"Expertness" from Structured Text? RECONSIDER: A Diagnostic Prompting Program

Mark S. Tuttle[a,b], David D. Sherertz[c], Marsden S. Blois[a], Stuart Nelson[d]

[a]Section on Medical Information Science, A-16, Univ. Cal. San Francisco, San Francisco, CA 94143
[b]Computer Science Division - EECS, Univ. Cal. Berkeley, Berkeley, CA 94720
[c](formerly) Section on Medical Information Science, UCSF
[c](currently) Tandem Computers, Inc., 19333 Vallco Parkway, Cupertino, CA 95014
[d]Dept. of Medicine, State Univ. of New York at Stony Brook, Stony Brook, NY 11794

Abstract:

RECONSIDER is an interactive diagnostic prompting program which uses simple information retrieval techniques to prompt a physician regarding possible diagnoses, given a list of positive patient findings. Its knowledge base consists of "structured text" definitions of 3262 diseases and a synonym dictionary. Patient findings, and their synonyms, are matched against inverted files of terms from the disease descriptions, the number and selectivity of the patient findings matching terms in a given disease description determine that disease's "score", and the matched diseases are sorted on this score to form a preliminary differential diagnosis. Definitions of diseases can be referenced for viewing by name, or by their position in a differential. While its first formal evaluation is not yet complete, the performance of RECONSIDER continues to exceed the expectations of user and designer alike.

1. Motivation & Background

A review of the various means by which medical knowledge is represented in symbolic form [6, 7] led us to formulate the following spectrum:

Human Processible	Machine Processible

Spectrum of
Medical Knowledge Representation
Schemes

The two endpoints of the spectrum represent the limiting cases wherein

> knowledge is difficult, or impossible, to process algorithmically, but transparent to medical personnel, e.g. free text;
>
> or easily processible algorithmically, and difficult to process by humans untrained in applied mathematics or computer science, e.g. a matrix of Bayesian probabilities, or a semantic network.

Those attending this conference will be familiar with work at both ends of the spectrum, if not in medicine, then in other knowledge domains. Most will concede that the greatest "successes" in the field of expert systems has been achieved by those working at or near the right-hand end of the spectrum; and that progress has been most difficult to achieve at the left-hand end of the spectrum. We concluded that, for the short run at least, those successes at the right-hand end would prove to be self-limiting - knowledge that was not readily accessible to and modifiable by the medical community at large could not remain in the mainstream of medical practice. Similarly, we saw no immediate prospects for a breakthrough in the algorithmic understanding of free text, though impressed with accomplishments in the area of natural language access to databases [9, 10].

The dilemma these observations implied led us to formulate the following question:

> *Can knowledge about diseases be represented in a form that is easily comprehended by physicians not trained in computer science or artificial intelligence, and yet still be algorithmically processible toward some medically useful end?*

Our initial attempts to answer this question led us to formulate yet another knowledge representation scheme, one which operated somewhat to the right of the human-processible end of the spectrum. Conceding the important role of *words* (rather than text) as conveyers of meaning in medicine, we focused on a hierarchical nominal-attribute model, wherein nominals (electrons, cells, lungs, etc) were "defined" in terms of attributes (spin, neoplastic, congested, etc.). Obviously, nominals could be attributes of other higher level nominals, and attributes could be nominals at a lower level. The principle result of this model was the observation that some words had meaning only at certain levels - electrons could not be congested or neoplastic, nor could lungs or cells have spin. While the idea of "levels of description" is not new, such levels were observed to be both well separated and powerful determiners of context in medicine.[1] In turn, well defined contexts implied, not

[1]This is not a tautology. In the world of artifacts (man-made nominals), levels are not so well separated or orderly. Until recently one would not ordinarily think of 'spark plug' and 'computer' as having closely connected meanings, but new electronic ignition systems in cars combine both in a single system. Biological systems are not so freely re-arranged.

surprisingly, well determined meanings for words, diminishing the need for syntax to clarify or disambiguate meaning.

Our search for an body of knowledge on which to explore certain hypotheses regarding such a nominal-attribute model in medicine led us to regard a familiar but little used resource in medicine in a new light. A corpus of computer readable disease definitions was seen to be a crude instantiation of the model. In this corpus each disease was given a name (a nominal), and defined by its (usually clinical) attributes - the original motivation for the corpus being the standardization of disease nomenclature. The attributes were written in a telegraphic, but otherwise easily readable style, and organized, for each disease, in a relatively stable format - a form we have chosen to call *structured text*.

