Visualization for Seeking and Comparing Clinical Trials

by

Maria-Elena Hernandez
B.Computer Science, Autonomous University of Puebla, 1994
M. Computer Science and Engineering, University of The Americas, 1996

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Computer Science

Visualization for Seeking and Comparing Clinical Trials

by

Maria-Elena Hernandez
B.Computer Science, Autonomous University of Puebla, 1994
M. Computer Science and Engineering, University of The Americas, 1996

Supervisory Committee

Dr. Margaret-Anne Storey, (Department of Computer Science)
Supervisor

Dr. Melanie Tory, (Department of Computer Science)
Departmental Member

Dr. Hausi A. Müller, (Department of Computer Science)
Departmental Member

Dr. Andre Kushniruk, (Health Information Science)
Outside Member

Supervisory Committee

Dr. Margaret-Anne Storey, (Department of Computer Science)
Supervisor

Dr. Melanie Tory, (Department of Computer Science)
Departmental Member

Dr. Hausi A. Müller, (Department of Computer Science)
Departmental Member

Dr. Andre Kushniruk, (Health Information Science)
Outside Member

ABSTRACT

The sheer quantity of information available on the Internet poses a challenge to users who need an efficient way of finding the information most relevant to their needs. One of the most frequent information-seeking activities of Internet users is the search for health and medical information. In this research, I focus on the user process of seeking information on clinical trials which are the only evidence-based source of information in the medical domain. Through my work, I show that improvements for seeking clinical-trial information could be made to enhance the effectiveness of seeking and gathering results from clinical trials. Unfortunately, little work has been reported on alternative methods and on visualization systems in particular for these enhancements. I suggest that this omission may be due to a lack of understanding of the particular information needs of users of clinical-trial data. Understanding the users' needs is the first step towards providing more effective interfaces to support them in the information seeking process.

In this dissertation, I investigate 1) how information is accessed in the medical domain; more specifically, how users seek clinical-trial information on the Internet and 2) how to improve current Web-seeking interfaces for clinical-trial users. I discuss my findings from three exploratory studies: moderated discussion with professional researchers of clinical trials, an online questionnaire with health professionals and patients who search the clinical trial domain, and a qualitative query log analysis of a popular medical search engine.

The results of this research indicate that most of the time users are successful in finding the information they require. However, the process of seeking clinical-trial information is tedious, frustrating and time consuming, because current search interfaces do not sufficiently support users seeking this kind of information. Based on the findings from my studies, I propose a set of design principles to design better seeking interfaces. I validate my findings and the set of design principles with two visualization tools that support users in performing information-seeking tasks in the clinical-trial domain. Finally, I provide initial evidence that my proposed designs are indeed helpful with finding, summarizing and comparing information in this complex medical domain.

Table of Contents

List of Tables

List of Figures

Acknowledgement

The work reported in this dissertation was possible thanks to the guidance I received from my supervisor Dr. Margaret-Anne Storey. Dr. Storey provided me with innumerable suggestions for improvement of my research. Her continuous support and feedback helped me to stay focused and excited about my research work. Her friendship and continuous interest in my well-being during these years will not be forgotten. I hope I can transmit well to my future students what I learnt from Dr. Storey about student motivation and encouragement. It has been an honor to work with her, and I feel a deep sense of satisfaction in having met her high standards.

Similarly, I want to acknowledge the valuable assistance I received from my committee members. Many thanks to Drs. Hausi Muller and Andre Kushniruk for their patience, feedback, and encouragement; to Dr. Melanie Tory for walking the extra mile with me and providing so many thoughtful comments and suggestions; and to Mark D. Wilkinson for acting as my external examiner.

I would like to acknowledge the remarkable team spirit in the Computer Human Interaction and Software Engineering Laboratory (CHISEL), where each one of my colleagues have played an important role in this work. I have rarely had the opportunity to collaborate with colleagues as supportive and talented as they have been. I am thankful to Dr. Sean Falconer not only for his collaboration on this research and for proofreading the final draft of my dissertation, but also for his exemplary persistence and discipline; to Jody Ryall for the countless hours he spent debating with me about my research; to Tricia Pelz for constantly looking over my shoulder and keeping me focused on my goals; to Chris Callendar and Chris Bennett for contributing their programming skills to my research; and to Christoph Treude for his feedback on my document.

I want to acknowledge Dr. Ida Sim and Simona Carini for introducing me to a new world of research opportunities and challenges in the areas of clinical-trial registration and knowledge management and for their patience in educating me in these intriguing domains.

To my colleagues and bosses from the Autonomous University of Tlaxcala, Mexico, who believed in me and were always confident that I would succeed in my doctoral program and return to my homeland to contribute to improving public education there. Antonio Durante Murillo has been an inspiring example of making things happen in an educational environment that is somewhat lacking in financial resources but not in human quality and enterprise. To Dr. Luciano Garcia-Banuelos and Marva Mora-Lumbreras I am greatly in debt for looking after my interests as if they were their own.

I want also to acknowledge the wonderful Canadian friends I have made over the years. Mary van Kooten and Vivian Roberts became second mothers to my daughter. Penelope Fenske and Michelle M. Irwin have taught me how to be overcome the challenges of single parenting as a graduate student. Dean Froese for the joy brought to the latest iterations of this document.

Finally, I extend my thanks to the Hernandez Clan–Mom, Celina, Aida, Alejandro, Angelica, and Adan, who have cherished my successes and cried over my losses; to my favorite person, my daughter Mariel, who has taught me to enjoy life; and to you, oh Lord, for granting me the desires of my heart!

Dedication

to my beloved daughter...

Chapter 1

Towards More Effective Support for Users of Clinical Trial Data

The Internet has become an important source of information for researchers and the general public alike. In the medical field, although users often find the information they need, the process of seeking the desired information is frequently slow and tedious. Human Computer Interaction (HCI) is a discipline that studies the design, implementation, and evaluation of interactive computer programs [DFA+98, ACM]. From an HCI perspective, when users are frequently frustrated and need to make a great cognitive effort to accomplish their tasks, this indicates a design problem in the user interface. User interfaces are more effective when they are developed based on the requirements of the targeted users and when they are tested with those users. Unfortunately, documented studies on the development of user interfaces frequently indicate that they are still developed according to common sense and the good intentions of software engineers and computer scientists, and are tested on people who are different from the target users, such as students of engineering or computer science.

