Office of Graduate Studies
University of South Florida
Tampa, Florida

CERTIFICATE OF APPROVAL

This is to certify that the thesis of

SAVVAS NIKIFOROU

in the graduate degree program of
Computer Science
was approved on January 30, 2002
for the Master of Science in Computer Science degree

Examining Committee:

Major Professor: Eugene Fink, Ph.D.

Member: Lawrence O. Hall, Ph.D.

Member: Dmitry B. Goldgof, Ph.D.

SELECTION OF CLINICAL TRIALS:

KNOWLEDGE REPRESENTATION AND ACQUISITION

by

SAVVAS NIKIFOROU

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Computer Science
Department of Computer Science and Engineering
College of Engineering
University of South Florida

May 2002

Major Professor: Eugene Fink, Ph.D.

Dedication

To Marina, for helping me fulfill my dreams, one at a time.

Acknowledgments

I gratefully acknowledge the help of Eugene Fink, who has supervised my thesis work and guided me through all steps of research and writing. I am grateful to Lawrence Hall and Dmitry Goldgof for their valuable guidance and feedback. I also thank Princeton Kokku, Tim Ivanovskiy, and Bhavesh Goswami for their comments and suggestions.

I have received invaluable support from my wife, Marina Nikiforou, who has provided help and encouragement through my graduate studies. I am also thankful to my parents, Christodoulos and Yiannoula Nikiforou, for their support and understanding. Finally, I would like to thank Josh Johnson and Mary Parrish for their feedback and help with grammar. The work has been partially sponsored by the Breast Cancer Research Program of the U.S. Army Medical Research and Materiel Command under contract DAMD17-00-1-0244.

Table of Contents

List of Figures

SELECTION OF CLINICAL TRIALS:

KNOWLEDGE REPRESENTATION AND ACQUISITION

by

SAVVAS NIKIFOROU

An Abstract

of a thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Computer Science
Department of Computer Science and Engineering
College of Engineering
University of South Florida

May 2002

Major Professor: Eugene Fink, Ph.D.

When medical researchers test a new treatment procedure, they recruit patients with appropriate health problems and medical histories. An experiment with a new procedure is called a *clinical trial.* The selection of patients for clinical trials has traditionally been a labor-intensive task, which involves matching of medical records with a list of eligibility criteria.

A recent project at the University of South Florida has been aimed at the automation of this task. The project has involved the development of an expert system that selects matching clinical trials for each patient. If a patient's data are not sufficient for choosing a trial, the system suggests additional medical tests.

We report the work on the representation and entry of the related selection criteria and medical tests. We first explain the structure of the system's knowledge base, which describes clinical trials and criteria for selecting patients. We then present an interface that enables a clinician to add new trials and selection criteria without the help of a programmer. Experiments show that the addition of a new clinical trial takes ten to twenty minutes, and that novice users learn the full functionality of the interface in about an hour.

Abstract Approved: ____

Major Professor: Eugene Fink, Ph.D.
Assistant Professor, Computer Science and Engineering

Date Approved: ______________________________

Chapter 1

Introduction

Cancer causes 550,000 deaths in the United States every year [Keppel *et al.*, 2002], and the treatment of cancer is an active research area. Medical experts explore new treatment methods, including drugs, surgery techniques, radiation therapies, and gene therapies. An experiment with a new treatment procedure is called a *clinical trial*. When researchers test a new procedure, they choose subjects with an appropriate cancer type and medical history. The selection of subjects has traditionally been a manual procedure, which involves significant human effort, and clinicians often miss eligible patients [Yusuf *et al.*, 1990; Kotwall *et al.*, 1992; Tu *et al.*, 1993; Séroussi *et al.*, 1999a; Gennari and Reddy, 2000].

A recent project at the University of South Florida has been aimed at automatic selection of patients for clinical trials. Fletcher and her colleagues have developed an expert system that prompts a clinician for a patient's data and identifies all matching trials [Bhanja *et al.*, 1998]. Experiments have shown that the system improves the matching accuracy and reduces human effort. Kokku *et al.* [2002a] have added a mechanism for ordering related medical tests; its purpose is to minimize the cost of tests involved in the selection process.

The system includes a knowledge base with information about available clinical trials, criteria for selecting patients, and related medical tests. When introducing new trials, clinicians need to add them to the knowledge base. Fletcher did not provide an interface for adding new trials, and she encoded the eligibility criteria in a special programming language. The time required to add a new trial varied from twenty to thirty hours. The language did not enforce standard encoding, and two programmers could produce incompatible descriptions of eligibility criteria.

We have designed a web-based interface that enables a clinician to add new trials without the help of a programmer. It has reduced the entry time from twenty hours to about twenty minutes. Furthermore, it converts the eligibility criteria into a standardized form and ensures compatibility of all knowledge in the system. We have used the interface to build a knowledge base for clinical trials at the Moffitt Cancer Center, located at the University of South Florida.

We begin with a review of previous work on medical expert systems (Chapter 2). We then explain the knowledge representation in the developed system (Chapter 3), describe the interface for adding new knowledge (Chapter 4), and conclude with a summary of the results and future challenges (Chapter 5).

Chapter 2

Previous Work

The automation of medical diagnosis and treatment selection is an important problem, and computer scientists have developed a variety of medical expert systems. They have created rule-based systems and Bayesian networks that capture expertise for several medical domains, covering bacterial diseases, cancer, asthma, liver diseases, and AIDS. We review some of these systems (Section 2.1) and related work on knowledge representation and acquisition (Sections 2.2 and 2.3).

2.1 Expert Systems

Researchers began to work on medical applications of artificial intelligence in the early seventies. Shortliffe and his colleagues developed the famous MYCIN system, which diagnosed bacterial diseases and suggested appropriate therapies [Shortliffe, 1974; Shortliffe *et al.*, 1975; Buchanan and Shortliffe, 1984]. It evolved from a chemical expert system, called DENDRAL [Lederberg, 1965; Buchanan *et al.*, 1969; Lederberg, 1987], that determined molecular structures based on spectrography results.

MYCIN's knowledge base consisted of *if-then* rules, which allowed the analysis of symptoms, selection of therapies, and evaluation of the selection certainty. For example, the system could determine that a patient with flu needed aspirin with 0.8 certainty. Experiments confirmed that MYCIN correctly diagnosed common diseases, which led to the development of other medical systems [Buchanan and Shortliffe, 1984; Musen, 1989], such as NEOMYCIN, PUFF, CENTAUR, and VM. Shortliffe *et al.* [1981]

created a system for selecting chemotherapy treatments, called ONCOCIN, which also evolved from MYCIN.

Lucas *et al.* [1989] constructed a rule-based system, called HEPAR, for diagnosing liver and biliary-tract diseases, but it often gave wrong diagnoses [Korver and Janssens, 1993; Onisko *et al.*, 1997]. Korver and Lucas [1993] converted the initial system into a Bayesian network, which improved its performance [Lucas, 1994].

Musen *et al.* [1996] built a rule-based system, called EON, that analyzed dependencies among the available data and assigned AIDS patients to clinical trials. For example, if an onset of low blood pressure coincided with the beginning of a new clinical trial, the system would notice that the trial may have caused the low pressure.

Ohno-Machado *et al.* [1993] developed the AIDS^2 system, which also matched AIDS patients to clinical trials. They integrated logical rules with Bayesian networks, which helped to make decisions in the absence of some data and to quantify the certainty of these decisions.

Bouaud *et al.* [1998; 2000] created a cancer expert system, called ONCODOC, that suggested alternative clinical trials for each patient, and allowed a physician to choose among them. It included a graphical interface for interactive entry of a patient's data and consideration of alternative trials. Séroussi *et al.* [1999a; 1999b; 2000; 2001a; 2001b] used ONCODOC to select participants for clinical trials at two hospitals, which helped to increase the number of selected patients by a factor of three.

Theocharous [1996] developed a Bayesian system that chose clinical trials for cancer patients. It learned conditional probabilities of medical-test outcomes and used them to evaluate the probability of a patient's eligibility for each trial [Papaconstantinou *et al.*, 1998]. On the negative side, the available medical records were insufficient for learning accurate probabilities. Furthermore, when a user added new clinical trials, he had to change the structure of the underlying Bayesian network, which was often a difficult task.

Hammond and Sergot [1996] built the OaSiS architecture, which combined the techniques from several earlier systems, including ONCOCIN and EON. It had a graphical interface for entering patients' data and extending the knowledge base.

Fallowfield *et al.* [1997] studied how physicians selected cancer patients for clinical trials, and compared manual and automatic selection. They showed that expert systems could improve the selection accuracy; however, their study also revealed that physicians were usually reluctant to use these systems. Carlson *et al.* [1995] conducted similar experiments with choosing participants for AIDS studies, and also concluded that expert systems could lead to a more accurate selection.

2.2 Knowledge Representation

Researchers have long realized the need for a general-purpose representation of medical knowledge [Clancey, 1993; Clancey, 1995] and investigated a variety of related representations.

In particular, Ohno-Machado *et al.* [1998] proposed a general format for medical knowledge, called the GuideLine Interchange Format. Their project involved researchers from Stanford, Harvard, and Columbia; they used the developed representation with a variety of algorithms, and concluded that it was sufficient for most medical knowledge. On the negative side, it did not enforce compatibility among knowledge bases developed by different researchers. Furthermore, it needed major improvements for representing conditional expressions, temporal reasoning, and uncertainty.

Lindberg *et al.* [1993] proposed an alternative general-purpose format, called the Unified Medical Language System, and developed tools for converting various medical databases into this format. Later, Le Duff *et al.* [2000] studied techniques for translating natural-language description of diseases into Lindberg's representation.

Rubin *et al.* [1999; 2000] analyzed selection criteria for clinical trials related to three cancer types and proposed a format for these criteria. They built a mechanism for encoding new criteria, which helped the users to avoid simple mistakes, such as missing or inconsistent selection rules.

Wang *et al.* [2001] compared eight previously developed formats and identified main elements of medical knowledge, which included patient data, treatment decisions, related actions, and a global state of an expert system. Wang also pointed out the need for abstraction and temporal reasoning.

2.3 Knowledge Acquisition

Early expert systems did not include knowledge-acquisition tools, and programmers hand-coded the related rules. To simplify knowledge entry, researchers implemented specialized tools for some systems. For example, Musen *et al.* [1988; 1989] developed the OPAL system for adding new knowledge to ONCOCIN, and Marcus and McDermott [1989] built the SALT system, which helped engineers to specify rules for elevator design.

Eriksson [1993] pointed out the need for general-purpose tools that would allow efficient knowledge acquisition, and described a system for building such tools. Tallis and his colleagues developed a library of scripts for modifying knowledge bases, which helped to enforce the consistency of the modified knowledge [Gil and Tallis, 1997; Tallis, 1998; Tallis and Gil, 1999; Tallis *et al.*, 1999]. Kim and Gil [2000a; 2000b] considered the use of scripts for building new knowledge-acquisition tools, and created a system for evaluating these tools. Blythe *et al.* [2001] designed a general knowledge-acquisition interface based on previous techniques.

Musen and his colleagues developed the PROTÉGÉ environment for creating knowledge bases [Musen, 1989]; later, researchers used it in the work on AIDS expert systems [Puerta *et al.*, 1992a; Puerta *et al.*, 1992b], asthma treatment selection [Johnson and Musen, 1996], and elevator-design rules [Rothenfluh *et al.*, 1996]. Musen *et al.* [2000] extended PROTÉGÉ and built a new version, called PROTÉGÉ-2000.

Several researchers have studied techniques for extracting medical knowledge from natural-language documents. In particular, Hahn and Schnattinger [1997a; 1997b; 1998] built a parser for processing German medical texts on gastro-intestinal diseases. Romacker and Hahn [2001] improved the parser and showed that the resulting accuracy of semantic representations was between 80% and 93%; however, its effectiveness in constructing knowledge bases was very low.

