Restricted Natural Language Processing for Case Simulation Tools

Christoph U. Lehmann, M.D.* ¶, Bach Nguyen¶, George R. Kim, M.D.*,
Kevin B. Johnson, M.D.*¶, and Harold P. Lehmann, M.D., Ph.D.*¶
***Department of Pediatrics and ¶Division of Biomedical Information Sciences**
Johns Hopkins University School of Medicine
Baltimore, Maryland

For Interactive Patient II, a multimedia case simulation designed to improve history-taking skills, we created a new natural language interface called GRASP (General Recognition and Analysis of Sentences and Phrases) that allows students to interact with the program at a higher level of realism. Requirements included the ability to handle ambiguous word senses and to match user questions/queries to unique Canonical Phrases, which are used to identify case findings in our knowledge database.
In a simulation of fifty user queries, some of which contained ambiguous words, this tool was 96% accurate in identifying concepts.

INTRODUCTION

Computers and especially simulations have been identified as effective tools for stimulating educational interest and increasing learning.[1] In medicine, computer-based case simulations of patient encounters, using varying degrees of verisimilitude, have been recognized as a popular and beneficial format for medical education.[2] Simulated patient encounters transmit cognitive and procedural information to medical trainees without having to expose real patients to the student.[3] Case simulations enable students to acquire evaluation and management skills and become familiar with current developments in the management of illnesses[4] in an efficient and reproducible manner.
Case simulations are representations of patient encounters with varying degrees of complexity. The needed complexity of the simulation corresponds to the desired educational goals. Essential to clinical learning is mastery of history-taking and physical examination skills to arrive at a differential diagnosis. To develop good history-taking skills, the trainee must learn to elicit responses from patients in a logical, sensitive and appropriate fashion. Prepared history menus[5], while popular, limit the learning experience by giving away clues through limited choices and do not reflect the open-endedness of a real patient encounter.[6] However, the use of Natural Language Processing (NLP) forces the student to formulate the appropriate options on their own and develop a line of "questioning" based upon their knowledge and observations. While more challenging to implement, NLP provides a higher degree of realism and an improved learning experience in this crucial area.

METHODS

Interactive Patient II

We are currently developing Interactive Patient II, a multimedia, Web-based clinical simulation environment designed to teach medical trainees history-taking and physical examination skills, as well as skills in efficient selection of diagnostic tests and treatment options. The Internet's ubiquitous access, relative ease of use, and platform independence made it the medium of choice for this application.[7] A key component of the design was the desire to allow users to formulate, type and submit natural language questions that will be "understood" by the simulated patient and result in appropriate multimedia responses. For the Interactive Patient II natural language interface, we envisioned the use of information retrieval technology to query our knowledge database and return responses relevant to the user's question.

Features

An Interactive Patient II user should be able to inquire about any history-related knowledge in a number of different ways. Inquiries are expected to be appropriate for a physician-patient relationship (e.g., a question that includes a profanity would be considered inappropriate). A user might be interested in the duration of a patient's headache:

1. How long has your head been hurting?
2. You have had the headache since when?
3. What is the duration of the pain in your head?

While these sample questions vary in syntax, choice of words and style, they all attempt to derive the same information. The minimum requirement for the natural language parsing is to relate all these queries to the same knowledge representation and ultimately the same finding from our knowledge database: "My headache has lasted for two days now."

For medical applications, different systems utilizing concept-based algorithms have been used in free-text processing[8-9] including history-taking[10] and medical literature indexing and retrieval.[11]

To accomplish our goals, a natural language parser in a case simulation must contain the following features: The ability to handle queries that contain words with ambiguous meanings (i.e. "drugs") and the capability to identify the meaning of interest for the user. While QMed[10] and CAPIS[8] do not allow for ambiguous words, SAPHIRE[11,12] returns in the case of ambiguity a weighted list of concepts based on the number of synonyms and their proximity to each other.

Because of our belief that understanding users' questions containing ambiguous words is of importance for this simulation, we developed a natural language parser called GRASP (*General Recognition and Analysis of Sentences and Phrases*) to fulfill this specification.

Design

The knowledge in Interactive Patient II is represented by a unique concept representation called *Canonical Phrase*. The core process of GRASP is to return a *Canonical Phrase* and the related case finding in response to any user query. This is done in two steps: First, word-level synonyms are identified and *Canonical Terms* are returned. Second, *Canonical Phrase* synonyms are used to determine the correct *Canonical Phrase*.