No one doubts that the Internet is a valuable resource which permits access to a vast amount of medical information; however, given the current search interfaces, finding the information one desires and making sense of it is still a challenge. In this dissertation, I approach this topic by focusing on how information is accessed in the medical domain; more specifically, on how users search for information on clinical trials.

Clinical trials are biomedical experiments designed to investigate the efficacy of a medical treatment with a sample of patients who present specific medical conditions. Most of the time, the treatment to be evaluated is based on administering drugs, but in some cases may consist of other therapies, such as surgery, radiotherapy, hospitalization, etc. A researcher performs experiments to find the best treatment for an illness or disease [EH05, Pia05, Poc84, WB06]. The users of clinical-trial information that I consider in this dissertation are a quite diverse population: medical researchers, clinicians, nurses, patients, and family of patients. Medical researchers and practitioners seek and compare clinical trials in order to find research opportunities, treatment strategies and similarities. Patients and family of patients are interested in finding what treatments have been tested for a given disease, what are the secondary effects of a treatment, and which trials are open for enrollment.

1.1 Motivation

Searching the Internet for clinical trials and understanding the relationship between the retrieved documents is an important, but cognitively complex process. Users of clinical trials search the Internet using generic search engines such as Google, AOL, and Yahoo! or specialized search engines corresponding to *The Journal of the American Medical Association* (JAMA), *The Lancet*, *Annals of Internal Medicine*, *ClinicalTrials.gov*, *MedLine*, and *The Cochrane Collaboration*. Typically in these search engines, users are provided with a search box to enter a query, and, having done so, they receive a large list of documents to explore, compare, and contrast. The number of retrieved documents may be overwhelming for users who still have to invest many hours in finding answers to their questions.

Often, the answers to the users' questions are contained in more than one document, or, more precisely, in the aggregation of segments of a particular collection of related documents. For example, a user who wants to know the most common treatment for cancer of the liver enters a search for "liver cancer treatment" into PubMed, a search engine for biomedical information. The answer this user is looking for may not be in a single clin-

ical trial, but in a tabulation of all the studies on treatments for this type of cancer that are listed in the query result. From Figure 1.1, one can see that the typical list of search results provides no clear indication of which documents are reporting on studies regarding the treatment of liver cancer, and thus being relevant to the posed question.

Figure 1.1. *PubMed Retrieved Documents*

Summarizing and comparing selected results is challenging due to the different contexts in which those trials occurred, and the diversity of variables taken into account in each particular study. For example, a study testing a specific drug on HIV conducted in Africa may have quite different results from a study conducted in North America testing the same drug. The ethnicities, genetics, dietary habits, resources, beliefs of these two populations are so diverse that the studies may seem hard to compare. Trials can differ in the criteria used to select participants, the drugs used, and the outcomes measured. To complicate matters further, because of a lack of regulation, published reports on trials have frequently focused on positive results, hiding or minimizing negative results [ZIT+07].

Moreover, many lay users do not know the technical terms used in the medical literature and may follow a tedious trial-and-error process, gradually learning and submitting more appropriate terms, until they acquire the desired results [BYRN99, HK97, Mar06, SBC98]. For example, users searching trials for cancer of the liver might not know that medical terms such as "liver carcinoma" or "hepatocellular carcinoma" refer to that illness.

1.2 Research Problem

This research was undertaken in collaboration with the National Center for Biomedical Ontology (NCBO), a research consortium that promotes the electronic dissemination of biomedical knowledge. It was motivated by discussions with two NCBO clinical-trial researchers based at the University of California at San Francisco (UCSF). They expressed a need for an efficient, intuitive, computer-supported tool to help them search for and analyze previously published clinical trials. They stated that searching for and analysing electronically-stored, related trials is a complicated and time-consuming process.

In this dissertation, I explore the problem of searching clinical trials on the Internet and comparing retrieved results with the purpose of proposing better user interfaces using Information Visualization (Infovis), a discipline that has emerged from HCI with the goal of supporting the cognitive process of understanding abstract data [CMS99, War00]. Visual representations may be used to support the user to search for relevant trials, to summarize large lists of results, and to support dynamic exploration of large amounts of clinical data.

Little work has been reported on alternative methods, and on visualization systems in particular, to enhance the effectiveness of seeking and gathering results from clinical trials. I believe that this omission may be due to a lack of understanding of the particular information needs of users of clinical-trial data. Consequently, I conducted a series of studies to investigate in depth the questions these users ask and the tasks they undertake, as well as how they believed current search interfaces could be improved to support such questions and tasks.

The ultimate goal of my research is to propose design principles to improve the users' experience of seeking and comparing collections of clinical trials. To this end, I investigate the behaviour of users searching for clinical trials on the Internet and suggest improvements to the user interface in the information-seeking process. Specifically, my research leads me to suggest the use of interactive visualization tools to support the user in the cognitively complex problem of finding, summarizing and comparing related clinical trials. This research goal raises the following research questions:

Q1: What are the information needs of users searching for clinical trial information on the Web?

Q2: How can we improve current Web-search interfaces for clinical trial information?

1.3 Research Design

The research questions approached in this dissertation are exploratory since little is known on the behaviour and needs of the clinical trial user. Thus, I follow a qualitative approach [Cre04, SSI07], which consists of the collection and analysis of qualitative data.

The research has three parts: literature review; exploratory studies of the user tasks and needs; and confirmatory studies:

Literature review: I perform a review of the scholarly literature which investigates users' behaviour while seeking health and medical information on the Internet, as well as that which explores the state of the art in visualization techniques used in searching electronic documents, insofar as these appeared pertinent to the seeking of specific information from clinical trials.

Exploratory studies of user tasks and needs: Since the existing literature is lacking in studies specifically devoted to the needs of Internet users seeking information on the experimental designs and results of clinical trials, I undertook some fresh research in this area. Specifically, I moderated discussions with two researchers from UCSF

who were experts on clinical trials, conducted an online questionnaire of both expert and lay users of clinical-trial data, and performed a query-log analysis of the leading electronic database of clinical-trial and other biomedical information. The findings of this research provide the answer to Question 1 above. Using the knowledge gained from this stage, I define a set of design principles that can be used to develop more effective user interfaces to support the information-seeking process in the domain on clinical-trials. These principles provide answer to Question 2 above.

Confirmatory studies: To confirm my findings and proposed design principles, I designed two visualization tools, namely, CTSearch and CTeXplorer. I conducted user studies to test the usefulness of the proposed visualizations. The user studies consisted on testing the visualizations with end-users engaged in tasks based in real settings. These studies were used to complement the answer for Question 2.