Researchers have also considered data-mining techniques for learning medical knowledge from clinical databases [Cimino *et al.*, 1988; Shusaku, 1998; Mendonça and Cimino, 2000]. Although these techniques generated basic diagnosis rules, they allowed the correct diagnosis only in 8% of the test cases.

The reader can find a more detailed review of the work on knowledge entry in the book by Boose and Gains [1990], who described knowledge-acquisition tools not only for medical systems but also for other applications. Ringland and Duce [1988] presented standard techniques for knowledge representation, including functional approaches and temporal reasoning. Price [1990] also reviewed general tools for knowledge representation and acquisition.

Chapter 3

Selection of Clinical Trials

Physicians at the Moffitt Cancer Center have about 150 clinical trials available for cancer patients. They have identified criteria that determine a patient's eligibility for each clinical trial, and they use these criteria to select appropriate trials for eligible patients. Traditionally, physicians have selected trials by a manual analysis of a patient's data. The review of resulting selections has shown that they usually do not check all clinical trials and occasionally miss an appropriate trial.

To address this problem, Fletcher and her colleagues built a system for automatic selection of clinical trials, and a knowledge base for breast cancer [Bhanja *et al.*, 1998]. Their system prompted a clinician to enter the results of medical tests for a patient, and identified appropriate trials. It selected the order of questions that minimized the expected amount of data entry. For instance, if some question could reveal that a patient was not eligible for any trial, the system asked it before the other questions. Experiments showed that the system reduced the human effort involved in trial selection and helped to avoid inaccuracies. Kokku has continued this work and added information about the costs of medical tests and the pain levels of related procedures [Kokku *et al.*, 2002b]. He has developed a technique for finding the order of test procedures that reduces the expected cost and pain.

We review Kokku's system for selection of clinical trials. We begin with an example of the selection process (Section 3.1), describe the main elements of the knowledge base (Section 3.2), and explain heuristics for test ordering (Section 3.3).

General information
1. The patient is female.
2. She is at most forty-five years old.

Mammogram: Cost is $150, pain level is 1
3. Cancer stage is either II or III.
4. Cancer is not invasive.

Electrocardiogram: Cost is $160, pain level is 1
5. The patient has no congenital heart disease.
6. The patient has no cardiac arrhythmias.

Biopsy: Cost is $200, pain level is 3
7. At most three lymph nodes have tumor cells.
8. All tumors are smaller than three centimeters.

Figure 3.1: Eligibility criteria for Clinical Trial A.

3.1 Example

In Figure 3.1, we give a simplified example of eligibility criteria for a certain clinical trial, called Trial A. We can use this trial for young and middle-age women with a noninvasive breast cancer at stage II or III. A patient is eligible if she has at most three affected lymph nodes, all her tumors are smaller than three centimeters, and she has no heart problems.

When a clinician tests a patient's eligibility for this trial, he has to order three medical tests. To check Conditions 3 and 4, a clinician sends a patient for a mammogram, which is almost painless and costs $150. If the patient meets these conditions, she needs an electrocardiogram, which is the next cheapest test. Finally, if she satisfies Conditions 5 and 6, the clinician sends her for a biopsy, which is an expensive and painful procedure.

The system first prompts a clinician to enter the patient's sex and age (Figure 3.2a). If the patient satisfies the corresponding conditions, the system asks for the mammogram results (Figure 3.2b), and the clinician orders a mammogram. Then, the system requests the electrocardiogram (Figure 3.2c) and biopsy (Figure 3.2d).

What is the patient's sex?	Female Male
What is the patient's age?	

(a) General questions.

What is the cancer stage?	I II III IV		
Does the patient have invasive cancer?	Yes	No	Unknown

(b) Mammogram results.

Does the patient have congenital heart disease?	Yes	No	Unknown
Does the patient have cardiac arrhythmias?	Yes	No	Unknown

(c) Electrocardiogram results.

How many lymph nodes have tumor cells?	
What is the greatest tumor diameter?	

(d) Biopsy results.

Figure 3.2: Example questions. The system guides a clinician through a series of questions, grouped by test procedures, and uses the answers to select appropriate clinical trials.

If the clinician has information about some of the patient's old tests, he may answer the corresponding questions along with entering personal data, before the system selects test procedures. For example, if he knows that the patient has invasive cancer, he may enter it along with sex and age, and then the system immediately rejects Trial A.

In Figure 3.3, we give another example of eligibility criteria, and we refer to the corresponding clinical trial as Trial B. If both trials are in the knowledge base, the system can check whether a patient is eligible for either of them. First, it prompts the clinician to enter the general information (Figure 3.4a), and then asks for the mammogram results, which are relevant to both trials (Figure 3.4b). If the results satisfy the eligibility criteria for Trial B, the system requests the liver-test data (Figure 3.4c), and then outputs the decision for Trial B. To determine the eligibility for Trial A, it requests the electrocardiogram data (Figure 3.4d). If the results satisfy the

General information
1. The patient is female.
2. She is at least twenty-seven years old.

Mammogram: Cost is $150, pain level is 1
3. Cancer stage is III.
4. Cancer is not recurrent.

Liver test: Cost is $150, pain level is 1
5. The patient has no hepatitis B.
6. The patient has no liver infections.

Figure 3.3: Eligibility criteria for Clinical Trial B.

eligibility criteria, the system asks for the biopsy data (Figure 3.4e) and then outputs the decision for Trial A.

3.2 Knowledge Base

The system's knowledge base includes questions, medical procedures, and logical expressions that represent eligibility conditions.

Questions. The system supports three types of questions. The first type takes a yes/no response, the second is a multiple choice, and the third requires a numeric answer. When the system asks a yes/no question, it accepts one of three answers: YES, NO, or UNKNOWN. The user can disable the *unknown* option for some questions; for example, we do not accept *unknown* for the electrocardiogram results in Figure 3.4(d). When the clinician gets a multiple-choice question, such as a cancer stage, he has to select one of the available answers (Figure 3.4b). An answer to a numeric question is a real value, which must be within the legal range for this question; for example, a patient's age is between 0 and 150 (Figure 3.4a), and a tumor diameter is between 0 and 25 centimeters (Figure 3.4e).

What is the patient's sex?	Female Male
What is the patient's age?	

(a) General questions.

What is the cancer stage?	I II III IV		
Does the patient have invasive cancer?	Yes ◯	No ◯	Unknown ◯
Does the patient have recurrent cancer?	Yes ◯	No ◯	Unknown ◯

(b) Mammogram results, relevant to both trials.

Does the patient have hepatitis B?	Yes ◯	No ◯	Unknown ◯
Does the patient have liver infections?	Yes ◯	No ◯	Unknown ◯

(c) Liver-test results, relevant to Trial B.

Does the patient have congenital heart disease?	Yes ◯	No ◯
Does the patient have cardiac arrhythmias?	Yes ◯	No ◯

(d) Electrocardiogram results, relevant to Trial A.

How many lymph nodes have tumor cells?	
What is the greatest tumor diameter?	

(e) Biopsy results, relevant to Trial A.

Figure 3.4: Checking the eligibility for two clinical trials. The system begins with the questions related to both trials.

Medical tests. The description of a medical test includes the test name, dollar cost, estimated pain level, and list of questions that can be answered based on the test results. For example, the mammogram in Figure 3.1 has a cost of $150 and pain level of 1, and it provides data for Criteria 3 and 4. Two different tests may answer the same question; for instance, both the mammogram and the biopsy show the cancer stage.

Eligibility criteria. We encode eligibility for a clinical trial by a logical expression that does not have negations, called the *acceptance expression.* It includes variables that represent the available data, as well as equalities, inequalities, "set-element"

sex = FEMALE `and`	*sex* = MALE `or`
age $\in [0, 45]$ `and`	*age* $\in (45, 150]$ `or`
cancer-stage $\in$ {II, III} `and`	*cancer-stage* $\in$ {I, IV} `or`
invasive-cancer = NO `and`	*invasive-cancer* $\in$ {YES, UNKNOWN} `or`
lymph-nodes $\in [0, 3]$ `and`	*lymph-nodes* $\in (3, 100]$ `or`
tumor-diameter $\in [0, 3]$ `and`	*tumor-diameter* $\in (3, 25]$ `or`
heart-disease = NO `and`	*heart-disease* $\in$ {YES, UNKNOWN} `or`
cardiac-arrhythmias = NO	*cardiac-arrhythmias* $\in$ {YES, UNKNOWN}
(a) Acceptance expression.	(b) Rejection expression.

Figure 3.5: Logical expressions for the criteria in Figure 3.1. The acceptance expression represents the eligibility conditions (a), whereas the rejection expression describes ineligible patients (b).

relations, conjunctions, and disjunctions. For example, we encode the criteria in Figure 3.1 by the expression given in Figure 3.5(a). In addition, the system uses the logical complement of the eligibility criteria, called the *rejection expression,* which also does not include negations (Figure 3.5b). It describes the conditions that make a patient ineligible for the clinical trial.

The system collects data until it can determine which of the two expressions is TRUE. For example, if the patient's sex is MALE, then the rejection expression in Figure 3.5(b) is TRUE, and the system immediately determines that this trial is inappropriate. On the other hand, if the sex is FEMALE, and the other values are unknown, then neither expression is TRUE, and the system has to ask more questions.

3.3 Order of Tests

When a clinician enters medical data for a patient, the system identifies all appropriate trials. The total cost and pain level of the tests involved in the trial selection may depend on their ordering. For instance, if we begin with the mammogram, and it shows that the cancer stage is I, then we can immediately reject the trial in Figure 3.1 and avoid more expensive tests.

Kokku *et al.* [2002a; 2002b] have studied heuristics for ordering the tests; their heuristics account for the cost and pain level of tests, the structure of acceptance and rejection expressions, and the number of expressions that require each test. The heuristics use a disjunctive normal form of acceptance and rejection expressions; that is, each expression must be a disjunction of conjunctions.

Kokku has defined the overall "payment" for medical tests as a linear combination of their costs and pain levels; that is, if a patient needs n tests, the payment is

$$a \cdot \sum_{i=1}^{n} cost_i + b \cdot \sum_{i=1}^{n} pain_i.$$

A user sets the values of a and b, and the system chooses the order of questions that reduces the expected payment. After getting the results of the first test, it re-evaluates the need for other tests and revises their ordering. The choice of the first test is based on three criteria.

1. *Cost and pain level of the test.* The system gives preference to tests with smaller payments. For example, it may start with the mammogram, which is cheaper and less painful than the other two tests in Figure 3.1.

2. *Number of clinical trials that require the test.* When the system checks a patient's eligibility for several trials, it prefers tests that provide data for larger number of trials. For example, if the electrocardiogram gives data for three trials, the system may prefer it to the mammogram despite its higher cost.

3. *Immediate decisions for some trials.* If a test can lead to an immediate acceptance or rejection of some trials, the system prefers it to other tests. For instance, if the liver test shows that the patient has hepatitis B, the system can immediately reject Trial B.

To select the first test, the system scores all required tests according to the three criteria. It computes a linear combination of these three scores for every test, and chooses the test with the highest score. Kokku *et al.* [2002b] have evaluated this strategy using retrospective data for 187 patients at the Moffitt Cancer Center, and demonstrated that it significantly reduces the cost.

Chapter 4

Entering Eligibility Criteria

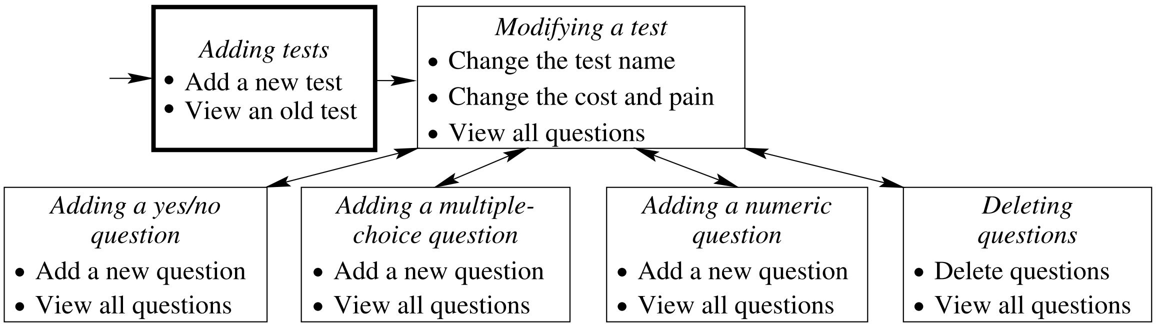
When Fletcher developed the initial system, she did not include an interface for adding eligibility criteria, and a programmer had to encode the criteria in a specialized language. We have designed a web-based interface for adding new criteria, which consists of two main parts; the first part is for entering information about medical tests (Figure 4.1), and the second is for specifying eligibility criteria (Figure 4.2).