Superficially, the corpus had but one level of description, attributes of diseases. But each disease definition was divided, explicitly, into "contexts" (etiology, symptoms, ... , lab, x-ray, ...), and each disease was place in one (or, at most, two) "body systems" (whole body, skin, ... , urogenital, ...). These contexts and systems were obviously strong, if imprecise, determiners of context.

Early experiments [3,4] with this corpus, the computer readable version of *Current Medical Information and Terminology (CMIT), 4th Edition* [*11*], explored the selective and associative power of the words it employed, and confirmed our hypothesis that word use in it was both relatively consistent and systematic. We soon realized that the sharpest test of the ability of words to convey meaning in this context was to evaluate the corpus as a knowledge base for a "diagnoses program" which would accept a description of the patient in the form of a list of words, such as 'pain, fever, jaundice, ...'. The specific diagnostic problem we addressed was that of formulating a "differential diagnosis"[2] [12, 15], which included, as alluded to by Scadding [21, 5], diseases that a physician might not *otherwise* think of, but, perhaps, should think of.

Important to our attempt to formulate a diagnostic prompting program was not only that the knowledge base should be readily comprehensible, but, if the disease "prompts" were to be credible, the "reasoning" by which diseases were retrieved and ranked had to be equally accessible - a consultative criterion noted by Shortliffe and co-workers [22, 23].

In addition, the availability of a knowledge base containing in excess of 3000 disease descriptions has allowed us to study phenomena that would be hard to reproduce in the context of most "expert systems".[3] For example, appended to this paper is a transcript of an interaction with *RECONSIDER* regarding a case of methanol poisoning supplied by one of the authors (SN). None of the patient findings are particularly specific, but *RECONSIDER* places the correct diagnosis in 6th place, and determines that most of the diseases near the top of the differential are "whole body" diseases, a group containing most toxicity diseases. If this differential were selected from among a few hundred diseases, or even from a knowledge base of toxicity diseases, the result would be more open to a variety of less favorable interpretations. Put differently, when one is retrieving from such a large knowledge base, one is more tolerant about the appearance of "false positives" (diseases that shouldn't be there) in the interests of minimizing the number of "false negatives" (diseases that should be there, but are not).

Finally, *RECONSIDER* provides a test bed for the evaluation of some hypotheses regarding the kind of problems encountered representing and utilizing knowledge about the 'natural', as opposed to 'artificial', world. Briefly, *RECONSIDER* benefits from the high degree of structure observable in diagnostic medicine, in spite of our ignorance in many areas, and the otherwise generally unappreciated stability and specificity of medical language regarding this structure.

2. Expectations?

Non-medical audiences should be reminded of differing expectations regarding such meaning representation experiments. As computer scientists, two of us (MST & DDS) "knew" that meaning could not be represented satisfactorily by words alone: words were ambiguous, in general, and, besides, syntax was a partner with semantics, and to separate the two was to grossly distort the meaning of either.[4] We regarded early efforts as potentially interesting from the point of view of statistical linquistics - how did words and contexts associate? However, the medically trained member of the initial team (MSB) predicted the successful performance of *RECONSIDER* once he saw the results of some early word-counting experiments. Later, SN, an internist with a background in mathematics, anticipated the performance limiting aspect of *RECONSIDER* without ever using the program! (He predicted that inadequacies in the knowledge base would be more important than any shortcomings in the algorithms by which descriptions of patients were "matched" with the descriptions of the diseases.)

[2]A "differential diagnosis" is usually a list of diseases which represents the current thinking of a physician regarding possible diagnoses for a given patient, at a given point in the diagnostic process.

[3]The best know diagnosis program, an expert system formerly named INTERNIST - now called CADUCEUS, currently "understands" a few hundred diseases in the field of internal medicine [19, 18, 20, 16, 14].

[4]A local example of failure in "full-text searching" was recently brought to our attention [13]. In a search of documents in a database collected for a suit regarding a large construction project, *precision* (the probability of a document being relevant) was no better than 80%, which might have been acceptable except for the fact that the *recall* (the probability that the relevant document will be retrieved) was no better than 20%!

3. An Example of 'Structured Text'

CMIT was designed first for human users, as a reference of standard disease names (in book form it is about the size of the World Almanac), and second for computer applications. (The *RECONSIDER*-formatted CMIT definition of *methyl alcohol, toxicity* appears in the appendix of this paper.) The "structure" imposed on CMIT definitions is largely external to the language of those definitions.