1.4 Scope

The focus of this dissertation is to define design principles to guide the design of effective user interfaces to explore clinical-trial data. To this end, I consider two groups of users: a) medical professionals (researchers, clinicians, etc.) and b) lay users (patients, family or friends of patients, etc.). Initially, we were focused on investigating the needs of medical professionals; however, lay users are a large and active group of users and as this research progressed, we realized the importance of including this population. The analysis of the differences between these two groups of users was not the focus of this dissertation. Nevertheless, some insight in this regard is discussed in Chapter 6.

The data considered in the dissertation consists of the experimental designs of clinical trials including information such as: medical condition under study, therapy to test, outcomes or effects measured, characteristics of the population, etc. The actual results of the clinical trials are not considered. This decision was made in virtue of the data that was available for this investigation.

1.5 Organization of Dissertation

The remaining chapters of this dissertation are organized as follows (cf. Figure 1.2):

- In Chapter 2, I provide a brief description of the history of clinical trials and introduce the definitions of important terms within the field which are used throughput this dissertation.
- In Chapter 3, I discuss theories and models that describe the process that users follow when seeking general information on the Internet.
- In Chapter 4, I describe results from current empirical studies that focus on the behavior of users searching for health and medical information on the Internet.
- In Chapter 5, I describe the design of my dissertation research, providing accounts of two stages of the research (exploratory and confirmatory) and of the experimental instruments used in both stages.
- In Chapter 6, I proceed to discuss in detail my exploratory studies and discuss their findings.
- In Chapter 7, I discuss the challenges that the information seeking process poses to users of clinical trials and propose a set of design principles and use cases that a visualization tool should incorporate in order to satisfy the information needs of these users.
- In Chapter 8, I discuss the design of a visualization tool, namely, CTSearch, to support information seeking in the clinical-trial domain. I also report on an observational laboratory study and a post-study questionnaire, that I used to test the usefulness of this application, and discuss the results.
- In Chapter 9, I discuss the design of a second visualization tool, namely, CTeXplorer, to support the comparison of data retrieved from a particular collection of related clinical trials. I also report on an observational laboratory study, a post-study

questionnaire, and a focus-group discussion, that I used to test the usefulness of this application, and discuss the results.

- In Chapter 10, I discuss the contributions of the research undertaken for this dissertation, its limitations, and make some suggestions for future research and development.

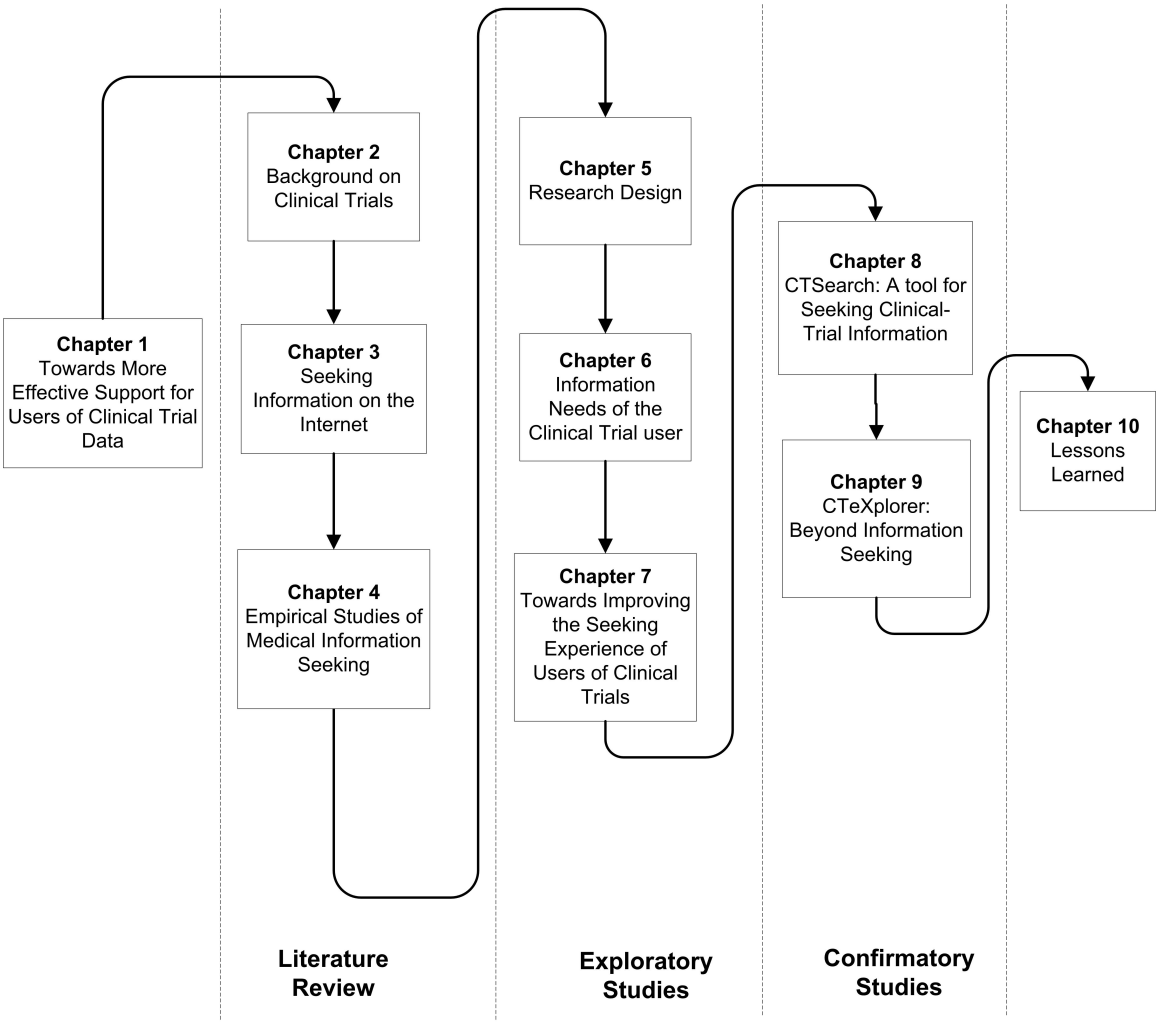

Figure 1.2. *Dissertation Organization*

Part I

Literature Review

Chapter 2

Background on Clinical Trials

This chapter provides an overview of basic methods and protocols of clinical-trial studies. I start with a brief discussion of the history of clinical trials, paying particular attention to methodological developments; then I describe the four phases of clinical-trial research; and, finally, I examine the data elements of which clinical trials are composed.