Any question submitted by a user is compared to a table of phrases. (Figure 1) These phrases can be single words (example: "pain" or "hurt") or combination of words (example: "how old"). If a phrase is found to be a substring of the query/question string, a related *Canonical Term* is added to the *Canonical Term Query*.

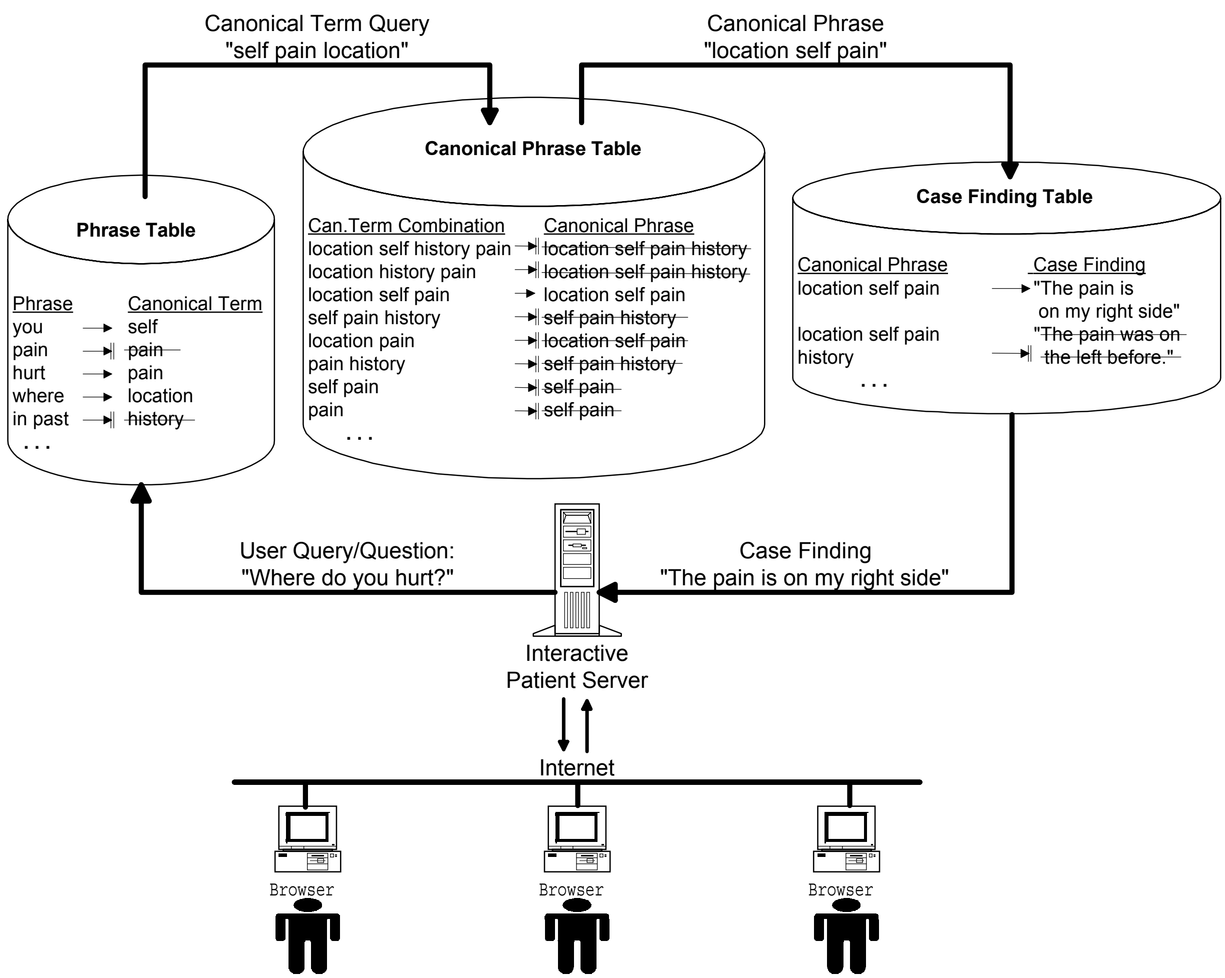

Figure 1: GRASP Data flow diagram (→| = no match generated)

In case a phrase is a combination of more than one word (i.e., "how old"), the phrase is considered a substring if all words contained in the phrase ("old" and "how") are substrings of the original user query/question. Any duplicate occurrences of *Canonical Terms* are removed from the *Canonical Term Query*.