The interface includes fifteen screens; three of them are "start screens," which can be reached directly from any other screen. For example, consider the *"Modifying a test"* screen in Figure 4.3, which allows changing the test name, cost, and pain level. It has four buttons at the bottom for moving to related screens, and three buttons on the left for moving to the start screens.

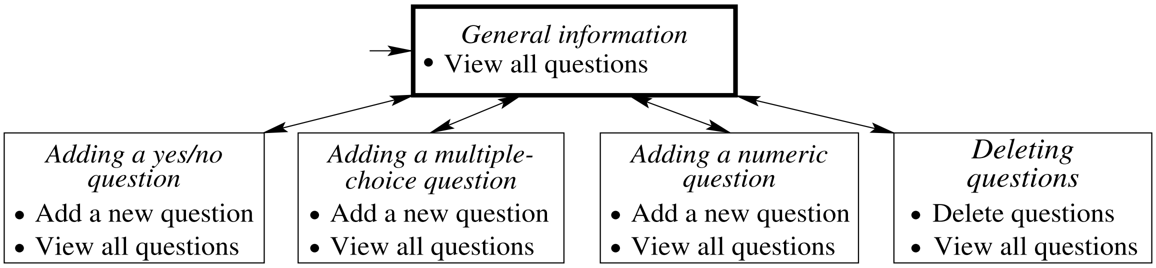
We give an example of entering eligibility criteria (Section 4.1), describe the two main parts of the interface (Sections 4.2 and 4.3), give algorithms for generating acceptance and rejection expressions (Section 4.4), and present experiments on the effectiveness of the interface (Section 4.5).

4.1 Example

Suppose that the user needs to enter the clinical trials in Figures 3.1 and 4.4, and the system initially has no information about the related tests. The user has to describe the tests and questions, and then specify the eligibility conditions. We assume that he first enters the trial in Figure 3.1, and later adds the trial in Figure 4.4.



(a) Tests and questions.



(b) General questions.

Figure 4.1: Entering tests and questions. We show the screens by rectangles and transitions between them by arrows. The bold rectangles are the start screens.

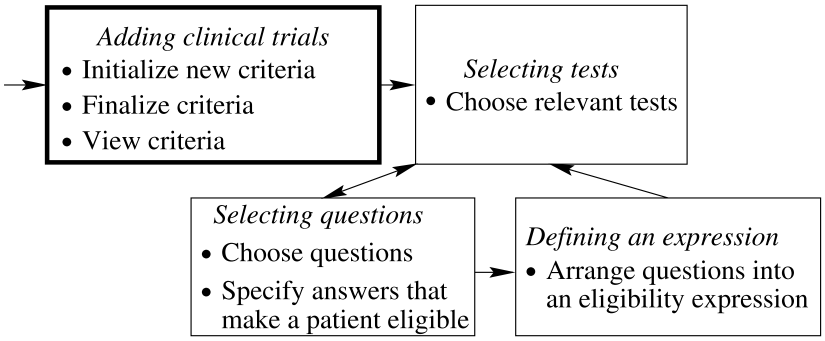


Figure 4.2: Entering eligibility criteria.

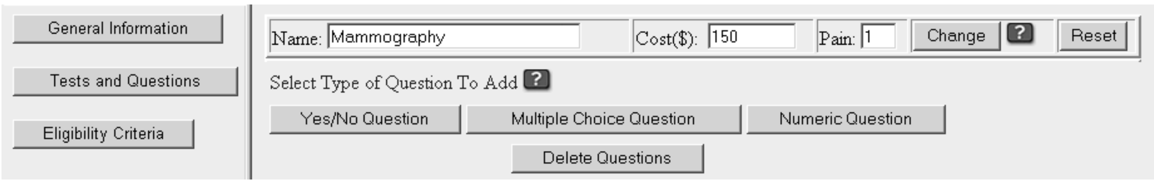


Figure 4.3: *"Modifying a test"* screen. The three buttons on the left are for moving to the start screens; every screen in the system has these buttons.

First, he uses the *"Adding tests"* screen to enter the new tests; we illustrate the entry of two tests in Figure 4.5. Then, he enters the related questions; to enter questions for a specific test, he selects the test and clicks *"Modify"* (Figure 4.6), which takes him to the *"Modifying a test"* screen (Figure 4.7).

To add a question, the user clicks the appropriate button at the bottom (Figure 4.7) and then types the question (Figure 4.8). For a multiple-choice question, he has to include the answer options (Figure 4.8b); for a numeric question, he needs to specify the range of allowed values (Figure 4.8c). If other tests provide data for the same question, the user has to select all related tests in the lower box (Figure 4.8b). The newly added questions appear on the *"Modifying a test"* screen (Figure 4.9).

After adding the questions for all tests, the user goes to the *"Adding clinical trials"* screen, initializes a new trial (Figure 4.10), and selects it for adding eligibility conditions. He gets the *"Selecting tests"* screen and chooses the tests related to the trial (Figure 4.11). Then, he selects relevant questions and the answers that make a patient eligible (Figure 4.12).

Now suppose that the user needs to add the clinical trial in Figure 4.4. The new eligibility conditions require a liver test, which is not in the knowledge base, and the user has to add the information related to this test. Furthermore, he has to add the question about recurrent cancer to the mammography test. After making these additions, he is ready to enter the eligibility criteria.

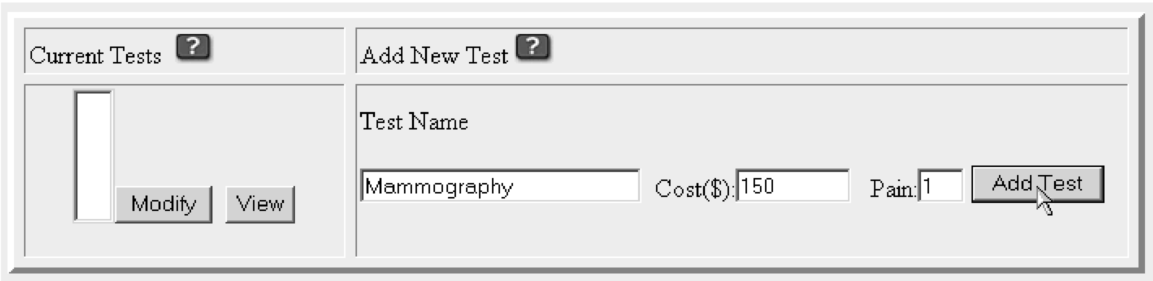
Condition 5 includes a disjunction, which requires the *"Combined question"* option at the bottom of the questions screen (Figure 4.12). The user checks the elements of the disjunctive question, marks the appropriate answers, and clicks *"Combined question,"* which takes him to the screen for composing logical expressions (Figure 4.13). After entering Condition 5, he adds the other criteria using the *"Simple questions"* option (Figure 4.12).

General information
1. The patient is female.
2. She is at least twenty-seven years old.

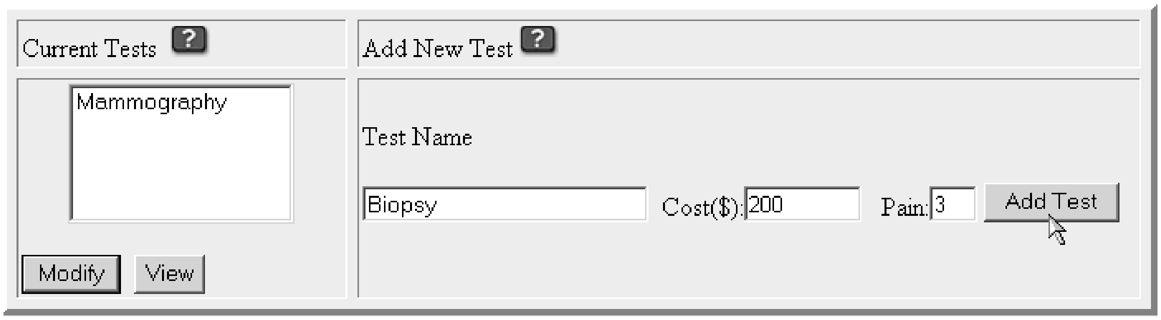
Mammogram: Cost is $150, pain level is 1
3. Cancer stage is III.
4. Cancer is not recurrent.
5. Either
   - the tumor is at least two centimeters, or
   - the cancer is not invasive and
     at least two lymph nodes have tumor cells.

Liver test: Cost is $150, pain level is 1
6. The patient has no hepatitis B.
7. The patient has no liver infections.

Figure 4.4: Eligibility criteria with a disjunctive condition.



(a) Mammography test.



(b) Biopsy test.

Figure 4.5: Adding new tests.

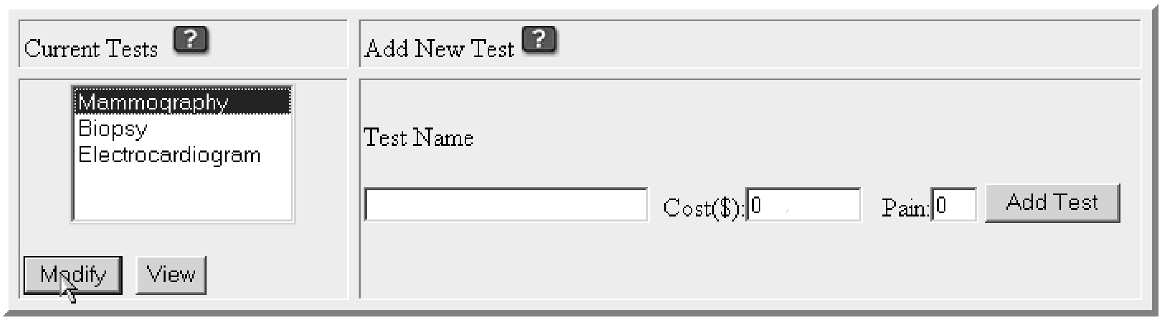


Figure 4.6: Selecting a test for entering the related questions.