Clinical trials are essential in discovering and evaluating therapies, as they allow medical researchers to better understand the course of a disease and the effects of various therapies upon it. Advances in medicine often result from clinical trials, which are especially important in the case of diseases, such as AIDS or fibromyalgia, that cannot be cured using current therapies.

2.1 History of Clinical Trials

In 1834, Pierre Charles A. Louis, argued for the importance of using the scientific method to support medical research. However, it took more than 100 years for researchers to practice their experiments with the rigor and systematization that the scientific method demands and for clinical trials to be widely accepted as the most valuable method in medical research [Poc84, Goo03]. With a view to standardizing the way in which medical experiments were conducted, Austin B. Hill published several articles in the 1950's detailing how a clinical trial should be designed and executed. He discussed such matters as patient selection, description of the treatment, and follow up [Poc84].

In their modern conception, clinical trials are scientific investigations that are systematic, controlled, randomized, and double-blind. Their purpose is to evaluate the effectiveness of a targeted treatment, and their results are published in unbiased, peer-reviewed journals. Four of the key terms that define a clinical trial are *control group*, *placebo*, *randomization*, and *double-blindness*. These terms are defined below in the context of the historical development of clinical-trial methodology from the 18th to the 20th century.

The 18th century: *Control groups* were first used in clinical research in the mid-18th century. A control group is a group of patients who do not receive the treatment under study. It is used as a baseline and allows the comparison and evaluation of different interventions. In 1747, James Lind conducted a study to treat scurvy, in which he introduced, for the first time, a control group to test the traditional therapy for this medical condition. Lind's study demonstrated the effectiveness of citrus fruits as a treatment for this disease. For this contribution he is considered the father of clinical trials [Dun97].

The 19th century: The 19th century saw the first use of *placebos* to compare the progression of an illness in the control group and the group receiving an intervention. A placebo is an inactive substance or treatment administered in place of the one whose effects are being tested. Pierre Charles A. Louis is remembered for conducting the first evaluation of the efficacy of *bloodletting*, a practice which had been taught for many centuries in the best medical schools and which was still used in Louis's day to treat a vast number of illnesses. He found no scientific evidence of the benefit of this therapy for most of the cases and demonstrated that it had merely a psychological effect [Mor06].

The 20th century: The remaining two aspects of modern clinical-trial research mentioned above, *randomization* and *double-blind* trials, were introduced during the 20th century. Randomization is the process of randomly assigning patients to one or another of the groups under study. This practice guarantees that no bias affects the selection.

Randomization was used for the first time in 1948 in a study on the effectiveness of streptomycin in pulmonary tuberculosis. Randomization can be complemented with a double-blind strategy. Double-blind research utilizes a design in which neither the patient nor the clinician knows whether the patient is receiving a treatment or a placebo. (A single-blind study utilizes a design in which patients do not know whether they are receiving the treatment or placebo). The first double-blind clinical trial was conducted in 1950 to study the effects of antihistamines on the common cold. During the 20th century, multi-centred studies also became common practice; this type of study takes place in more than one medical centre and has the purpose of increasing the size of the trial sample and, thus, the reliability of the findings [Poc84].

Modern trials testing drugs are conducted at different stages. The following section introduces the four phases of a clinical trial.

2.2 The four phases of Clinical Trials

There are four successive phases of clinical trials through which a new drug therapy passes while being evaluated for efficacy and safety [WB06, Pia05, Poc84]:

Phase I: This phase deals with healthy individuals typically hired by pharmaceutical companies to participate in the experiment. The purpose of this phase is to understand how a drug affects the human body in the short term and to discover a range of dosage that is not harmful to the participant. Once a safe dosage has been identified, more testing is done with real patients, usually in groups of 20 to 80 participants. When dealing with a highly toxic treatment, such as chemotherapy for cancer, real patients, not hired participants, are used.

Phase II: In this phase, a wider sample is recruited, usually between 100 to 200 participants. The purpose of this phase is to determine the benefits, side effects, and optimal dosage of the drug.

Phase III: The purpose of this phase is to compare the efficacy and safety of a treatment against other drugs or against a placebo. When the trial is conducted by a pharmaceutical company, typically this phase tests a new drug in comparison to older and better known drugs. This phase involves from a few hundred to a few thousand participants. Phase III studies are Randomized Clinical Trials (RTCs). Successful RCTs are required to obtain approval from regulatory agencies for a drug to be released onto the market. Most of clinical trials in this stage are double-blind.

Phase IV: Phase IV trials, also called post-marketing trials, begin after a drug has been approved for marketing. Their intent is to investigate the drug's effects on a massive population over a longer period of time, its interaction with other drugs, and its effects on diseases other than the originally targeted ones.

To ensure that a new therapy is safe and effective, it is essential not only that trials be conducted during each of the four stages, but also that multiple trials be conducted on different populations and in different socio-cultural contexts during each stage. This is evident from news reports regarding the many drugs that have been found highly toxic. For example, *Lisinopril* tablets for hypertension, produced by the Spanish company Normon SA, were linked to the death of 20 consumers in Panama in late 2006 [RLM$^+$08]; and the drug TGN1412 (*CD28-SuperMAB*), developed by TeGenero Immuno Therapeutics for the treatment of rheumatoid arthritis and chronic lymphoid leukemia, was linked to failure of vital organs in a clinical trial with six human participants in early 2006 [DVE06].

2.3 Data and Datasets of Clinical Trials

From the standpoint of designing an information system, there are several objects of information (or data elements) that compose a clinical trial. Examples of such data components are: *eligibility criteria*, *interventions*, *outcomes*, *countries of recruitment*, *size of the sample*, *start date*, *end date*, *blinding method*, and *statistical methods*. These are defined below:

Eligibility Criteria: These criteria are the conditions or attributes that a patient must possess in order to be eligible to participate in a medical trial. Most of the eligibility criteria are defined in terms of medical qualifying conditions; for example, in a trial of a treatment for diabetes, these conditions might include being over 40 years of age and suffering from Type II diabetes. Along with these conditions, other medical factors, such as interventions received prior to the trial, or *co-morbidities* may influence the decision of the clinician to accept a participant (such as not having undergone surgery).