First, the entire text of CMIT is organized in the aforementioned *nominal-attribute* form, the disease names being the nominals and the descriptions consisting of the attributes of the disease.[5]

Second, each disease is assigned to one, or possibly two, *body systems:*

whole body
skin
musculoskeletal
respiratory
cardiovascular
hemic & lymphatic
gastrointestinal
urogenital
endocrine
nervous
special sense organs

Third, each disease is described in *parts:*

additional terms (synonyms & eponyms)
etiology
symptoms
signs
complications
laboratory
pathology
x-ray
references[6]

Fourth, within each *part*, the descending hierarchy of *sentences*, *clauses*, and *phrases* (all inferrable from punctuation) are used relatively consistently to denote appropriate "chunks" of meaning.

Thus, in this instance, *structured text* is tightly edited prose written in nominal-attribute form, employing external markers, and relatively consistent punctuation, style, and vocabulary. Put differently, CMIT can be "structurally" parsed without the need to *infer* any of the semantics from the text. (Again, a portion of this "parse" is what produces the "display" of the definition of *methyl alcohol, toxicity* shown in the appendix.)

[5]As we are learning from our evaluation, the *names* of diseases, even when they are descriptive names (as CMIT is designed to encourage), are not always sufficient to determine which disease is being spoken of. Without the descriptions (attributes) physicians would be unable to resolve the problems created by different systems of disease nomenclature.

[6]An important feature of the computer readable version of CMIT is that it contains references, mention of which is not made in the printed version.

4. The Current RECONSIDER Implementation

4.1. The Inverted File

Using abstract syntax to represent the structure in the text, CMIT was scanned and "parsed"[7] to produce a sequence of *terms*, each with the following attributes:

ordinal position of term in phrase
ordinal position of phrase in clause
ordinal position of clause in sentence
ordinal position of sentence in part
name of part
disease
body system(s) of disease

Thus, a dictionary (containing in excess of 20,000 such terms) was formed and CMIT "inverted", so that each dictionary entry was followed by pointers to every occurrence of that entry in CMIT. Included with every pointer were the seven attributes associated with each occurrence of that term. There are 333,211 term occurrences in CMIT, for an average of about 102 terms per disease, or 79 unique terms per disease, the difference being terms that are used more than once in a given definition. In principle, this "dictionary" could be used to reconstruct CMIT, *as it represents, in alternative format, exactly the same information!*

This large inverted file allows efficient searching for terms in the text. The searches can be (1) constrained to a context (diseases of the skin), (2) constrained to textual proximity (adjacency, or membership within a clause), or (3) constrained to a definition part (symptoms only).

4.2. Synonym Dictionary

A 15,388 term "synonym" dictionary[8], includes words not in CMIT which are synonyms of words used in the CMIT definitions and words already in CMIT that are synonyms of each other (e.g. *pruritus* and *itching)*. These are partitioned amongst 4,165 "synonym classes" (the two or more words within each class are synonyms of each other). Search options allow searches with or without equivalencing the synonyms, and with or without invoking hierarchical synonyms. The term "synonym" is used generously, as the dictionary is actually functioning as a kind of semantic net - connecting words with strong conceptual links. It should also be noted that *RECONSIDER* does not employ "stemming". All variants of a term (and some phrases, e.g. *abdominal pain*), including, in some cases, mis-spellings, appear within a single "synonym class". Though we have not proven this, it is our opinion that this synonym dictionary is what converts an interesting tool for research into medical term-use, into something

[7]Once again, this parse is not identifying "parts of speech" in the conventional sense. Rather the abstract syntax (a BNF grammar akin to those defining programming languages) encodes the meaning of the external markers and punctuational conventions employed in CMIT.

[8]Constructed by Rodney Ludwig, M.D. and Hyo Kim, M.D..

that functions not unlike an expert system.

4.3. Searches

Searches for a set of terms can require a match on every term, or a match on one or more of the terms in the set. In the latter case, matches are scored in a manner reminiscent of techniques used for literature and information retrieval by Salton, Sparck-Jones and others, and in particular Doszkocs [8]. The scoring algorithm is illustrated in the next section.

4.4. The User-Interface

RECONSIDER is an interactive user interface running on top of the inverted file and the search algorithms. It accepts terms, search modifiers, and requests for one of the two matching algorithms, formulates the appropriate query, searches the inverted files, computes the score of the diseases retrieved (if requested), constructs a body-system histogram (if requested), ranks the diseases if appropriate, and displays any disease definitions selected for viewing or browsing by the user.

