, but would provide more accurate recommendations if ideal priors are used. More precisely, our experiments assume that all the diseases have the same incidence rate. The prior probability that a searcher will select a symptom or sign is the same for all the symptoms and signs covered in iMed's questionnaire. For a question asked by iMed, a searcher will select all the candidate answers with equal likelihood.

In our experiments, a user has up to 60 minutes to perform iterative search for each medical case. At the end of the search process, the user can list up to three diseases that he thinks best match the medical case's situation description. If any of these diseases is among the correct diagnoses accompanying the data set, the search is considered successful. We allow users to search for a relatively long time, because medical information searchers care about their health and often spend hours on searching. We allow users to list multiple diseases as their findings, because even doctors sometimes cannot make precise diagnosis without lab test results.

Similar to the TREC interactive track [15], we use two sets of measures as the performance metrics for medical search engines: one set is objective while the other set is subjective. The objective performance measures include the success rate, the number of search iterations, the number of search result Web pages viewed, and the time spent on the search process. The subjective performance measures include the users' perceptions of ease of using the system, ease of understanding the system, and overall satisfaction with the system. All these subjective performance measures are on a 7-point scale, with 1=low and 7=high [15]. They were obtained from a brief questionnaire that users filled out after using the systems. For each objective or subjective performance measure, we average it over all the 30 medical cases and all the users, and report both its mean and its standard deviation when

appropriate. We used ANOVA as the significance test. Our experiments were performed on a computer with two 3GHz processors, 2GB memory, and one 111GB disk.

5.2 Overall Results

iMed's ISA is efficient. On average, it takes less than two seconds to make suggestions in one iteration of the search process. iMed's recommendations are also very helpful for searchers to find the correct diagnosis, where most of the searchers' time is spent on reading the search result Web pages. The objective performance measures in Table 1 show that the ISA helps the user find results in fewer iterations, view fewer search result Web pages, spend less time on the search process, and achieve a higher success rate. All these differences are statistically significant.

Table 1. Objective performance measures (* means significant at <0.05 level).

mean (standard deviation)	without advisor	with advisor
success rate	25% (8%)	34%* (10%)
number of iterations	5.7 (1.5)	4.2* (1.4)
number of search result Web pages viewed	20 (7)	15* (6)
time (minutes)	45 (13)	33* (11)

Table 2. Subjective performance measures (* means significant at <0.05 level).

mean (standard deviation)	without advisor	with advisor
ease of using	4.7 (1.1)	5.5* (1.1)
ease of understanding	5.8 (1.1)	5.5 (1.2)
satisfaction	4.9 (0.9)	5.7* (1.0)

Table 2 shows the subjective performance measures. As it takes time to become familiar with the ISA, users consider the iMed without the ISA slightly easier to understand, while the difference is not statistically significant. Nevertheless, once users understand the ISA, they can use it without difficulty. With the ISA, users can obtain much help and thus feel that iMed becomes easier to use. Overall, the ISA greatly improves user satisfaction as it helps produce more useful search results with less user effort. These differences caused by the ISA are statistically significant.

5.3 Two Detailed Examples

To give the reader a feeling of the suggestions made by iMed, we present detailed results of iMed's suggestions for the two example medical cases shown in Figures 2 and 3.

The first example medical case is about a female complaining abdominal pain. For this medical case, one user writes "intermittent abdominal pain crampy diarrhea constipation" as the description of the patient's situation. Figure 4 shows the corresponding suggestions provided by iMed. The top six symptoms and signs suggested by iMed are focal abdominal swelling, generalized abdominal swelling, rectal bleeding, rectal pain, vulval vaginal ulcerations, and vulval vaginal mass. Depending on the specifics of the medical case, they can all be relevant to the patient's situation. The first recommended alternative answer is "persistent" to the question "Is the pain colicky or persistent?". As mentioned in Section 2, the user can find the correct diagnosis irritable bowel syndrome by choosing this alternative answer.

The top five medical phrases suggested by iMed are dyspepsia, torsades de pointes, cutaneous fistula, vaginal fistula, and postoperative complications. Dyspepsia is chronic or recurrent pain

or discomfort centered in the upper abdomen. A cutaneous fistula is an abnormal passage or communication leading from an internal organ to the surface of the body. A vaginal fistula is an abnormal passage that connects the vagina to other organs. Postoperative complications can cause abdominal pain. Depending on the specifics of the medical case, the first, third, fourth, and fifth suggested medical phrases can be relevant to the patient's situation.

The second example medical case is about a female complaining vaginal bleeding. For this medical case, one user writes "vaginal bleeding pelvic pain" as the description of the patient's situation. The first symptom or sign suggested by iMed is dyspareunia. As mentioned in Section 2, the user can find the correct diagnosis endometriosis by choosing this symptom.

The top five medical phrases suggested by iMed are prolapse, hyperplasia, dysmenorrhea, hemolysis, and uterine hemorrhage. Prolapse means that an organ, such as the uterus, falls down or slips out of place. Dysmenorrhea represents painful menstruation. Uterine hemorrhage is the same as uterine bleeding. Depending on the specifics of the medical case, the first, third, and fifth suggested medical phrases can be relevant to the patient's situation.

In general, for a medical case, iMed can make several useful suggestions on the relevant symptoms and signs, alternative answers to the questions asked by iMed, and related MeSH medical phrases.

6. ACKNOWLEDGEMENTS

7. REFERENCES

[1] Data & Statistics, Centers for Disease Control and Prevention. http://www.cdc.gov/DataStatistics, 2007.
[2] Conditions by Incidence. http://www.wrongdiagnosis.com/lists/incid.htm, 2007.
[3] R.D. Collins. Algorithmic Diagnosis of Symptoms and Signs: Cost-Effective Approach. Lippincott Williams & Wilkins, 2002.
[4] Family Medicine Online homepage. http://www.hmc.psu.edu/ume/fcmonline/index.htm, 2007.
[5] A.I. González-González, M. Dawes, and J. Sánchez-Mateos et al. Information Needs and Information-Seeking Behavior of Primary Care Physicians. Annals of Family Medicine 5: 345-352, 2007.
[6] Healthline homepage. http://www.healthline.com, 2007.
[7] G. Luo. iMed: An Intelligent Medical Web Search Engine. Available at http://pages.cs.wisc.edu/~gangluo/imed.pdf, 2008.
[8] G. Luo, C. Tang, and H. Yang et al. MedSearch: A Specialized Search Engine for Medical Information. WWW 2007: 1175-1176.
[9] MeSH homepage. www.nlm.nih.gov/mesh/meshhome.html, 2007.
[10] Medstory homepage. http://www.medstory.com, 2007.
[11] Merck Manual Home Edition homepage. http://www.merck.com/mmhe/index.html, 2007.
[12] J.M. Ponte, B.W. Croft. A Language Modeling Approach to Information Retrieval. SIGIR 1998: 275-281.
[13] P. Ramnarayan, A. Tomlinson, and G. Kulkarni et al. A Novel Diagnostic Aid (ISABEL): Development and Preliminary Evaluation of Clinical Performance. Medinfo 2004: 1091-1095.
[14] C. Sherman. Curing Medical Information Disorder. http://searchenginewatch.com/showPage.html?page=3556491, 2005.
[15] TREC interactive track homepage. http://trec.nist.gov/data/interactive.html, 2007.
[16] USMLE homepage. http://www.usmle.org, 2007.
[17] WebMD homepage. http://www.webmd.com, 2007.
[18] C. Zhai, J.D. Lafferty. A Study of Smoothing Methods for Language Models Applied to Ad Hoc Information Retrieval. SIGIR 2001: 334-342.