Interventions: A treatment, or intervention, consists of the administration of one or more drugs that are tested to compare their efficacy and safety. Each trial population is divided into one or more intervention groups and a control group, where the intervention group or groups typically receive the drug being tested and the control group receives a different drug or a placebo. Interventions are described in terms of the drugs, dosage, and frequency of administration.

Outcomes: Outcomes are recorded along a timeline and are classified as either efficacy outcomes or safety outcomes. Efficacy outcomes identify the quality of the effects of a given treatment, while safety outcomes consider the integrity of the patient.

Countries of recruitment: The countries and cities where the study was conducted and from which participants originated is also tracked. Additionally, information about ethnicity is sometimes collected.

Sample size: The sample size is the number of participants in the study.

Start date: The start date is the date when the trial begins.

End date: The end date is the date when the trial ends.

Blinding method: This refers to the blinding method used in the trial, that is, either blinding or double blinding.

Statistical methods: This indicates the statistical methods used in the trial.

The digital collections I am addressing are centralized repositories that store large collections of semi-structured text-based documents (articles, books, journal papers, and clinical-trial protocols). These documents are semi-structured because they are only partially described according to a defined model or schema. Most of these digital collections are annotated using controlled vocabularies to facilitate storage and retrieval of the documents. For example, ClinicalTrials.gov annotates its electronic documents using the Unified Medical Language System (UMLS), and MedLine indexes its database using the Medical Subject Headings (MeSH) [ZKA+02].

Such semi-structured information resources are not unique but are very prevalent on the Web. Non-medical examples include scientific publication archives such as CiteSeerX[1], SCIRUS[2], and DBLP[3]; online stores such as Amazon[4]; and online library systems.

2.4 Summary

Clinical trials are essential in discovering and evaluating new therapies to treat disease. The importance of using a systematic method to conduct trials is clear in the light of historical evidence: in particular, the fact that up until 1950 therapies tended to be arbitrarily developed and introduced, and their proponents always claimed outstanding results [Poc84]. To test a drug or therapy, it is necessary to conduct exhaustive research in each of the four essential phases. Replication of trials is necessary to test different populations and contexts. Clinical-trial data can be accessed from open databases on the Internet and is often stored in a semi-structured organization.

[1] http://citeseerx.ist.psu.edu/
[2] http://www.scirus.com/
[3] http://dblp.uni-trier.de/
[4] http://www.amazon.com/

Chapter 3

Seeking Information on the Internet

Before investigating the information-seeking process as applied specifically to the area of clinical-trial data, it is necessary to gain a theoretical understanding of that process more generally. In this chapter, I give an overview of theories and models of users seeking general information on the Internet. This discussion will help us develop an initial understanding of the information-seeking process and will be used in later chapters to identify potential challenges for users of clinical trials. I begin by introducing some basic concepts, and then move on to discuss specific theories and models: namely, the information-foraging theory, the traditional-information seeking models, and the berry-picking model.

Information seeking is the iterative process of searching and browsing for information, which is conducted by users who often change strategies during such a process [Mar95, BYRN99, Bat89]. Research on information seeking, as well as the theories and methods on which it is based, derives from the information sciences and views the searching problem as having both human and technological aspects. Information seeking includes evaluation criteria based on human needs such as learnability, efficiency, effectivity, accessibility, and feedback; these are concepts commonly used to evaluate systems from the human-computer interaction (HCI) point of view.

3.1 Basic definitions

To begin discussing the behaviour of users seeking information on the Internet, I start with basic definitions such as **searching** and **browsing**. In the literature, these terms are used with slightly different connotations, as we shall see in the following examples.

1. Baeza-Yates and Ribeiro-Neto [BYRN99] define *searching* as the process of retrieving information with clear goals by submitting a set of words that describe the information needed. *Browsing* is defined as a process of retrieving information with a goal that is not clearly specified at the beginning of the interaction with the system. The user may iteratively switch from searching to browsing to satisfy an information need (cf. Figure 3.1).

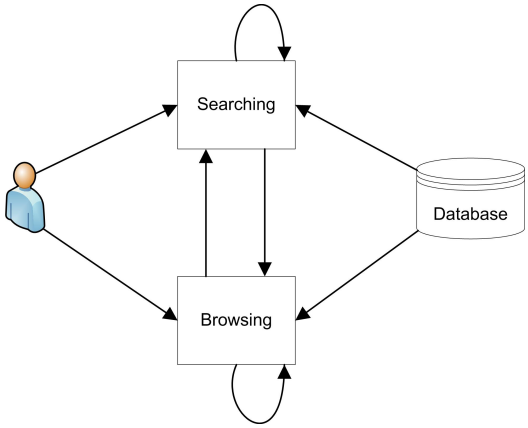

Figure 3.1. *Searching and Browsing Interaction by Baeza-Yates and Ribeiro-Neto [BYRN99]*

2. Furnas [JF97] defines searching and browsing as **tasks** that can be accomplished by two **tactics**: querying or navigating. *Searching* is the task of looking for known information, *browsing* is the task of looking to see what information is available, *querying* consists of submitting a set of keywords describing the desired information into a search engine, and *navigating* consists of moving sequentially to find specific information and deciding where to go next based on what has been seen so far (cf. Figure 3.2). Searching tasks can be accomplished by querying or navigating, while

browsing is usually done by navigating alone. However, browsing can involve querying when users pose broad queries to get an overview of what is contained in a digital collection.

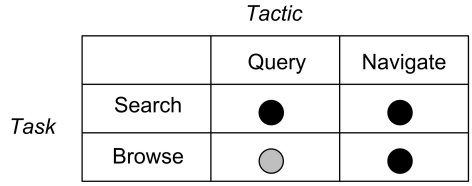

Figure 3.2. *Tasks and Tactics by Furnas [JF97]. Dark shaded circles indicate strong support for tasks; light shaded circles indicate weak support.*

3. Marchionini [Mar06] defines *searching* as a fundamental activity consisting of seeking to fulfill an information need. He defines three searching **tasks**, for which he uses the terms *lookup*, *learn*, and *investigate*, as well as two **strategies** to accomplish those tasks, an *analytical strategy* and a *browsing strategy* (cf. Figure 3.3) [Mar06, Mar95]. A summary of the search activities is presented below:

   Lookup: This is a basic task of finding information to answer questions such as who, when, and where. This is sometimes interpreted as fact retrieval or question answering. Users start with specific goals and the task requires minimal comparison of the retrieved documents.