5. Performance

5.1. A Comparison with two Diagnostic Expert Systems

When applied to the published cases diagnosed by *INTERNIST* and PIP [20, 17, 16], *RECONSIDER* produced the correct diagnosis (or diagnoses) at, or near, the top of the disease list produced by entering the positive findings given to these programs [5]. (Again, CADUCEUS considers 300 diseases from internal medicine, and PIP considers 20 diseases featuring edema.) While these cases were often complex, a large amount of clinical information was available for each patient.

5.2. Diagnostic Prompting: An Example

We believe that *RECONSIDER* performs better, and much more usefully, at an earlier point in the diagnostic process, at a time prior to any extensive patient work-up, when the physician's "cognitive span" is widest [2].

For example, a patient presents with findings as noted at the beginning of the appendix. *RECONSIDER* begins by prompting for terms. The prefix *ss/* is used by the physician-user to indicate that the succeeding terms are to be searched for in either the *symptoms*, or *signs* portions of the disease descriptions. This grouping, a union of the two vocabularies, was necessitated by the non-consistent usage of terms in these contexts.[9] The phrase *abdominal pain* will match (given the *RECONSIDER* options selected to run this case) any co-occurrence of these two words (or its synonyms) within a single clause. *RECONSIDER* responds with the synonyms it knows for the terms entered, and the number of diseases containing one of more occurrences of each of the terms within the *ss/* context. The response *abdominal pain[191+80]* indicates that the pair *abdominal pain* occurs in 191 diseases and that 80 additional diseases have been retrieved by the synonyms for *abdominal pain*, namely *colic[35]*, *colicky[16]*, and *pain in abdomen[48]*. The fact that 35+15+48 exceeds 80, and 191+35+18+48 exceeds 191+180, indicates that some disease definitions contain more than one term from this synonym class.

The *score* (a measure of selectivity) for *abdominal pain* is

$$0.917 = 1 - (271/3262)$$

where 271 is the number of "disease occurrences" of *abdominal pain*, and 3262 is the total number of diseases in CMIT. A disease's score is the sum of the scores of the terms its description matched.

Most physicians would probably conclude that the observation that the patient smoked was not relevant to the patient's illness, but the term *smoking* was entered here to show its obvious effect on the disease list (it brings *nicotine, toxicity* and *drug dependence*, *marihuana* nearer to the top, partly because it is so "selective"). It is not clear which 'part' of the disease descriptions the term *smoking* will be found in, so its search context is *all/*, and the same decision is made with respect to *acidosis*. *Anion gap acidosis* is not used in CMIT, so we enter the more general form.[10] Entering *smoking* in the *all/* context has the disadvantage that it brings in a reference to *smoky*, which is used as an adjective.

The histogram displays the body system frequencies for the diseases near the top of the disease list (the top 4% was selected by the user to include about the first "screen's worth" of the disease list - 679 diseases containing one or more of the terms entered, or their synonyms).

A physician-user viewing the first screenfull of this list (the portion shown in the appendix) would next formulate a strategy for resolving it, assuming the diagnosis was still not immediately apparent. A methodical approach would note first that no disease matched all five entries (as no disease has a score of 4.738). Similarly, diseases #1, #2, and #3 would be ruled out by asking the patient appropriate questions. (If the patient were from Marin County, here in the Bay Area, we might focus our initial attention on #2, *mushroom, toxicity*, in response to recent news reports of cases of it there -

[9] The use of terms within CMIT did not follow the medical dogma as to what was a *symptom*, and what was a *sign*.

[10] An attempt on the part of the user to enter *anion gap acidosis*, while laudable (it would be very selective), would be greeted by a message that the term was not found in CMIT or its synonym dictionary - in this case because CMIT predates wide use of this test. At this point the physician-user must use his or her own knowledge of medicine, to know that the term *acidosis* is the best substitute under these circumstances. Looked at differently, our evaluation seems to confirm that, in general, more medical knowledge makes one a more effective *RECONSIDER* user. If true, we regard this as a positive feature of *RECONSIDER*.

knowledge that is not available to *RECONSIDER*.) Disease #4, *eclampsia*, raises a more interesting issue. *RECONSIDER* does not have a model of gender (or of anything else), so a disease that occurs during pregnancy is not automatically ruled out when the patient is male. While understandably distracting at first, users are soon comfortable ignoring such inclusions, especially since it's easy to understand *why RECONSIDER* put the disease there. Viewing the CMIT definition of disease #5, *nephritis, salt losing* reveals that it is usually accompanied by a rich complex of symptoms, so while it can not be ruled out at this point, it becomes extremely unlikely. Since the patient is not an alcoholic, the definition of disease #6, *methyl alcohol, toxicity*, suggests the possibility of occupational exposure (perhaps percutaneous or respiratory). Once considered, an appropriate test would confirm the existence of the toxic substance in the body.