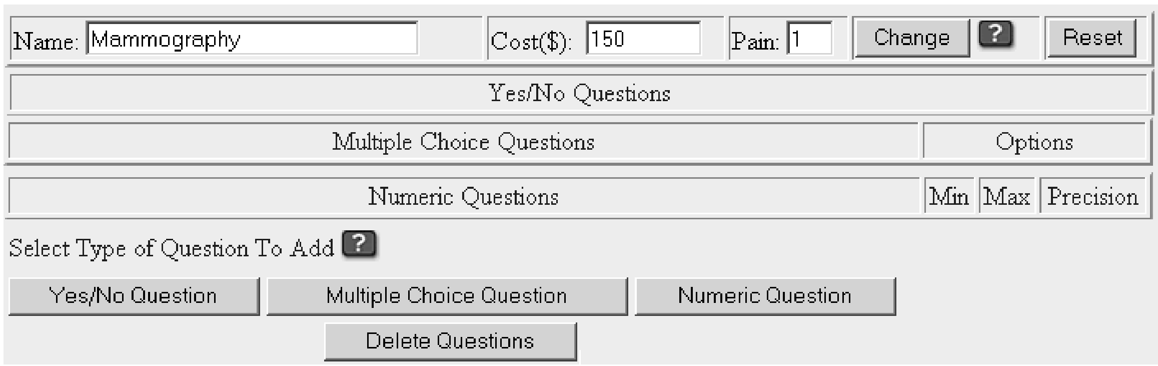
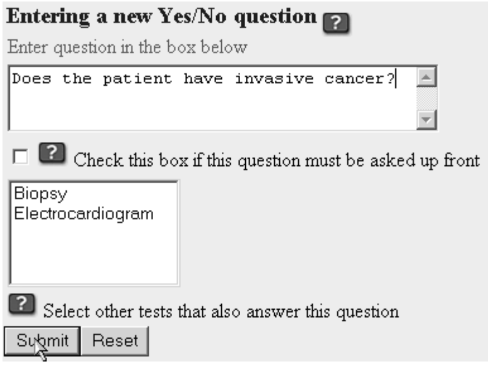
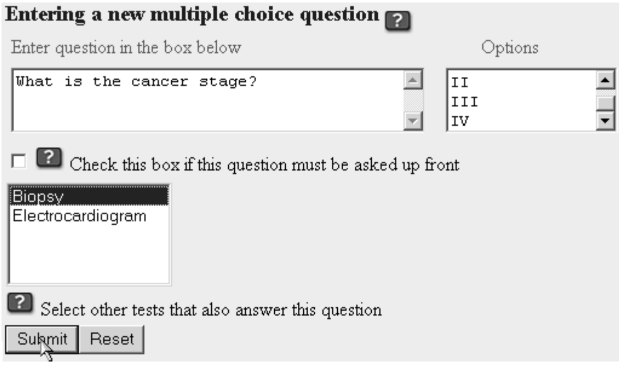


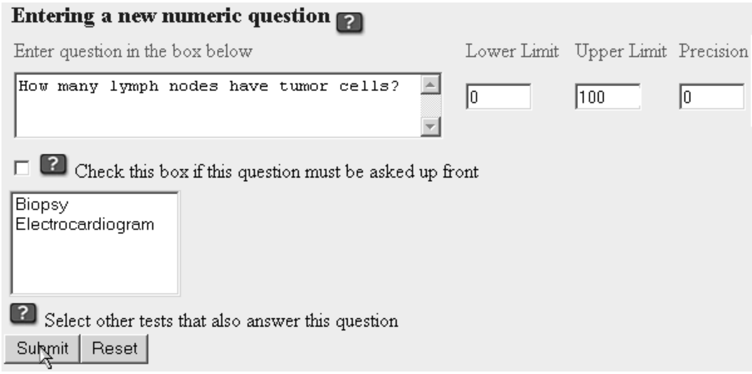
Figure 4.7: *"Modifying a test"* screen. The system has no information about related questions, and the user clicks one of the bottom buttons for moving to a question-entry screen.



(a) Yes/no question.



(b) Multiple-choice question.



(c) Numeric question.

Figure 4.8: Adding new questions. The user types a question and answer options. If the question is related to several tests, the user should check all these tests.

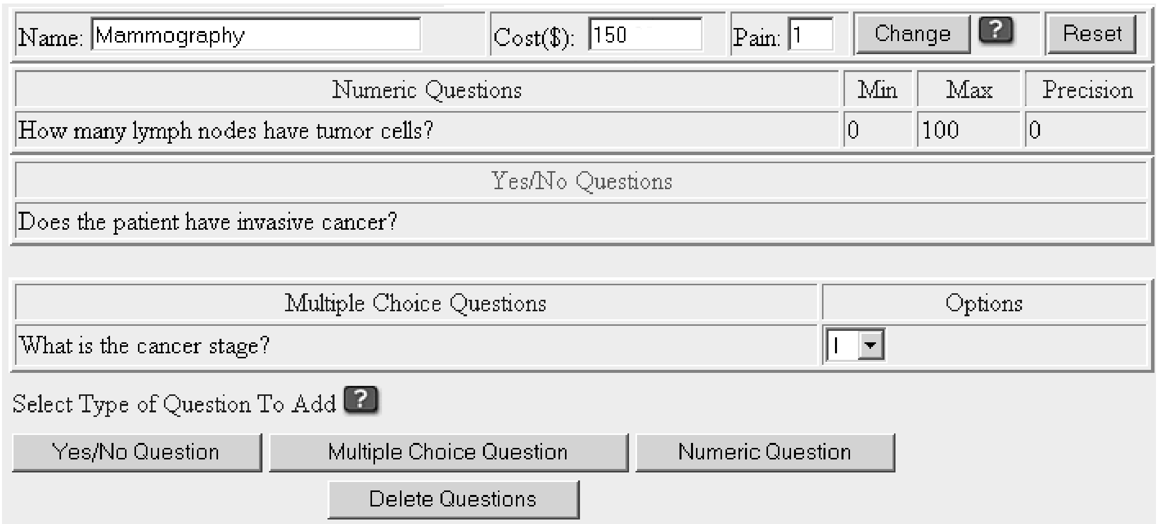


Figure 4.9: *"Modifying a test"* screen with a list of questions.

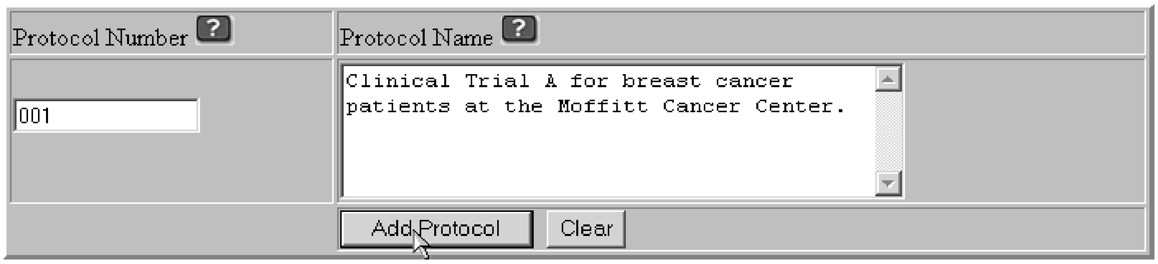


Figure 4.10: Adding a new clinical trial.

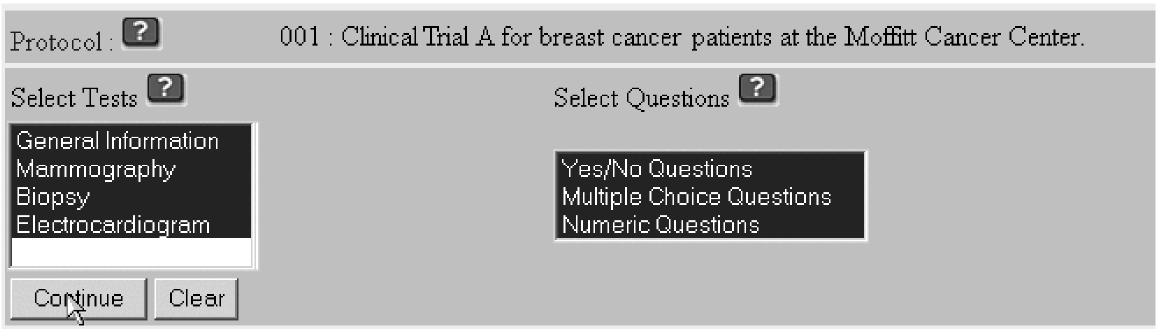


Figure 4.11: Choosing tests and question types.

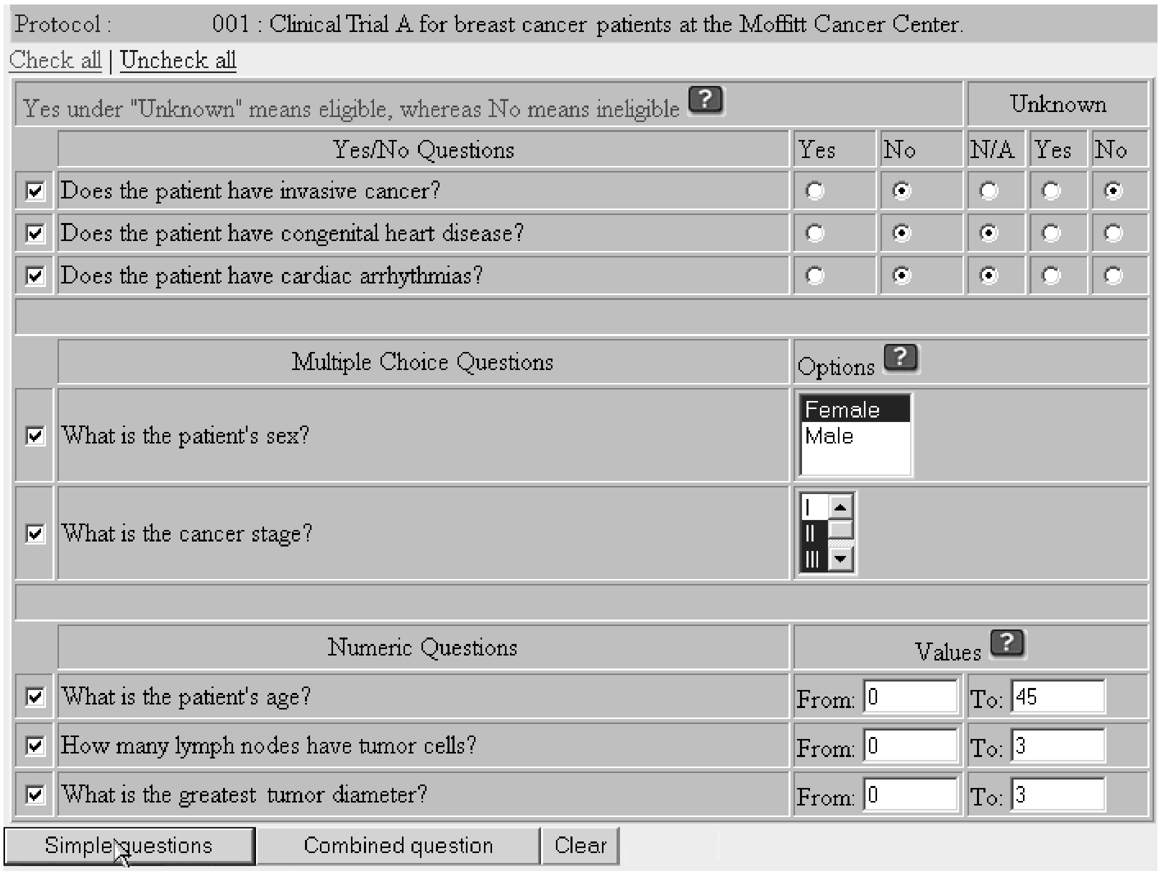


Figure 4.12: Selecting questions and answers. The user checks the questions for the current clinical trial and marks the answers that satisfy the eligibility criteria.

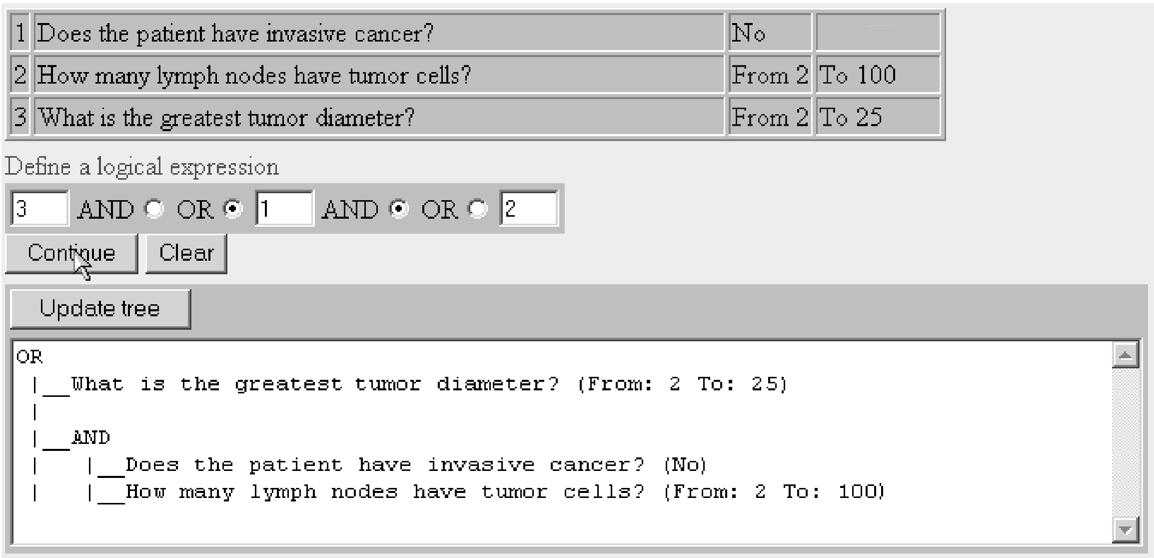


Figure 4.13: Combining questions into a logical expression.