   Learn: This type of search requires multiple interactions or reformulations and requires a significant cognitive effort to locate, interpret, and compare the retrieved results. This type of search aims at knowledge acquisition, comparison, and aggregation of information.

   Investigate: As compared to learning searches, this task involves longer periods of time and more interactions. It requires the highest cognitive effort to analyze, synthesize, and evaluate the retrieved results. The process involves the creation of annotations and artifact generation. Such artifacts become part of the search results. This type of search aims at finding new information and at finding gaps

in the existing information.

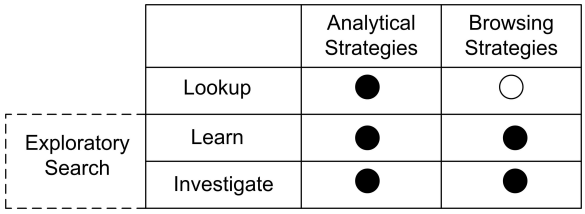

Figure 3.3. *Searching Tasks by Marchionini. Dark shaded circles indicate strong support for tasks; unshaded circles indicate absence of support.*

Turning to strategies, an *analytical search strategy* consists of carefully planned steps consisting of query formulation, query reformulation, and evaluation of results. Queries are formulated by recalling the terms to describe the information need. Sometimes using a precise syntax is required, as in a database-management system (DBMS). A *browsing strategy* is informal and opportunistic and depends on recognition of terms. A browsing strategy is appropriate when problems are not well defined or when the goal is to get an overview on a topic.

A lookup search or fact retrieval is adequate for analytical search strategies, whereas learn and investigate searches can use a mixture of both strategies. A lookup task is not adequate for a browsing strategy (cf. Figure 3.3). Learn and investigation tasks are called an *exploratory search* in recent publications by White, Kules, Drucker, and schraefel [WKDs06] and Marchionini [Mar06].

4. Spencer [Spe06] defines searching as a task with four modes: known-item, exploratory, don't know what you need to know, and re-finding.

   Known-item: In this mode, the users know what they want and know the terms to describe it. They may have an approximate idea of appropriate sources of information. The users' needs do not change much during the information-seeking process.

   Exploratory: Users have an idea of what information they need, but they do not know the right terms to formulate their need. Also, the users would not know

what relevant sources of information exist. These users can recognize information that corresponds to their needs. Users gain knowledge as they progress in their information seeking, and their needs evolve.

Don't know what you need to know: Users have an idea of what they want, but they are not aware of what they need to know. There is no clear goal to reach and users do not know when to stop searching; this is sometimes called browsing.

Re-finding: Users are seeking information they have accessed previously. They may or may not remember where the source of information was located. An example of this is when a doctor frequently revisits specific websites to find out the potential secondary effects of drugs.

The analyses presented in these four sources have much in common. They share some concepts, definitions, and categories; however, the frameworks into which these concepts are organized differ. In Table 3.1 we can see the overlap among existing searching tasks. Fact retrieval or question answering is a basic task acknowledged by Baeza-Yates, Ribeiro-Neto, Furnas, Marchionini, and Spencer. Marchionini and Spencer identify exploratory searches, although Spencer defines exploratory search without specifying the goals or on the cognitive effort required to accomplish the task. Marchionini splits exploratory search according to the users' cognitive efforts to reach their search goals. If there is a great cognitive effort, then it is an investigative search; otherwise, it is a learning search. Baeza *et al.*, Furnas, and Spencer define browsing as a task in which users engage when they do not have a specific goal. Interestingly, of all the four sources, only Spencer defines *re-finding* as a separate category of search task, despite the fact that it is a very common task among Internet users.

Although there is a general consensus that searching and browsing are different, delimiting their boundaries is not easy. For example, trying to differentiate browsing from an exploratory search can be difficult. For the purposes of this dissertation, I define *searching* as a task with a "predefined" goal whereas *browsing* is a task with an "emergent" goal.

Baeza & Ribeiro (1999)	Furnas (1997)	Marchionini (2006)	Spencer (2006)
Searching	Searching	Lookup	Known-item
		Learn	Exploratory
		Investigate	
Browsing	Browsing		Don't know
			Re-finding

Table 3.1. *Comparison of Information Seeking Tasks*

Also, I define searching and browsing on the Internet as tasks that occur interchangeably and as tasks where the desired information may not exist. I will use the term *seeking* as the task of looking for information regardless of the specificity of the users' goals. Seeking for information implies a spectrum of tasks having searching and browsing as the ends at each side of the spectrum. With these definitions in mind, we can start discussing information-seeking theories and models to explain the behavior of users seeking information on the Internet. I present the information-foraging theory proposed by Peter Pirolli in the early 1990's, followed by some traditional information-seeking models as interpreted by Shneiderman *et al.*, Marchionini, and Hearst, and lastly I describe the berry-picking model proposed by Bates.

3.2 Information Foraging Theory

Pirolli [Pir07] develops his information-foraging theory by way of an analogy with the optimal foraging theory of biologists. Optimal foraging theory explains that animals have developed strategies to hunt for food with the highest amount of calories in the least time and that animals behave as if they make a cost-benefit analysis to decide whether to move to the next patch of food. Pirolli applies this concept to the behaviour of humans seeking, gathering, evaluating, and assessing information. Humans use information-seeking strategies adapted to the rate or flow of information in a given context and make cost-benefit analyses to decide whether to move to the next patch of information. An *information patch* is a small portion of a large collection of information. Humans seek to maximize the amount of

useful information found in the least amount of time. In information-foraging theory, users seeking information as a means of adapting to the world are called *informavores*. Informavores use *information scents* to find "profitable" information patches. Just as animals use scent to guide them to their prey, humans use analogous stimuli to find relevant subsets of the information they are looking for. When information scent is strong, users follow it, and when it diminishes, they stop seeking in the current patch and move to a different one. The information-seeking process consists of two activities: searching for information patches and extracting the information found.

One problem with the information-foraging theory is that people sometimes are hunting for good patches where they can find relevant information, but on other occasions they might search for information that is the easiest to digest, the most popular, the most attractive, and so forth. One may think, for example, of users going initially to Wikipedia in search of academic information rather than being drawn to the densest patches, such as specialized journals in the target area. They may do so not because they regard Wikipedia as the most reliable source, but because it is the most convenient.