The resulting *Canonical Term Query* is then compared to entries in a table containing all *Canonical Term* combinations for which a concept exists in the case knowledge database. The *Canonical Term* combinations are sorted according to their priority which is determined by the number of *Canonical Terms* that are contained in any given combination as well as some special knowledge rules. The special knowledge rules allow the developer to alter the priority of certain concepts (e.g., the concept "profanity" has the highest priority and overrides all other concepts contained in the question, effectively working as a filter). Once a *Canonical Term* combination is found to be a subset of the *Canonical Term Query*, the search is interrupted and the corresponding *Canonical Phrase* representing the knowledge concept is returned.

The *Canonical Phrase,* in turn, is used to return the case findings from a finding table, which may include text findings, images, sound files and small movies. All user queries/questions as well as the resulting *Canonical Phrases* and findings are saved and periodically reviewed to evaluate GRASP's accuracy and to utilize user requests that were not adequately matched to enlarge our knowledge database.

Implementation

GRASP was programmed as a Java server application. Embedded in an HTML interface, a client-side Java applet allows users to submit their question to the server via a socket connection. GRASP uses server-side JDBC[13] to interact with the database on the Interactive Patient II server and to return the resulting case finding to the user.

The phrase table, *Canonical Term* Combination table as well as the knowledge tables were developed in Microsoft Access. GRASP and the corresponding database are housed on a Windows NT Server on a Pentium II 200 MHz computer.[14]

The first case of Interactive Patient II is currently being developed based on an Adolescent Medicine case. The history knowledge base is divided into a default finding table and a case-specific table. Any query resulting in a *Canonical Phrase* returns a finding in the adolescent case including multimedia files and a text response. If there is no related finding in the adolescent case table, the corresponding finding from the default table is used. This will facilitate the development of new cases in the future. (Figure 2)

RESULTS

To evaluate GRASP against its specifications, fifty random findings from the knowledge database were chosen and questions/queries formulated by the authors that a user might use to access these findings through GRASP (Table 1). All queries were entered into GRASP and in 48 cases (96%), GRASP returned a *Canonical Phrase* that correlated to the desired finding. In one case, a misspelled phrase in the Phrase list resulted in a mismatch, while in the other case, a related *Canonical Phrase* was returned which had a higher priority in the sorting of *Canonical Term* combinations. While the resulting *Canonical Phrase* was not an exact match, the concept returned was similar enough to result in a meaningful answer to a user's query.

The delay between a query and the return of a *Canonical Term* and the related finding in the knowledge database varied significantly. Detailed and specific queries (i.e. "Does your chest pain radiate into your arm?") resulted in significantly faster (less than 2 seconds) responses than broader and more generalized queries (i.e. "Do you have pain?") (3 - 5 seconds).

DISCUSSION

We developed a restricted natural language processing tool for a clinical case simulation that will allow students to interact with a simulation through "questioning" the program. In tests, this tool has proven to efficiently produce accurate phrase identification while allowing usage of ambiguous words.

While word sense disambiguation[15] has developed as an important area of natural language processing, we developed a natural language interface for a clinical simulation that rejects the notion of word sense disambiguation in favor of sense discrimination.[16] Word sense disambiguation and discrimination both identify distinct senses and classify the occurrence of the word as belonging to one of these senses.[15] Word discrimination however does not label the senses or associates them with an external knowledge source (in our case this knowledge is intrinsic to our *Canonical Term* combination table). By allowing phrases to be linked to multiple *Canonical Terms* we allowed for the recognition of ambiguous terms. Using these *Canonical Terms* with terms generated by other phrases in the same query, GRASP was able to identify the correct *Canonical Phrase* and implicitly to determine the appropriate word sense. We demonstrated this with a number of sample questions containing ambiguous words ("arms", "drugs").

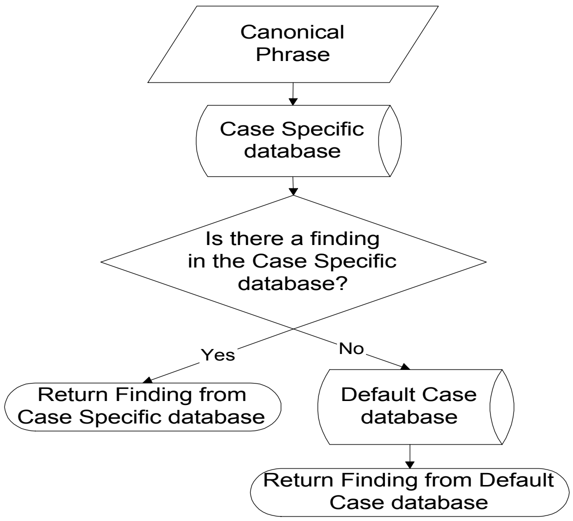

Figure 2: Finding Retrieval Data flow

GRASP's ability to discriminate word sense is in part secondary to the fact that the vast majority of ambiguity found in the English language is of little or no significance in our application. First, our case simulation deals with a very specific context, the history and examination of a patient. Since a specific sublanguage is used, the occurrence of an ambiguous word means that it will most likely be in that context. Second, GRASP is capable of interpreting only those queries for which there exists a specific finding in our knowledge database.