4.2 Tests and Questions

We now describe the six-screen interface for adding tests and questions (Figure 4.1a). The start screen allows viewing the available tests and defining new ones, whereas the other screens are for modifying tests and adding related questions.

Adding tests. We show the start screen in Figure 4.5; its left-hand side allows viewing questions and going to a modification screen. If the user selects a test and clicks *"View,"* the system shows the related questions at the bottom of the same screen (Figure 4.14). If he clicks *"Modify,"* it displays the *"Modifying a test"* screen (Figure 4.7). The right-hand side of the start screen allows adding a new test by specifying its name, cost, and pain level.

Modifying a test. The test-modification screen shows the information about a specific test, which includes the test name, cost, pain level, and related questions (Figure 4.9). The user can change the test name, cost, and pain level by entering new values and clicking *"Change."* The four bottom buttons allow moving to the screens for adding new questions and deleting old questions.

Adding a question. We show the screen for adding yes/no questions in Figure 4.8(a), multiple-choice questions in Figure 4.8(b), and numeric questions in Figure 4.8(c). The user can enter a new question for the current test, along with a set of allowed answers. If the question is also related to other tests, the user has to mark them in the lower box (Figure 4.8b).

Deleting questions. This screen (Figure 4.15) is for removing old and incorrectly entered questions. The user has to mark unwanted questions and click *"Delete."*

General information. The general questions include sex, age, and other personal data, collected without medical tests. The mechanism for adding such questions consists of five screens (Figure 4.1b), and the user adds general questions in the same way as test-related questions.

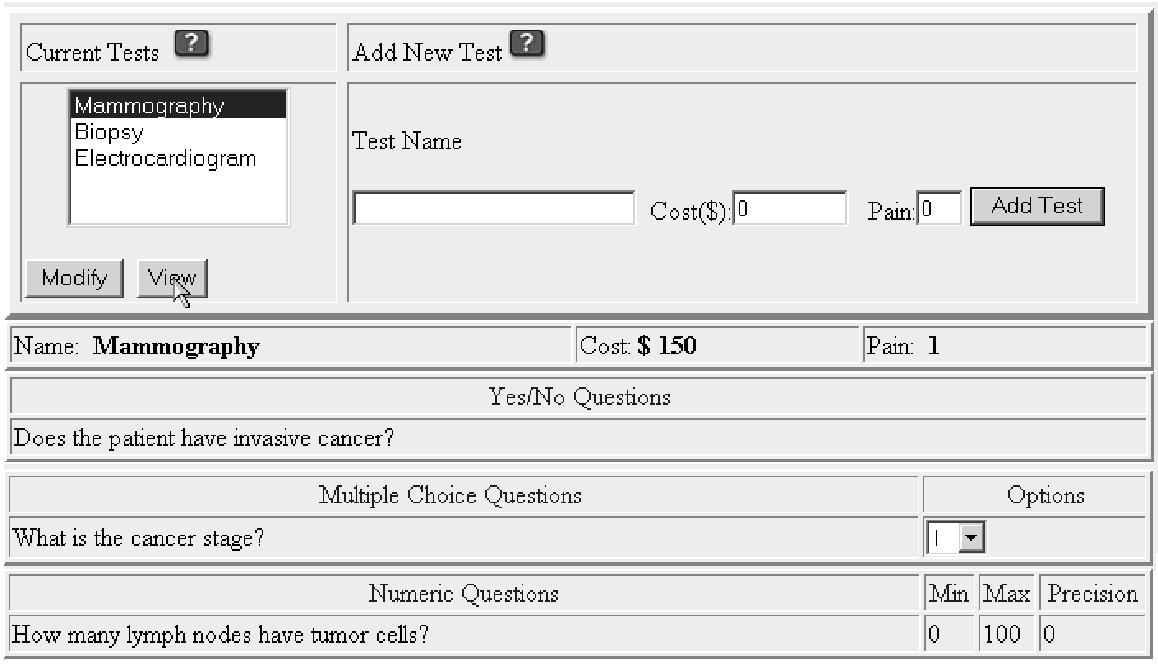


Figure 4.14: Viewing the questions for a specific test.

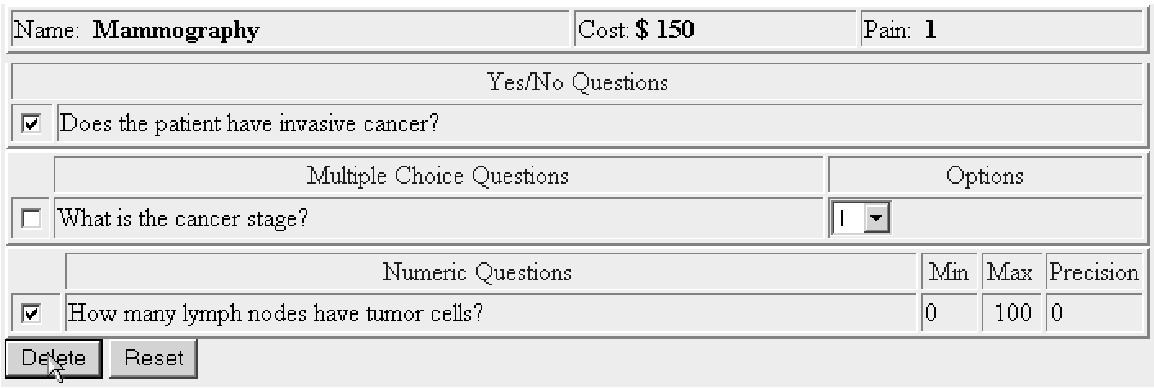


Figure 4.15: Deleting questions.

4.3 Eligibility Conditions

We next describe the mechanism for entering eligibility criteria, which consists of four screens (Figure 4.2).

Adding clinical trials. The start screen (Figure 4.16) allows the user to initialize a new clinical trial, view the criteria for old trials, and finalize completed trials. The lower part of the screen is for initializing a new trial, which requires entering the trial's name and unique number. The upper part is a list of trials with unfinished eligibility criteria. The user can view the questions for an unfinished trial by clicking *"View,"* and he can go to a modification screen by clicking *"Modify."* After completing the eligibility criteria, the user finalizes the trial by clicking *"Activate."* The list of finalized trials is in the middle of the screen; the user can view these trials, but he cannot modify them.

Selecting tests. If the user clicks *"Modify"* on the start screen, the system displays the test-selection screen (Figure 4.11). The user then chooses related tests and question types, and clicks *"Continue"* to get the question list. For instance, if he chooses mammogram and biopsy on the left, and the top two question types on the right, then he gets a list of all yes/no and multiple-choice questions related to the mammogram and biopsy.

Selecting questions. The next screen (Figure 4.12) allows the user to select specific questions and mark answers that make a patient eligible. For a multiple-choice question, the user may specify several eligibility options; for example, a patient may be eligible if her cancer stage is II or III. For a numeric question, the user has to specify a range of values; for example, a patient may be eligible if her age is between 0 and 45 years. If the user clicks *"Simple questions,"* the system generates a conjunction of the selected criteria. If the eligibility conditions involve a more complex expression, the user has to click *"Combined question,"* which takes him to the screen for composing logical expressions.

Defining an expression. This screen (Figure 4.13) allows the user to arrange the selected questions into an expression that includes nested conjunctions and disjunctions; however, the system does not allow negations.

4.4 Logical Expressions

When the user finalizes a clinical trial, the system combines the eligibility criteria into an acceptance expression, and then generates the corresponding rejection expression. In Figure 4.17, we give an algorithm that constructs the rejection expression by recursive application of DeMorgan's laws; the resulting expression does not include negations.

If the system uses the ordering heuristics described in Section 3.3, it has to convert the acceptance and rejection expressions into a disjunctive normal form, that is, a disjunction of conjunctions; we use a standard conversion algorithm [Kenneth, 1988; Crama and Hammer, 2001], summarized in Figure 4.18. For instance, if the eligibility criteria are as shown in Figure 4.19(a), the system generates the acceptance expression in Figure 4.19(b) and the rejection expression in Figure 4.19(c).

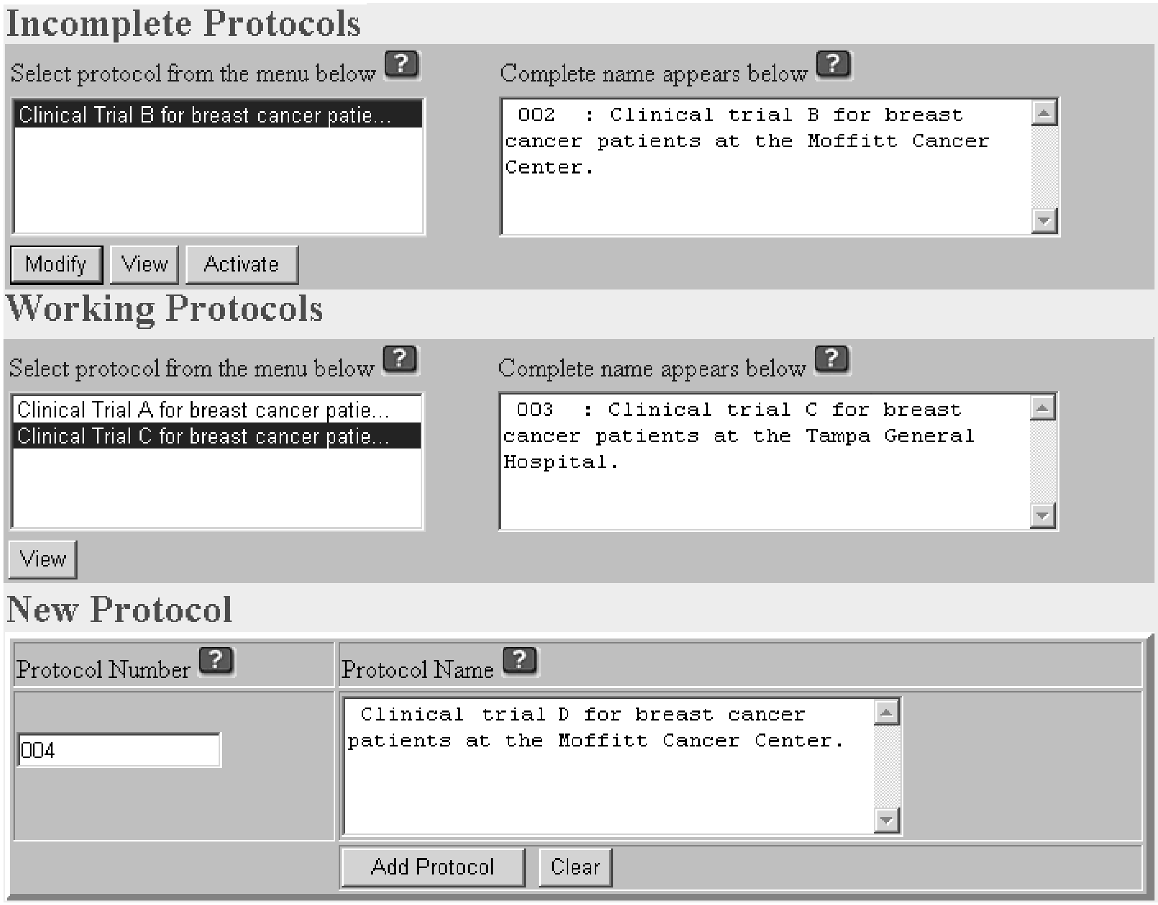


Figure 4.16: *"Adding clinical trials"* screen. It allows the user to add new trials (bottom part), modify and finalize eligibility criteria (top), and view the finalized criteria (middle).