6. End-User Experience

We have not permitted *RECONSIDER* to be used "live" in a clinical context. In addition to the fact that evaluation of the program is not complete, the knowledge base is known to be out of date. Nonetheless since we have been able to move *RECONSIDER* to the MIS-UCSF VAX 11/750 running UNIX* (Berkeley 4.1) students, post-doctoral fellows and some faculty have been able to use the program. The initial reaction usually consists of the following three observations: (1) "Why is that disease there?" (sometimes it's there legitimately, and sometimes not), (2) "How does such a dumb program do so well?" (referring to *RECONSIDER*'s lack of evident reasoning power), and (3) "What I need to be able to do now is ..." (fill in your favorite interactive-knowledge-base user-feature).

We tolerate the problem alluded to by question (1) because it is more important, at this stage of development, not to miss important diseases, and because it is easier for a physician-user to reject totally inappropriate diseases than it is for the program to do so. Question (2) alludes to the point raised by the title of this paper. *RECONSIDER* can only be considered an "expert" (if at all) because its knowledge base is so large (relative to what a physician can keep readily available in his or her head), and because of its performance. It is obviously not like a human "expert" in the way it *arrives* at the disease list. And question (3) we take to be a compliment that reveals, among other things, that occasionally the utility of RECONSIDER is limited not by the knowledge it contains, but by the means we currently have of accessing it through the narrow window of a 23-line CRT terminal.

Question (1) deserves further comment. The author (MST) has observed considerable user-discomfort caused by CMIT mixing diseases from several body systems near the top of a sorted disease list. Apparently, the cognitive dissonance is usually avoided by thinking about diseases by system, as the discomfort can be relieved by restricting the search (and thus the sorted list) to a single body system. The problem with the latter practice is that the preliminary results of our evaluation reveals that contextless (*all*/searches) are the most efficacious, on average. As this is also the opposite of the behavior predicted by our model of context in a nominal-attribute knowledge base, further study is suggested. In any case, it may prove necessary to re-design the user-interface to accomodate some users' need to view deseases by system, within a contextless search.

*UNIX is a product of Bell Telephone Laboratories, Inc.

7. Evaluation

A formal evaluation of *RECONSIDER* on 100 serial admissions to a tertiary care medical ward, is in progress (and will be reported elsewhere), but the preliminary results are both encouraging and interesting. They are encouraging because the correct diagnoses is included so often in the first frame or two (and usually higher), and interesting because the difference between diagnostic programs, and diagnostic *prompting* programs is made quite clear. The former have a very specific goal, and it is easy to tell whether it is reached or not. A prompting program is evaluated against a different standard; not whether it is correct but whether it is helpful. And judging whether something is helpful or not may be a subtle matter. If the correct diagnosis is included high on the list, the performance can be given a high score. But if, instead, a listed disease closely related to the correct one has the result of directing the physician's attention to the correct body system, and finally the correct diagnosis, how is this to be scored?

8. Suspected Limitations:

8.1. The Knowledge Base

As has been the experience with similar projects, computer processing subjects "knowledge" to a harsh and unyielding light. We anticipate that a half a man-year of "tuning" would significantly improve *RECONSIDER*'s performance, but that the next and much more serious limitation will be the quality, uniformity, completeness, and timeliness of CMIT and the synonym dictionary. Given the opportunity to rewrite CMIT (and continue to do so on an on-going basis), or introducing AI techniques to *RECONSIDER* (we have received many suggestions), we would choose the former.

8.2. Other Limitations

Our experience to date has taught us that, in this context, negatives are important. Terms such as *fever absent* are teated as if *fever* were a positive finding; while not fatal, such retrievals increase the number of false positives. Also users often wish to search using "rule-out", e.g. eliminate all diseases from consideration containing a certain term, or terms. Especially tricky would be interactions between these two uses of negation.