The Interactive Patient prototype[17-18] utilized a single table *Canonical Phrase* look-up, limiting the number of potential user queries by excluding word level synonyms. In the Interactive Patient II, the use of word and *Canonical Phrase* synonyms[11] allowed us to represent a large number of possible queries that could be accurately identified by GRASP, despite a limited number of concepts represented to date.

Furthermore, in our experience with the Interactive Patient prototype[17-18], problems did not occur (according to user feedback) secondary to word sense ambiguity but due to incomplete knowledge representations. For instance, compound queries / conjunctions (e.g., "Do you have vomiting AND diarrhea") that are not represented in the knowledge base, will result in only one finding ("self vomiting" or "self diarrhea"). Utilizing user queries/questions and the resulting *Canonical Phrases* and findings to evaluate the NLP tool for omissions and inadequacies has proven very efficient in the Interactive Patient prototype.[17-18] Approximately 40% of case findings in the prototype were added after users were allowed to interact with the simulation. At the time of this study, the navigational interface for the Interactive Patient II was only partially developed, forcing us to generate testing questions from random findings in the knowledge database introducing an element of bias. After completion of the first case, we will repeat this test with questions formulated by actual users.

As demonstrated by the sample questions, GRASP is capable of determining relevant findings in the knowledge database with a high degree of accuracy. The reason for this accuracy is two-fold. First, due to the format presented to the user, only one query is submitted at any given time. This allowed us to omit parsing for sentences[8,10] and to assume that all *Canonical Terms* identified in this query are related and may be used for word sense discrimination. Second, since all user queries are assumed to be questions to a patient, there is no need to evaluate for negations[8] in these queries since the resulting knowledge findings would be similar. ("Do you have a headache?" vs. "Don't you have a headache?") While in an ideal environment differences between these two queries would be identified since they represent subtle differences in the physician-patient relationship; this differentiation would be unnecessary in a learning tool as envisioned by us.

We found a significant difference in the GRASP's response time for general versus specific concepts. This difference can be explained by the fact that more general concepts usually are matched to *Canonical Term* combinations with low priority and a subsequent longer duration of searching.

The goal for the development of GRASP was to allow users of a case simulation to interact on a more realistic level with the program using a natural language interface. This interface had to fulfill levels of word sense discrimination that existing parsers did not provide.[8-11] Since the order of questions during history-taking depends on the interviewer who is responsible for "completeness", we were able to omit, the need to generate the large complex grammers[8-10] that describe the relationships of the different concepts in the knowledge database.

ACKNOWLEDGEMENTS

This work was funded in part by the National Library of Medicine (Grant # F38LM00064) and by the National Board of Medical Examiners (Project # 34-9798). The authors are indebted to Joan Freedman, M.S. and Bonnie Cosner from the Office of Medical Informatics Education at Johns Hopkins University.

Query/Question	Canonical Phrase
How long have you had the headache?	self head pain duration
Do you own any arms?	self gun
Do you have pain in your arm?	self arm pain
Do you smoke?	self smoke
Do you smoke cigarettes or cigars?	self smoke smok_object