Negate-Expression(*bool-exp*)
The input is a logical expression, *bool-exp*, which represents eligibility criteria.

Determine whether *bool-exp* is a conjunction, disjunction,
yes/no question, multiple-choice question, or numeric question.
Call the appropriate subroutine below and return the resulting expression.

Negate-Conjunction(*bool-exp*)
The input is a conjunctive expression; that is, *bool-exp* is "*sub-exp* and *sub-exp* and ..."

New-Exps := $\emptyset$.
For every term *sub-exp* of the conjunction *bool-exp*:
New-Exps := *New-Exps* $\cup$ {*Negate-Expression*(*sub-exp*)}.
Return the disjunction of all terms in *New-Exps*.

Negate-Disjunction(*bool-exp*)
The input is a disjunctive expression; that is, *bool-exp* is "*sub-exp* or *sub-exp* or ..."

New-Exps := $\emptyset$.
For every term *sub-exp* of the disjunction *bool-exp*:
New-Exps := *New-Exps* $\cup$ {*Negate-Expression*(*sub-exp*)}.
Return the conjunction of all terms in *New-Exps*.

Negate-Yes-No(*bool-exp*)
The input is a yes/no question.

If *bool-exp* is "*Variable* = YES," then return "*Variable* $\in$ {NO, UNKNOWN}."
If *bool-exp* is "*Variable* = NO," then return "*Variable* $\in$ {YES, UNKNOWN}."
If *bool-exp* is "*Variable* $\in$ {YES, UNKNOWN}," then return "*Variable* = NO."
If *bool-exp* is "*Variable* $\in$ {NO, UNKNOWN}," then return "*Variable* = YES."

Negate-Multiple-Choice(*bool-exp*)
The input is a multiple-choice question; that is, *bool-exp* is "*Variable* $\in$ *Option-Set*."

Let *All-Options* be the set of all answer options for *Variable*.
New-Options := *All-Options* − *Option-Set*.
(This set difference includes all answers that are not in *Option-Set*.)
Return "*Variable* $\in$ *New-Options*."

Negate-Numeric(*bool-exp*)
The input is a numeric question; that is, *bool-exp* is "*Variable* $\in [Min, Max]$."

Let "[*Lower*, *Upper*]" be the range of allowed values for *Variable*;
that is, the value of *Variable* is always between *Lower* and *Upper*.
If *Min* = *Lower*, then return "*Variable* $\in$ (*Max*, *Upper*]."
If *Max* = *Upper*, then return "*Variable* $\in$ [*Lower*, *Min*)."
Return "*Variable* $\in$ [*Lower*, *Min*) $\cup$ (*Max*, *Upper*]."

Figure 4.17: Constructing a rejection expression. The *Negate-Expression* procedure inputs an acceptance expression and recursively processes its subexpressions.

Normalize(*bool-exp*)
The input is a logical expression, *bool-exp*; the output
is an equivalent expression in disjunctive normal form.

If *bool-exp* is an equality, inequality, or "set-element" test,
  then return *bool-exp*.
If *bool-exp* is a disjunction $sub\text{-}exp_1 \lor sub\text{-}exp_2$,
  then $norm\text{-}exp_1 := Normalize(sub\text{-}exp_1)$;
    $norm\text{-}exp_2 := Normalize(sub\text{-}exp_2)$;
    return $norm\text{-}exp_1 \lor norm\text{-}exp_2$.
If *bool-exp* is a conjunction $sub\text{-}exp_1 \land sub\text{-}exp_2$,
  then $norm\text{-}exp_1 := Normalize(sub\text{-}exp_1)$;
    $norm\text{-}exp_2 := Normalize(sub\text{-}exp_2)$;
    return $Merge(norm\text{-}exp_1, norm\text{-}exp_2)$.

Merge($norm\text{-}exp_1$, $norm\text{-}exp_2$)
The input is two logical expressions, $norm\text{-}exp_1$ and $norm\text{-}exp_2$, in disjunctive normal form;
the output is a disjunctive normal form of their conjunction, $norm\text{-}exp_1 \land norm\text{-}exp_2$.

$New\text{-}Exps := 0$.
For every term $sub\text{-}exp_1$ of $norm\text{-}exp_1$:
  For every term $sub\text{-}exp_2$ of $norm\text{-}exp_2$:
    $New\text{-}Exps := New\text{-}Exps \cup \{sub\text{-}exp_1 \land sub\text{-}exp_2\}$.
Return the disjunction of all terms in *New-Exps*.

Figure 4.18: Converting an expression into a disjunctive normal form. The *Normalize* procedure inputs an expression without negations, which represents acceptance or rejection conditions, and generates an equivalent expression in disjunctive normal form.

$$\begin{array}{l}
\textit{sex} = \text{FEMALE}\ \texttt{and} \\
\textit{age} \in [27, 150]\ \texttt{and} \\
\textit{cancer-stage} = \text{III}\ \texttt{and} \\
\textit{recurrent} = \text{NO}\ \texttt{and} \\
(\textit{tumor-size} \in [2, 25]\ \texttt{or} \\
\quad (\textit{invasive} = \text{NO}\ \texttt{and}\ \textit{lymph-nodes} \in [2, 100]))\ \texttt{and} \\
\textit{hepatitis} = \text{NO}\ \texttt{and} \\
\textit{liver-infections} = \text{NO}
\end{array}$$

(a) Eligibility criteria.

$$\left(\begin{array}{l}
\textit{sex} = \text{FEMALE}\ \texttt{and} \\
\textit{age} \in [27, 150]\ \texttt{and} \\
\textit{cancer-stage} = \text{III}\ \texttt{and} \\
\textit{recurrent} = \text{NO}\ \texttt{and} \\
\textit{tumor-size} \in [2, 25]\ \texttt{and} \\
\textit{hepatitis} = \text{NO}\ \texttt{and} \\
\textit{liver-infections} = \text{NO}
\end{array}\right)\ \texttt{or}\ \left(\begin{array}{l}
\textit{sex} = \text{FEMALE}\ \texttt{and} \\
\textit{age} \in [27, 150]\ \texttt{and} \\
\textit{cancer-stage} = \text{III}\ \texttt{and} \\
\textit{recurrent} = \text{NO}\ \texttt{and} \\
\textit{invasive} = \text{NO}\ \texttt{and} \\
\textit{lymph-nodes} \in [2, 100]\ \texttt{and} \\
\textit{hepatitis} = \text{NO}\ \texttt{and} \\
\textit{liver-infections} = \text{NO}
\end{array}\right)$$

(b) Acceptance expression.

$$\begin{array}{l}
\textit{sex} = \text{MALE}\ \texttt{or} \\
\textit{age} \in [0, 27)\ \texttt{or} \\
\textit{cancer-stage} \in \{\text{I, II, IV}\}\ \texttt{or} \\
\textit{recurrent} \in \{\text{YES, UNKNOWN}\}\ \texttt{or} \\
(\textit{tumor-size} \in [0, 2)\ \texttt{and}\ \textit{invasive} \in \{\text{YES, UNKNOWN}\})\ \texttt{or} \\
(\textit{tumor-size} \in [0, 2)\ \texttt{and}\ \textit{lymph-nodes} \in [0, 2))\ \texttt{or} \\
\textit{hepatitis} \in \{\text{YES, UNKNOWN}\}\ \texttt{or} \\
\textit{liver-infections} \in \{\text{YES, UNKNOWN}\}
\end{array}$$

(c) Rejection expression.

Figure 4.19: Acceptance and rejection expressions for the eligibility criteria in Figure 4.4. We represent both expressions as disjunctive normal forms without negations.

4.5 Entry Time

To evaluate the interface, we have run experiments with seven novice users. All participants have been undergraduate students, who had no prior experience with the interface. First, every user has entered four sets of medical tests; each set has included three tests and ten questions. Then, each user has added eligibility expressions for ten breast-cancer trials used at the Moffitt Cancer Center; the number of questions in an eligibility expression has varied from ten to thirty-five.

We have measured the entry time for each test set and each clinical trial (Tables 4.1 and 4.2). We show the mean time for every test set in Figure 4.20(left), and the time per question for the same sets in Figure 4.20(right). All users have entered the test sets in the same order, from 1 to 4; since they had no prior experience, their performance has improved during the experiment. In Figure 4.21, we give similar graphs for the entry of clinical trials.

In Figure 4.22, we plot the dependency of the entry time on the size of an eligibility expression, for the eight trials entered after the initial learning period. The results suggest that the time linearly depends on the number of questions, which means that the time per question does not depend on the complexity of an expression.

The experiments have shown that novices can efficiently use the interface; they quickly learn its full functionality, and their learning curve flattens after about an hour. The average time per question is 31 seconds for the entry of medical tests and 37 seconds for eligibility criteria, which means that a user can enter all breast-cancer trials used at Moffitt in about nine hours. On the other hand, coding the same trials without the interface is projected to take seven weeks of full-time work.

Table 4.1: Time to add medical tests and related questions. We give the times for seven users, who have entered four sets of tests. Every set includes three tests and ten questions.

Num. of a test set	Entry time (seconds)							
	User A	User B	User C	User D	User E	User F	User G	*Mean*
1	575	726	575	874	412	420	468	*579*
2	348	505	375	430	383	300	345	*390*
3	339	430	345	338	323	275	321	*339*
4	303	355	382	302	336	205	314	*316*

Table 4.2: Time to add eligibility criteria. We show the results for seven users; each user has constructed eligibility expressions for ten clinical trials. The number of questions in an expression varies from ten to thirty-five.

Num. of a trial	Num. of questions	Entry time (seconds)							
		User A	User B	User C	User D	User E	User F	User G	*Mean*
1	10	1380	590	406	566	970	420	563	*586*
2	12	225	322	580	700	640	437	475	*526*
3	15	443	466	570	340	775	300	507	*493*
4	18	622	443	812	712	1080	497	570	*686*
5	21	630	602	746	722	1230	828	760	*815*
6	27	683	597	700	612	972	882	579	*724*
7	28	753	742	1032	880	995	950	889	*915*
8	29	763	634	860	722	1020	763	865	*811*
9	30	431	561	623	460	765	443	605	*576*
10	35	1168	900	1265	1085	1555	1007	1160	*1162*

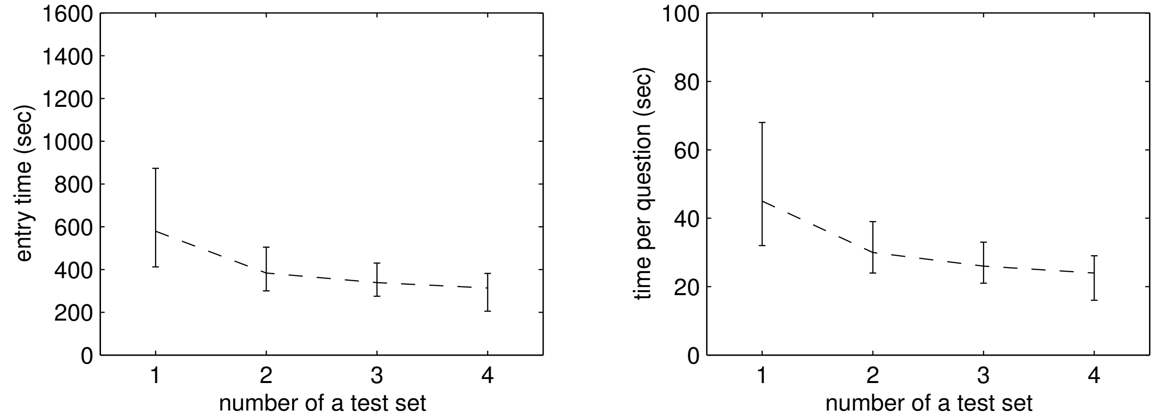


Figure 4.20: Entry time for test sets (left) and the mean time per question for each set (right). We plot the average performance (dashed lines) and the time of the fastest and slowest users (vertical bars).