Another problem with information foraging theory is that it depicts the users as browsing for information rather than searching. Considering that the information seeking process comprises both activities, the information foraging theory is representing only one side of the spectrum. Describing only one aspect makes the information foraging theory an unbalanced view for the information-seeking process.

Having gained a theoretical understanding of the tasks and goals of users seeking general information on the Internet, we now turn to an examination of models that describe that seeking in step-by-step detail, specifically the traditional information-seeking models and the berry-picking model.

3.3 Traditional Information Seeking Models

The information-seeking process has been variously modelled with reference to the different steps within it. The resulting models have been useful in designing information-seeking systems. Of these, the most widely discussed in the literature have been those proposed by Shneiderman *et al.* [SBC98], Marchionini [Mar95], and Hearst [BYRN99]. They have become known as the "traditional" information-seeking models.

Shneiderman *et al.* describe information seeking as a process comprising four stages: *formulation*, *action*, *review of results*, and *refinement* (cf. Figure 3.4).

Formulation: This stage comprises all steps taken before submitting a query to a search engine. It includes the selection of the collection of information in which the search is to be conducted; the decision as to how the search should be limited by selecting the attributes (or tags) of the desired documents, such as author, year of publication, and so on, and possibly assigning a specified range for each of these attributes; the decision as to the search terms to be used; the decision as to whether to apply advanced features such as Boolean operators and "wildcard" characters; and the decision as to what related terms are acceptable, or what range of values is acceptable for terms that refer to variables.

Action: This is the stage in which a user formulates and submits a query to the system and waits for the results.

Review of results: This is the stage of reviewing the retrieved documents. It might include re-ordering the displayed documents by author, journal, and so on; clustering results by related themes; and exploring various values of the attributes of the retrieved set of documents.

Refinement: This is what happens when a user has finished reviewing the results of a search and before returning to the formulation stage to conduct a refined search. It includes keeping a history of previous searches, saving intermediate results, and

providing feedback to the system as to which search terms have proven relevant to the query (as well as related terms that have turned up during the search) .

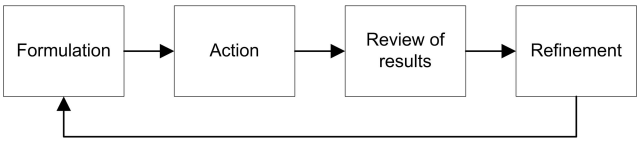

Figure 3.4. *Shneiderman's Information Seeking Model*

Marchionini proposes a more detailed model to represent the information-seeking process, comprising a larger number of more finely discriminated stages. His model has some similarity with Shneiderman's model. Marchionini's model begins by adding an initial stage in which users recognize that there is a need for information. Then, he decomposes the formulation stage into *define problem*, *select source*, and *formulate query*. He also splits the review of results stage into *examine results* and *extract information*. Finally, he adds a *reflect or stop* stage (cf. Figure 3.5). The stages of Marchionini's model, described in more detail, are as follows:

Recognize and accept an information problem: In this stage users become aware that there is an information need and decide whether to pursue an answer or ignore the need.

Define and understand the problem: Users identify key terms relevant to the information need. This stage includes delimiting the scope of the problem and formulating hypotheses as to the possible answers. The result of this stage is a called a "formalized need". In this stage users decide on a plan to solve the information problem.

Choose a search system: Users select the digital collection to search from. They base their decision on previous experience and on their knowledge of the information domain.

Formulate a query: This stage consists of matching the information need to the syntax and semantics of the selected search engine.

Execute search: In an electronic environment this stage consists of submitting a query to the system or browsing the system following hyperlinks to find the desired information.

Examine results: The retrieved documents are viewed as intermediate results and examined to determine the accuracy and usefulness of the information they contain.

Extract information: This stage consists of extracting information from the documents selected as relevant in the previous stage (examine results). Typical activities to extract information include reading or skimming the documents and classifying, copying, and storing the information extracted from them.

Reflect/iterate/stop: At this stage users decide whether they have found what they were looking for or whether they will refine their query and iterate the seeking process.

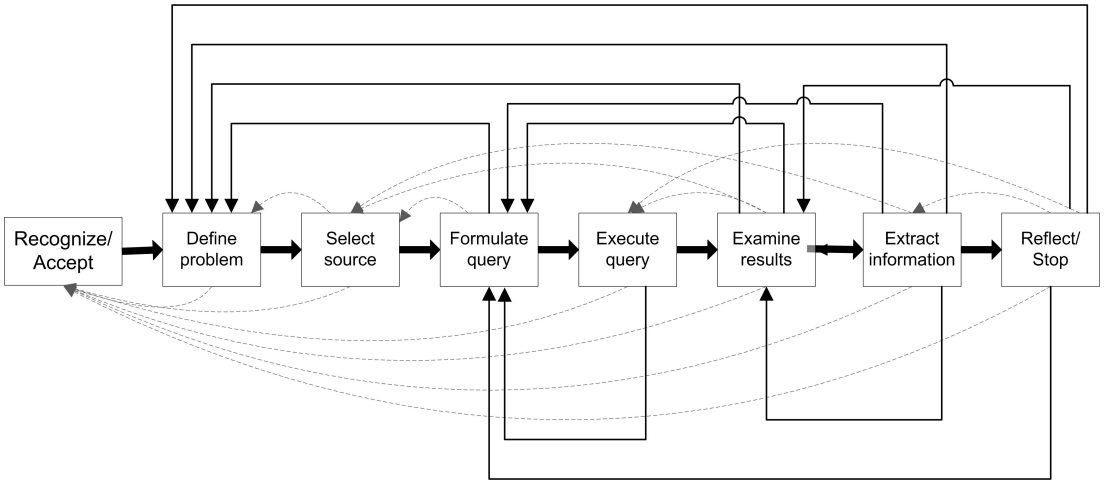

Figure 3.5. *Marchionini's Information Seeking Model [Mar95]. Boldface arrows indicate logical sequence of steps; solid lightface arrows indicate probable iteration; dashed lightface arrows indicate possible iteration.*

Hearst's model of the information-seeking process is represented in Figure 3.6. Her model overlaps considerably with Marchionini's; however, she omits the extract information and reflect stages. On the other hand, similar to Shneiderman's model, Hearst's model

includes *reformulate* as an explicit stage, while in Marchionini's model it is merely assumed. Hearst's components of the information-seeking process are as follows: *information need, query, send to the system, receive results, evaluate results, done,* and *reformulate*.