On a more global level, CMIT's homogenization of diseases contributes to confusion and loss of information. Congestive heart failure is listed as a disease under *heart, failure, congestive*, as a symptom under *heart, hypertensive, disease*, as a sign under *heart, hypertrophy, heart, fatty degeneration* and *aortic stenosis, subvalvular*, and as a complication in, for example, *trypanosomiasis, American*. And to illustrate the stress on the process of attempting to form a closed set of synonyms, the symptoms and signs of *congestive heart failure* are described at various points as in *cardiomyopathy*, but the phrase *congestive heart failure* does not occur in that description.

9. Future Implementations

Given an opportunity to re-implement CMIT, we would retreat from our original notion that it should not be modified (so as to prove that structured text could be used, intact, as a knowledge base). Rather we would maintain the inverted files dynamically, in a relational database, so as to facilitate modifications, and experiments with alternative knowledge representations and retrieval techniques. Specifically, we would investigate the difficulty of re-writing CMIT to improve the quality and timeliness of the information it contained, to use a more standard model of disease nomenclature [1], to evaluate alternative ways of handling negation (such as *jaundice absent*), and the allow users to specify *necessity* (a term *must* occur, or not occur, in a disease description for it to be retrieved).

RECONSIDER currently requires some 20 MB of disk space. A dynamically revisable version would require at least twice that, making *RECONSIDER* a little like an orphan elephant in already pressed medical computing environments. A "production" version of *RECONSIDER* might fit in 15 MB, leaving two alternatives for the future: running *RECONSIDER* on the large address-space micro-based systems now available with large hard disks, or making it available on a network. We are looking into both possibilities.

10. Conclusions

In the context of medical diagnoses, and perhaps in other application areas, "structured text", as defined here, has been neglected as a means of representing information in a form accessible to both humans and algorithms. If as Minsky has put it, "For a program, being smart is knowing a lot.", then carefully edited and constructed natural language text, available in computer-readable form, may facilitate the process by which programs come to "know a lot" and continue to "know a lot" as the knowledge evolves over time.

We conclude by noting that ultimately the usefulness of diagnostic aids such as *RECONSIDER*, must await the verdict of users. If the cost and bother of their use is less than the benefit they are found to provide, we can expect them to make their way into clinical practice. Up until the present time, no diagnostic support program seems to have accomplished this.

11. Acknowledgements

Future reports will include the performance of the case "enterers" who have labored to complete the task of formulating differentials for 100 cases. As some of their reactions are included here they are acknowledged below. Those case-enterers who are not co-authors are Mark Erlbaum, M.D., Peter Harrison, M.D., Hyo Kim, M.D., Pauline Velez, medical student, Bernard Winklmann, M.D., Dale Yamashita, M.D.

12. References

1. , *ICD.9.CM - The International Classification of Diseases, 9th Revision*, Commission on Professional and Hospital Activities, Green Road, Ann Arbor, Michigan (1978).
2. Blois, M. S., "Clinical judgment and computers," *New Eng. Jour. Med.* 303 pp. 192-197 (1980).
3. Blois, M. S., D. D. Sherertz, and M. S. Tuttle, "Word and Object in Disease Descriptions," *Proc. of the 18th Ann. Meeting of Assoc. for Computational Linguistics (Philadelphia, June, 1980)*, Assoc. for Computational Linguistics, (1980).
4. Blois, M. S., D. D. Sherertz, and M. S. Tuttle, "The Algorithmic Processing of Medical Text Utilizing Context," *Proc. of the 4th Ann. Symposium on Computer Applications in Medical Care (Washington, D.C., November, 1980)*, IEEE, (1980).
5. Blois, M. S., M. S. Tuttle, and D. D. Sherertz, "RECONSIDER: A Program for Generating Differential Diagnoses," in *Proc. of the 5th Ann. Symposium on Computer Applications in Medical Care (Washington, D.C., November, 1981)*, IEEE, New York (1981).
6. Blois, M. S., "Conceptual Issues in Computer-Aided Diagnosis and the Hierarchial Nature of Medical Knowledge," *Jour. Med. Phil.* 8 pp. 29-48 (1983).
7. Blois, M. S., *Information and Medicine; An Hierarchial View*, Univ. of California Press, Berkeley (1983 (in press)).
8. Doszkocs, T.E., "An Associative Interactive Dictionary (AID) for On-Line Bibliographic Searching," *Proc. American Society for Information Science* 15 pp. 105-109 (November, 1978).
9. Epstein, Martin, "Natural Language Access to a Clinical Database," TR-51, Section on Medical Information Science, Univ. Cal. San Francisco, San Francisco, CA (1980).