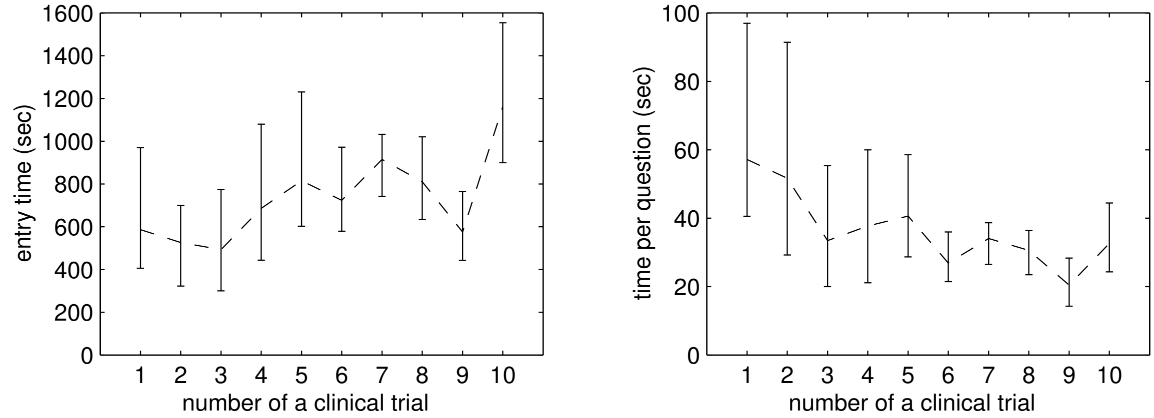


Figure 4.21: Entry time for eligibility criteria. We show the average time for each clinical trial and the corresponding time per question (dashed lines), along with the performance of the fastest and slowest users (vertical bars).

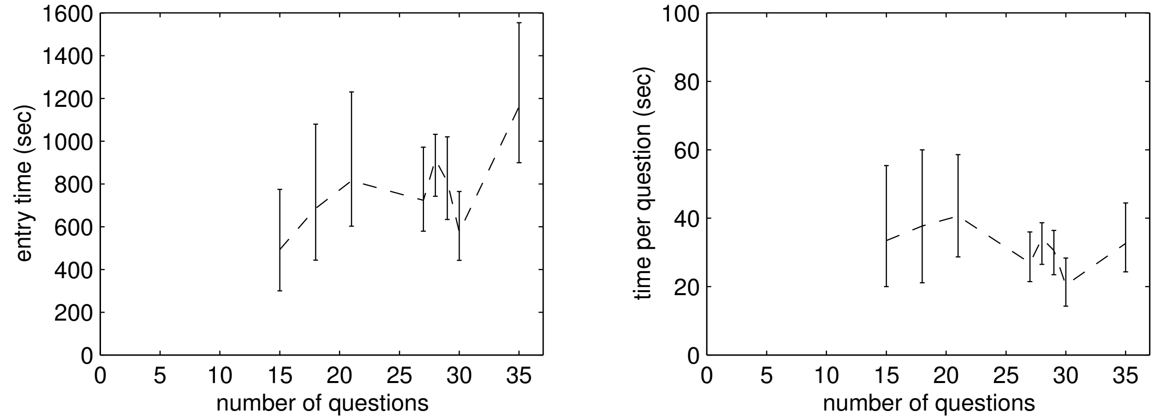


Figure 4.22: Dependency of the time on the number of questions in an eligibility expression. We plot the time of entering eligibility expressions (left) and the corresponding time per question (right). The results show that the time per question does not depend on the complexity of an expression.

Chapter 5

Concluding Remarks

We have developed knowledge-acquisition tools for an expert system that selects clinical trials for cancer patients. We have described the representation of selection criteria and a web-based interface for adding new trials. Although cancer research at Moffitt has provided the motivation for this work, the developed tools are not limited to cancer, and we can use them to enter selection criteria for clinical trials related to other diseases.

The experiments have shown that a user can enter a clinical trial in ten to twenty minutes, whereas programming the same knowledge without the interface takes about twenty hours (personal communication with Kokku). Novices can readily use the interface without prior instructions, and they reach their full speed after about an hour.

The experiments have also revealed several limitations of the developed tools, and addressing them may be a subject of future work. First, the expert system does not estimate probabilities of medical-test results. We conjecture that integration of probabilistic methods with the current heuristics may further reduce the cost of selected tests. Second, the system does not parse the text of questions, and it cannot recognize identical or related questions. If a user accidently enters the same question twice, the system will treat it as two different questions. Third, the interface does not allow a user to encode logical relationships among questions. For example, we cannot specify that a post-menopausal woman is never pregnant. If the knowledge base includes menopause and pregnancy questions, the system may ask about pregnancy even after learning that a patient is post-menopausal.

References

[Bhanja *et al.*, 1998] Sanjucta Bhanja, Lynn M. Fletcher, Lawrence O. Hall, Dimtry B. Goldgof, and Jeffrey P. Krischer. A qualitative expert system for clinical trial assignment. In *Proceedings of the Eleventh International Florida Artificial Intelligence Research Society Conference*, pages 84–88, 1998.

[Blythe *et al.*, 2001] Jim Blythe, Jihie Kim, Surya Ramachandran, and Yolanda Gil. An integrated environment for knowledge acquisition. In *Proceedings of the International Conference on Intelligent User Interfaces*, pages 13–20, 2001.

[Boose and Gains, 1990] John H. Boose and Brian R Gains. *The Foundations of Knowledge Acquisition.* Academic Press, San Diego, CA, 1990.

[Bouaud *et al.*, 1998] Jacques Bouaud, Briggite Séroussi, Éric-Charles Antoine, Mary Gozy, David Khayat, and Jean-Franşois Boisvieux. Hypertextual navigation operationalizing generic clinical practice guidelines for patient-specific therapeutic decisions. *Journal of the American Medical Informatics Association*, 5(suppl.):488–492, 1998.

[Bouaud *et al.*, 2000] Jacques Bouaud, Briggite Séroussi, Éric-Charles Antoine, Laurent Zelek, and Marc Spielmann. Reusing ONCODOC, a guideline-based decision support system, across institutions: A successful experiment in sharing medical knowledge. In *Proceedings of the American Medical Informatics Association Annual Symposium*, volume 7, 2000.

[Buchanan and Shortliffe, 1984] Bruce G. Buchanan and Edward H. Shortliffe. *Rule-Based Expert Systems: The* MYCIN *Experiments of the Stanford Heuristic Programming Project.* Addison-Wesley, Reading, MS, 1984.

[Buchanan *et al.*, 1969] Bruce G. Buchanan, Georgia L. Sutherland, and Edward A. Feigenbaum. Heuristic DENDRAL: A program for generating explanatory hypotheses in organic chemistry. In Bernard Meltzer and Donald Michie, editors, *Machine Intelligence*, volume 4, pages 209–254. Edinburgh University Press, Edinburgh, United Knigdom, 1969.

[Carlson *et al.*, 1995] Robert W. Carlson, Samson W. Tu, Nancy M. Lane, Tze L. Lai, Carol A. Kemper, Mark A. Musen, and Edward H. Shortliffe. Computer-based screening of patients with HIV/AIDS for clinical trial eligibility. *Online Journal of Current Clinical Trials*, 4(179), 1995.

[Cimino *et al.*, 1988] James Cimino, L. J. Mallon, and Octo G. Barnett. Automated extraction of medical knowledge from MEDLINE. In *Proceedings of the Twelfth*

Annual Symposium on Computer Applications in Medical Care, pages 180–184, 1988.

[Clancey and Soloway, 1990] William J. Clancey and Elliot Soloway. *Artificial intelligence and learning environments*. Elsevier, Amsterdam, Netherlands, 1990.

[Clancey, 1993] William J. Clancey. Notes on epistemology of a rule-based expert system. *Artificial Intelligence*, 59(1/2):191–204, 1993.

[Clancey, 1995] William J. Clancey. The learning process in the epistemology of medical information. *Methods of Information in Medicine*, 34(1/2):122–130, 1995.

[Crama and Hammer, 2001] Yves Crama and Peter L. Hammer. *Boolean Functions*. Sringer-Verlag, Berlin, Germany, 2001.

[Duff *et al.*, 2000] Franck Le Duff, Anita Burgun, Mireille Cleret, Bruno Pouliquen, Victoire Barac'h, and Pierre Le Beux. Knowledge acquisition to qualify unified medical language system interconceptual relationships. In *Proceedings of the American Medical Informatics Association Annual Fall Symposium*, 2000.

[Eriksson, 1993] Henrik Eriksson. Specification and generation of custom-tailored knowledge-acquisition tools. In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, volume 1, pages 510–518, 1993.

[Fallowfield *et al.*, 1997] Lesley Fallowfield, D. Ratcliffe, and Robert Souhami. Clinicians' attitudes to clinical trials of cancer therapy. *European Journal of Cancer*, 33(13):2221–2229, 1997.

[Gennari and Reddy, 2000] John H. Gennari and Madhu Reddy. Participatory design and an eligibility screening tool. In *Proceedings of the American Medical Informatics Association Annual Fall Symposium*, pages 290–294, 2000.

[Gil and Tallis, 1997] Yolanda Gil and Marcelo Tallis. A script-based approach to modifying knowledge bases. In *Proceedings of the Fourteenth National Conference on Artificial Intelligence*, pages 377–383, 1997.

[Hahn and Schnattinger, 1997a] Udo Hahn and Klemens Schnattinger. An empirical evaluation of a system for text knowledge acquisition. In *Proceedings of the Tenth European Knowledge Acquisition Workshop*, pages 129–144, 1997.

[Hahn and Schnattinger, 1997b] Udo Hahn and Klemens Schnattinger. Knowledge mining from textual sources. In *Proceedings of the Sixth International Conference on Information and Knowledge Management*, pages 83–90, 1997.

[Hahn and Schnattinger, 1998] Udo Hahn and Klemens Schnattinger. A text understander that learns. In *Proceedings of the Thirty-Sixth Annual Meeting of the Association for Computational Linguistics and Seventeenth International Conference on Computational Linguistics*, volume 1, pages 476–482, 1998.

[Hammond and Sergot, 1996] Peter Hammond and Marek J. Sergot. Computer support for protocol-based treatment of cancer. *Journal of Logic Programming*, 26(2):93–111, 1996.

[Johnson and Musen, 1996] Peter D. Johnson and Mark A. Musen. Development of a guideline authoring tool with PROTÉGÉ-II based on the DILEMMA generic protocol and guideline model. Technical Report SMI-96-0620, Stanford Medical Informatics Group, Stanford School of Medicine, 1996.

[Kenneth, 1988] Rosen H. Kenneth. *Discrete Mathematics and Its Applications*. McGraw-Hill, New York, NY, third edition, 1988.

[Keppel *et al.*, 2002] Kenneth G. Keppel, Jeffrey N. Pearcy, and Diane K. Wagener. Trends in racial and ethnic-specific rates for the health status indicators: United states, 1990–98. *Healthy People 2000, Statistical Notes*, 23, 2002.

[Kim and Gil, 2000a] Jihie Kim and Yolanda Gil. Acquiring problem-solving knowledge from end users: Putting interdependency models to the test. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence*, pages 223–229, 2000.

[Kim and Gil, 2000b] Jihie Kim and Yolanda Gil. User studies of an interdependency-based interface for acquiring problem-solving knowledge. In *Proceedings of the International Conference on Intelligent User Interfaces*, pages 165–168, 2000.