How much do you smoke?	self smoke amount
How long have you had the pain?	self pain duration
Do you have any burning on urination?	self urination burn
* Does it hurt when you pee?	*when self urination pain
Does urination hurt?	self urination pain
What color is your urine?	self urination color
Do you have any blood in your urine?	self urination blood
Do you have any discharge from your penis?	self malegenitalia discharge
* Do you have a history of kidney stones?	*self family kidney stone history
Any other problems?	other self problem
What makes your problem better?	better self problem
How was your health in the past?	history quality self health
Are you exposed to asbestos?	exposure self asbestos
Did you have an injury at work?	self injury occupation
What kind of heat do you have at home?	self home heat
Do you have a gun at home?	self home gun
Do you use a seatbelt?	self seatbelt
Is your mother alive?	self mother alive
How old is your mother?	self mother age
Where does your father live?	where self father home
Did your mother graduate from college?	complete self mother school
Are your parents alive?	self father mother alive
Are your parents married or divorced?	quality self father mother marital status
How long does the nausea last?	self nausea duration
Do you get dizzy drinking alcohol?	self dizziness alcohol sub_abuse
Do you have vomiting and diarrhea?	self vomit diarrhea
Is your stool soft?	self stool consistency
Do you have blood in your stool?	self stool blood
Any heartburn?	self heartburn
Do you take drugs for your cough?	self cough medication
How long have your lymphnodes been swollen?	self lymphnode swelling duration
Do you friends have a problem with drugs?	self friend o_drugs problem
Are you sexually active?	self sexual activity
Do you have any pain?	self pain
Where do you hurt?	where self pain
Does the pain radiate into your back?	self back pain radiate
Is the pain stabbing?	self pain sharp
Do you have any pain in your extremities?	self arm leg pain
When did the pain start?	self pain when start
When did your pain stop?	self pain when stop
Is your pain sharp?	self pain sharp
Does your pain radiate?	self pain radiate
Where does the pain radiate?	where self pain radiate
Does your chest pain radiate into your arm?	self chest pain radiate arm
Do you have left-sided chest pain?	self chest pain left

Table 1: Queries and resulting *Canonical Phrases*.
* Unexpected *Canonical Phrases*.

REFERENCES

[1] Bahr JP. Computer-aided instruction in perinatal education. Am J Perinatol 1986 Apr; 3(2):147-50.

[2] Wu AH, LaRocco M, Fath SJ, Simon FA. Evaluation of computer case simulations for teaching clinical pathology to second-year medical students. Ann Clin Lab Sci 1990 Mar-Apr; 20(2):154-60.

[3] Mistry FD. Using case presentations to augment instruction in laboratory medicine. J Med Educ 1982 Oct;57(10 Pt 1):804-6

[4] Piemme TE. Television and computer technology. In: Rosof AB, Felch WC, editors. Continuing Medical Education: A Primer. 2nd ed. New York: Praeger Publishers: 215-222, 1991.

[5] Myers JH, Dorsey JK. Using diagnostic reasoning (DxR) to teach and evaluate clinical reasoning skills. Acad Med 1994 May; 69(5):428-9.

[6] Friedman CP. Anatomy of a clinical simulation. Acad Med 1995 Mar; 70(3):205-9.

[7] Lehmann HP, Lehmann CU, Freedman JA. The use of simulations in computer-aided learning over the World Wide Web. JAMA 1997 Dec;278(21): 1788.

[8] McCray AT. Extending a natural nanguage narser with UMLS knowledge. Proc Annu Symp Comput Appl Med Care, 1991;:194-8

[9] Lin R, Lenert L, Middleton B, Shiffman S. A free-text processing system to capture physical findings: Canonical Phrase Identification System(CAPIS). Proc Annu Symp Comput Appl Med Care 1991;:843-7

[10] Johnson K, Poon A, Shiffman S, Lin R, Fagan L . A history-taking system that uses continuous speech recognition. Proc Annu Symp Comput Appl Med Care 1992;:757-61

[11] Hersh WR, Hickam DH, Haynes RB, McKibbon KA. Evaluation of SAPHIRE: An automated approach to indexing and retrieving medical literature. Proc Annu Symp Comput Appl Med Care 1991;: 808-812.

[12] Hersh W, Leone TJ. The SAPHIRE server: a new algorithm and implementation., Proc Annu Symp Comput Appl Med Care 1995;:858-62.

[13] Sun Microsystems. The JDBC Data Access API [online]. 1998 Dec [cited 1999 Mar 3]. Available from: URL: http://java.sun.com/products/jdbc/.

[14] Lehmann CU, Nguyen B, Kim GR, Johnson KB, Lehmann HP. GRASP [online]. 1998 Nov [cited 1999 Mar 3]. Available from: URL: http://omie.med.jhmi.edu/IPServer/.

[15] Kilgarriff A. What is word sense disambiguation good for? Proc. Natural Language Processing Pacific Rim Symp. Phuket, Thailand. 1997 Dec; p. 209-14.

[16] Schütze H. Automatic Word Sense Discrimination, Computational Linguistics 1998; 24(1): 97-124.

[17] Lehmann CU, Hayes KA. The Interactive Patient [online]. 1995 May [cited 1999 Mar 3]. Available from: URL: http://medicus.marshall.edu/medicus.htm.

[18] Hayes KA, Lehmann CU. The Interactive Patient. MD Computing. 1996 Jul/Apr; 13(4): 330 - 4.