10. Epstein, Martin N. and William Lewis, "Methodology for Creation of and Access to a Clinical Database," *Proceedings of the First International Conference on Medical Computer Science*, IEEE Computer Society, (September, 1982).

11. Gordon, B. L.(Ed.), *Current Medical Information and Terminology, 4th Edition*, American Medical Association, Chicago (1971).

12. Lindberg, D. A. B., L. R. Rowland, C. R. Buch, jr., W. F. Morse, and S. S. Morse, "CONSIDER: A Computer Program for Medical Instruction," in *9th IBM Medical Symposium*, (1968).

13. Maron, M.E., "Problems with Full-Text Searching," *Office Automation Conference Digest*, AFIPS, (April, 1982).

14. Miller, R. A., H. E. Pople, and J. D. Myers, "INTERNIST-1, An Experimental Computer-Based Diagnostic Consultant For General Internal Medicine," *New England Journal of Medicine* **307, No. 8** pp. 468-476 (Aug. 19, 1982).

15. Morse, S. S., L. R. Rowland, and D. A. Lindberg, "CONSIDER Implementation," Technical Report MOU-IS-TR-6, Univ. of Missouri, Columbia, MO (1971).

16. Myers, J. D., H. E. Pople, and R. A. Miller, "INTERNIST: Can Artificial Intelligence Help?," pp. 251-269 in *Clinical Decisions and Laboratory Use*, ed. D. Fenderson,Univ. Minnesota Press, Minneapolis (1982).

17. Pauker, S. G., G. A. Gorry, J. P. Kassirer, and W. B. Schwartz, "Towards the simulation of clinical cognition: Taking a present illness by computer," *Amer Jour Med.* **60** pp. 981-996 (1976).

18. Pople, H. E., "The formation of composite hypotheses in diagnostic problem solving: an exercise in synthetic reasoning," in *Proc. of the 5th Int. Joint Conf. on Artif. Intell.*, Cambridge, Mass. (1977).

19. Pople, H. J., J. Myers, and R. Miller, "DIALOG: A Model of Diagnostic Logic for Internal Medicine," *Proc. Int. Joint Conf. A.I., Tbilsi, USSR, 1975*, ().

20. Pople, H. J., *Heuristic Methods for Imposing Structure on Ill Structured Problems: The Structuring of Medical Diagnostics (1980)*, American Association for the Advancement of Science, Washington, D.C. (in press).

21. Scadding, D., "Diagnosis: The clinician and the computer," *Lancet*, p. 876 (1967).

22. Shortliffe, E. H., *Computer-Based Medical Consultations: MYCIN*, Elsevier, New York (1976).

23. Shortliffe, E. H., B. G. Buchanan, and E. A. Feigenbaum, "Knowledge engineering for medical decision making: a review of computer-based clinical decision aids," *Proc. IEEE* **67** pp. 1207-1224 (1979).

Appendix: A Case of Methanol Poisoning

A 26 year old male was admitted to the medical ward of the SUNY Stony Brook hospital complaining of abdominal pain, confusion, and vomiting. It was noted that the patient was a smoker. A lab test had revealed anion gap acidosis.

```
Enter terms: ss/abdominal pain,confusion,
             vomiting

Signs or Symptoms: abdominal pain[191+80]
(colic[35], colicky[16], pain in abdomen[48]);
confusion[85+7] (confused[7]);
vomiting[425+1] (emesis[2], hyperemesis[2],
hyperemisis[1], vomitus[9]).

Enter terms: all/smoking,acidosis

Signs or Symptoms: abdominal pain[191+80]
(colic[35], colicky[16], pain in abdomen[48]);
confusion[85+7] (confused[7]);
vomiting[425+1] (emesis[2], hyperemesis[2],
hyperemisis[1], vomitus[9]).
all: smoking[23+8] (smoke[8], smoky[1]) ;
acidosis[37+1] (acidemia[1]).

Command: s
Computing scores for Signs or Symptoms terms
   Finished abdominal pain, selectivity = 0.917
   Finished confusion, selectivity = 0.972
   Finished vomiting, selectivity = 0.870
Computing scores for all terms
   Finished smoking, selectivity = 0.991
   Finished acidosis, selectivity = 0.989
Sorting totaled disease scores...
       4.738 - maximum total score
       679 diseases in this list

Histogram for the top 4% of the list
(27 diseases)
      Whole body *******************************************
            Skin ***
 Musculoskeletal *****
     Respiratory **
  Cardiovascular **
  Hemic-Lymphatic *
Gastrointestinal **********
     Urogenital ************
       Endocrine *
         Nervous ****
   Special sense

Signs or Symptoms: abdominal pain[191+80];
       confusion[85+7]; vomiting[425+1].
all: smoking[23+8]; acidosis[37+1].

        4.738 - maximum total score
        679 diseases in this list
    1   3.750  nicotine, toxicity   00
    2   3.748  mushroom, toxicity   00
    3   2.833  drug dependence, marihuana   00
    4   2.830  eclampsia   07
    5   2.830  nephritis, salt losing   07
```