[Kokku *et al.*, 2002a] Princeton K. Kokku, Lawrence O. Hall, Dimitry B. Goldgof, Eugene Fink, and Jeffrey P. Krischer. A cost-effective agent for clinical trial assignment. Unpublished Manuscript, 2002.

[Kokku *et al.*, 2002b] Princeton K. Kokku, Savvas Nikiforou, Lawrence O. Hall, Dimitry B. Goldgof, Eugene Fink, Jeffrey P. Krischer, and Tim Ivanovkiy. A cost-effective agent for clinical trial assignment. Unpublished Manuscript, 2002.

[Korver and Janssens, 1993] M. Korver and A. R. Janssens. Development and validation of HEPAR, an expert system for the diagnosis of disorders of the liver and biliary tract. *Medical Informatics*, 16(3):259–270, 1993.

[Korver and Lucas, 1993] M. Korver and Peter J. F. Lucas. Converting a rule-based expert system into a belief network. *Medical Informatics*, 18(3):219–241, 1993.

[Kotwall *et al.*, 1992] Cyrus Kotwall, Leo J. Mahoney, Robert E. Myers, and Linda Decoste. Reasons for non-entry in randomized clinical trials for breast cancer: A single institutional study. *Journal of Surgical Oncology*, 50:125–129, 1992.

[Lederberg, 1965] Joshua Lederberg. DENDRAL-64: A system for computer construction, enumeration and notation of organic molecules as tree structures and cyclic graphs. Part II. Technical Report N66-14074, NASA Scientific and Technical Aerospace Reports, 1965.

[Lederberg, 1987] Joshua Lederberg. How DENDRAL was conceived and born. In *Proceeding of the ACM Symposium on the History of Medical Informatics*. National Library of Medicine, 1987.

[Lindberg *et al.*, 1993] Donald A. B. Lindberg, Betsy L. Humphreys, and Alexa T. McCray. The unified medical language system. *Journal of Methods of Information in Medicine*, 4(32):281–291, 1993.

[Lucas *et al.*, 1989] Peter J. F. Lucas, R. W. Segaar, and A. R. Janssens. HEPAR: An expert system for the diagnosis of disorders of the liver and the biliary tract. *Liver*, 9:266–275, 1989.

[Lucas, 1994] Peter J. F. Lucas. Refinement of the HEPAR expert system: Tools and techniques. *Journal of Artificial Intelligence in Medicine*, 6(2):175–188, 1994.

[Marcus and McDermott, 1989] Sandra Marcus and John P. McDermott. SALT: A knowledge acquisition language for propose and revise systems. *Artificial Intelligence*, 39(1):1–37, 1989.

[Mendonça and Cimino, 2000] Eneida A. Mendonça and James J. Cimino. Automatic knowledge extraction from MEDLINE citations. In *Proceedings of the American Medical Informatics Assocation Annual Fall Symposium*, pages 575–579, 2000.

[Musen *et al.*, 1988] Mark A. Musen, David M. Combs, Edward H. Shortliffe, and Lawrence M. Fagan. OPAL: Toward the computer-aided design of oncology advice systems. In Perry L. Miller, editor, *Selected Topics in Medical Artificial Intelligence*, pages 166–180. Springer-Verlag, New York, NY, 1988.

[Musen *et al.*, 1996] Mark A. Musen, Samson W.Tu, Amar K. Das, and Yuval Shahar. EON: A component-based approach to automation of protocol-Directed therapy. *Journal of the American Medical Informatics Association*, 3(6):367–388, 1996.

[Musen *et al.*, 2000] Mark A. Musen, Ray W. Fergerson, William E. Grosso, Natalya F. Noy, Monica Crubzy, and John H. Gennari. Component-based support for building knowledge-acquisition systems. In *Proceedings of the Conference on Intelligent Information Processing of the International Federation for Information Processing World Computer Congress*, 2000.

[Musen, 1989] Mark A. Musen. *Automated generation of model-based knowledge acquisition tools*. Morgan Kaufmann, San Mateo, CA, 1989.

[Ohno-Machado *et al.*, 1993] Lucila Ohno-Machado, Eduardo Parra, Suzanne B. Henry, Samson W. Tu, and Mark A. Musen. AIDS^2: A decision-support tool for decreasing physicians' uncertainty regarding patient eligibility for HIV treatment protocols. In *Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care*, pages 429–433, 1993.

[Ohno-Machado *et al.*, 1998] Lucila Ohno-Machado, John H. Gennari, Shawn Murphy, Nilesh L. Jain, Samson W. Tu, Diane E. Oliver, and Edward Pattison-Gordon. The guideline interchange format: A model for representing guidelines. *Journal of American Medical Informatics Association*, 5(4):357–372, 1998.

[Onisko *et al.*, 1997] Agnieszka Onisko, Mark J. Druzdzel, and Hanna Wasyluk. Application of Bayesian belief networks to diagnosis of liver disorders. In *Proceedings of the Third Conference on Neural Networks and Their Applications*, pages 730–736, 1997.

[Papaconstantinou *et al.*, 1998] Constantinos Papaconstantinou, Georgios Theocharous, and Sridhar Mahadevan. An expert system for assigning patients into clinical trials based on Bayesian networks. *Journal of Medical Systems*, 22(3):189–202, 1998.

[Price, 1990] Chris J. Price. *Knowledge Engineering Toolkits*. Ellis Horwood, Chichester, West Sussex, United Knigdom, 1990.

[Puerta *et al.*, 1992a] Angel R. Puerta, John W. Egar, Samson W. Tu, and Mark A. Musen. A multiple-method knowledge-acquisition shell for the automatic generation of knowledge-acquisition tools. *Knowledge Acquisition*, 4(2):171–196, 1992.

[Puerta *et al.*, 1992b] Angel R. Puerta, Henrik Eriksson, John W. Egar, and Mark A. Musen. Generation of knowledge-acquisition tools from reusable domain ontologies. Technical Report KSL-92-81, Computer Science Department, Stanford University, 1992.

[Ringland and Duce, 1988] Gordon A. Ringland and David A. Duce. *Approaches to Knowledge Representation. An Introduction.* Research Studies Press, Letchworth, United Knigdom, 1988.

[Romacker and Hahn, 2001] Martin Romacker and Udo Hahn. Semantic interpretation of medical language—quantitative analysis and qualitative yield. In Silvana Quaglini, Pedro Barahona, and Steen Andreassen, editors, *Artificial Intelligence in Medicine*, pages 258–267. Springer-Verlag, Berlin, Germany, 2001.

[Rothenfluh *et al.*, 1996] Thomas E. Rothenfluh, John H. Gennari, Henrik Eriksson, Angel R. Puerta, Samson W. Tu, and Mark A. Musen. Reusable ontologies, knowledge-acquisition tools, and performance systems: PROTÉGÉ-II solutions to Sisyphus-2. *International Journal of Human-Computer Studies*, 44(3–4):303–332, 1996.

[Rubin *et al.*, 1999] Daniel L. Rubin, John H. Gennari, Sandra Srinivas, Allen Yuen, Herbert Kaizer, Mark A. Musen, and John S. Silva. Tool support for authoring eligibility criteria for cancer trials. In *Proceedings of the American Medical Informatics Association Annual Symposium*, pages 369–373, 1999.

[Rubin *et al.*, 2000] Daniel L. Rubin, John H. Gennari, and Mark A. Musen. Knowledge representation and tool support for critiquing clinical trial protocols. In *Proceedings of the American Medical Informatics Association Fall Symposium*, pages 724–728, 2000.

[Sager *et al.*, 1994] Naomi Sager, Margaret Lymanand Christine Bucknall, Ngo Nha, and Leo J. Tick. Natural language processing and the represantation of clinical data. *American Medical Informatics Association*, 1(2):142–160, 1994.

[Séroussi *et al.*, 1999a] Briggite Séroussi, Jacques Bouaud, and Éric-Charles Antoine. Enhancing clinical practice guideline compliance by involving physicians in the decision process. In Werner Horn, Yuval Shahar, Greger Lindberg, Steen Andreassen, and Jeremy C. Wyatt, editors, *Artificial Intelligence in Medicine*, pages 76–85. Springer-Verlag, Berlin, Germany, 1999.

[Séroussi *et al.*, 1999b] Briggite Séroussi, Jacques Bouaud, and Éric-Charles Antoine. Users' evaluation of ONCODOC, a breast cancer therapeutic guideline delivered at the point of care. *Journal of the American Medical Informatics Association*, 6(5):384–389, 1999.

[Séroussi *et al.*, 2001a] Briggite Séroussi, Jacques Bouaud, and Éric-Charles Antoine. ONCODOC: A successful experiment of computer-supported guideline development and implementation in the treatment of breast cancer. *Artificial Intelligence in Medicine*, 22(1):43–64, 2001.

[Séroussi *et al.*, 2001b] Briggite Séroussi, Jacques Bouaud, Éric-Charles Antoine, Laurent Zelek, and Marc Spielmann. Using ONCODOC as a computer-based eligibility screening system to improve accrual onto breast cancer clinical trials. In Werner Horn, Yuval Shahar, G. Lindberg, Steen Andreassen, and J. Wyatt, editors, *Artificial Intelligence in Medicine*, pages 413–430. Springer-Verlag, Berlin, Germany, 2001.

[Shortliffe *et al.*, 1975] Edward H. Shortliffe, Randall Davis, Stanton G. Axline, Bruce G. Buchanan, Cordell C. Green, and Stanley Cohen. Computer-based consultations in clinical therapeutics: Explanation and rule acquisition capabilities of the MYCIN system. *Computers and Biomedical Research*, 8:303–320, 1975.

[Shortliffe *et al.*, 1981] Edward H. Shortliffe, A. Carlisle Scott, Miriam B. Bischoff, William van Melle, and Charlotte D. Jacobs. ONCOCIN: An expert system for oncology protocol management. In *Proceedings of the Seventh International Joint Conference in Artificial Intelligence*, pages 876–881, 1981.

[Shortliffe, 1974] Edward H. Shortliffe. MYCIN*: A Rule-Based Computer Program for Advising Physicians Regarding Antimicrobial Therapy Selection.* PhD thesis, Computer Science Department, Stanford University, 1974.

[Shusaku, 1998] Tsumoto Shusaku. Automated extraction of medical expert system rules from clinical databases based on rough set theory. *International Journal in Information Sciences*, 112:67–84, 1998.

[Tallis and Gil, 1999] Marcelo Tallis and Yolanda Gil. Designing scripts to guide users in modifying knowledge-based systems. In *Proceedings of the Sixteenth National Conference on Artificial Intelligence*, pages 242–249, 1999.

[Tallis *et al.*, 1999] Marcelo Tallis, Jihie Kim, and Yolanda Gil. User studies of knowledge acquisition tools: Methodology and lessons learned. In *Proceedings of the Twelfth Workshop on Knowledge Acquisition, Modeling and Management*, 1999.

[Tallis, 1998] Marcelo Tallis. A script-based approach to modifying knowledge-based systems. In *Proceedings of the Tenth Innovative Applications of Artificial Intelligence Conference*, pages 1183–1195, 1998.

[Theocharous, 1996] Georgios Theocharous. An expert system for assigning patients into clinical trials based on Bayesian networks. Master's thesis, Computer Science and Engineering Department, University of South Florida, 1996.

[Tu *et al.*, 1993] Samson W. Tu, Carol A. Kemper, Nancy M. Lane, Robert W. Carlson, and Mark A. Musen. A methodology for determining patients' eligibility for clinical trials. *Journal of Methods of Information in Medicine*, 32(4):317–325, 1993.

[Wang *et al.*, 2001] Dongwen Wang, Mor Peleg, Samson W. Tu, Edward H. Shortliffe, and Robert A. Greenes. Representation of clinical practice guidelines for computer-based implementations. *Medical Informatics*, 10:285–289, 2001.

[Yusuf *et al.*, 1990] Salim Yusuf, Peter Held, K. K. Teo, and Elizabeth R. Toretsky. Selection of patients for randomized controlled trials: Implications of wide or narrow eligibility criteria. *Statistics in Medicine*, 9:73–86, 1990.