6	2.776	methyl alcohol, toxicity 00
7	2.776	food poisoning, staphylococcal 00-06
8	2.776	coma, diabetic 08
9	2.759	thallium, toxicity 00
10	2.759	arsenic, toxicity 00
11	2.759	migraine syndrome 00-09
12	2.759	porphyria, acute intermittent 06
13	1.961	psychosis with metabolic, nutritional disorder 00
14	1.961	carbon dioxide, narcosis 03
15	1.961	consumption coagulopathy 05
16	1.961	coma, hepatic 06-09
17	1.906	fanconi syndrome, adult, without cystinosis 02-07
18	1.906	diarrhea, chronic 06
19	1.906	kidney, calculus 07

6 2.776 methyl alcohol, toxicity 00
(ss) abdominal pain[0.917],
(ss) vomiting[0.870],
(al) acidosis[0.989],
END

(6) methyl alcohol, toxicity 00

Alternate terminology [at]
toxicity, methyl alcohol
wood alcohol, toxicity;
methanol, toxicity.

Etiology [et]
- Inhalation of vapor, ingestion, percutaneous absorption of flammable liquid widely used in industry;
- effect of metabolization by body to formaldehyde and formic acid, with depressant action on cns;
- tlv, 200 ppm of air;
- internal lethal dose, 60-250 ml or 2-8 oz.

- Occupational exposure: dry cleaning, organic synthesis;
- manufacture of antifreeze, dyes, explosives, fuel, leather, plastics.

Symptoms [sm]
- Acute poisoning from ingestion, inhalation, or percutaneous absorption: fatigue;
- headache;
- nausea;

==> vomiting;
- vision impaired;
- photophobia;
- dizziness;

==> in exposure to high concentration or ingestion of high dose, manifestations more marked as severe upper abdominal colicky pain, sweating, possibly blindness.

- Chronic poisoning from inhalation, percutaneous absorption: vision impaired initially, progressive;
- fatigue;
- nausea.

Signs [sg]
- Acute poisoning: with ingestion, onset within 6-48 hours;
- cyanosis;
- cold, clammy skin;
- euphoria;
- respiration shallow;
- blood pressure low;

==> features of acidosis;
- cns depression;
- convulsions;
- coma.

- Chronic poisoning: eczematoid dermatitis;
- conjunctivitis;
- tracheitis;
- bronchitis;
- unsteady gait.

- Course: in severe acute poisoning, mortality rate 25-50 percent;
- in milder forms, recovery within weeks to months;
- vision, renal function possibly impaired permanently.

- Treatment: administration of sodium bicarbonate orally or sodium lactate

==> intravenously for acidosis;
- irrigation of eyes with water;
- washing contaminated areas of body with soap, water;
- combating shock with oxygen, stimulants;
- oral administration of whiskey or intravenous administration of 10 percent ethanol possibly inhibiting oxidation of methanol to its toxic intermediates.

Laboratory [lb]
- Methyl alcohol in expired air, urine, blood;
- formic acid in urine.

- Ophthalmoscopy: in acute poisoning, dilatation of pupils, contraction of visual fields, hyperemia of optic disk, retinal edema;
- blind white discs, attenuated [illegible] of optic atrophy.

Pathology [pa]
- Meningeal petechia;
- cerebral edema;
- necrosis of retinal neurons;
- submucosal, subepicardial [illegible] hemorrhage;
- parenchymatous degenerat[illegible] kidney.

References [rf]
Dreisbach 131/32
Hunter 561 ff
Johnstone-miller 158/59
Plunkett 250/51
Thienes-haley 68