Information need: This stage describes the beginning of the process, when users recognize they have an information problem and select an information collection to search from.

Query: In this stage users formulate a query.

Send to the system: This is the stage when users submit the query to the system.

Receive results: This stage is when users retrieve the results of their query.

Evaluate results: In this stage, the users browse and assess the retrieved documents.

Stop: If the users have found the needed information, they end the process at this stage.

Reformulate: If the users are not satisfied with the information they have retrieved, they re-specify the query and iterate the seeking process.

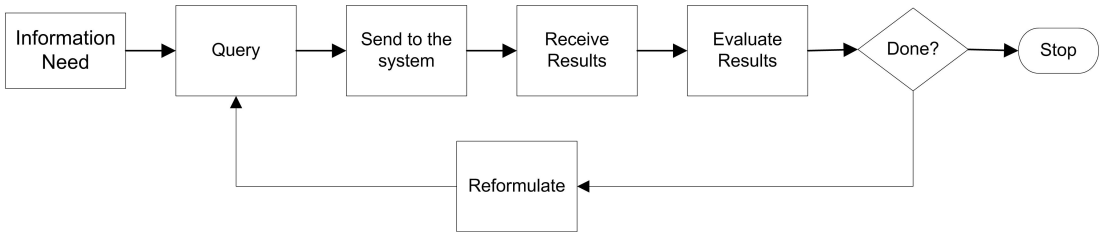

Figure 3.6. *Herst's Information Seeking Model [BYRN99]*

A fundamental problem with traditional information-seeking models is that they imply two unwarranted assumptions: that an answer to the information need exists, and that the seeking process will stop when users find it [BYRN99, Bat89]. Regarding the first assumption, it may be the case that no research has been published on the question or that the results of any such research have been inconclusive. Regarding the second, the users' information needs do not necessarily remain static throughout the process. Users increase

their understanding of the underlying domain searched as they scan titles of retrieved results, explore hyperlinks, read content, and view lists of related terms. As a consequence, their needs change throughout the seeking process.

In addition to the fundamental problem, there is another small issue with these models. They imply that users fulfill their information need by iterating on searching tasks (query formulation and reformulation). This assumption leaves aside the case when users begin by browsing and then switch to searching when they have acquired the vocabulary that they need. These models also ignore the case when users alternate between searching and browsing as suggested in Figure 3.1.

3.4 Berry-picking Information Seeking Model

The berry-picking model was originally developed exclusively for scholarly use, although it is also suitable for more general application. It receives its name in analogy with the action of picking berries. The model was proposed by its creator, Bates [Bat89], to describe an *evolving search* rather than a static search as implied by the traditional information-seeking models (cf. Figure 3.7). In the traditional models, the query is represented as being formulated in response to a static information need, although it may be iteratively rephrased to find better results. In the berry-picking model the information need is not static but changes dynamically throughout the process, the queries are adjusted to these changes, and users employ a variety of strategies during the process interchangeably. The information need changes in response to the documents retrieved in the evolving search or to the discovery of new ideas to explore. The users seek and extract information from different sources in the information space in which they are interested. In the berry-picking model the goal is not to find one target result set but rather to extract bits of information during the process. For Bates, the most important aspect of the process is the learning that occurs while it is under way. Users employ alternative techniques iteratively when seeking information, such as searching and browsing. Bates applies her model to manual searching

and identifies the following strategies employed by users during the information-seeking process:

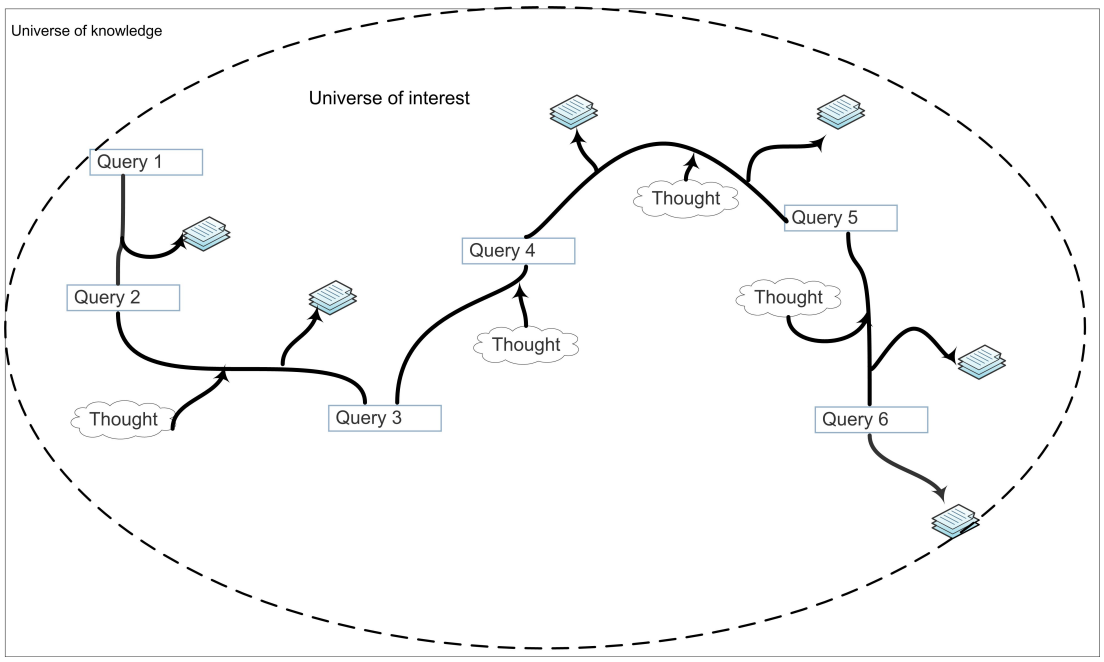

Figure 3.7. *Bate's Berrypicking Model [Bat89]. Rectangular boxes represent queries formulated in chronological order; clouds represent thought processes; stacked documents represent query results.*

Subject searches: This type of search consist of a description of the subject sought.

Author searching: This is a "known-item" search consisting in users looking for the published work of a particular author.

Chaining: This strategy consists of the following citations or footnotes. It takes two forms: forward chaining and backward chaining. Forward chaining occurs when users seek additional documents that cite a document they have already retrieved. Backward chaining means that users follow the bibliographic references they find in a particular document.

Journal run: This strategy consists in browsing a journal known to be relevant to the topic of inquiry. Users identify particular numbers of interest and skim through them.

Area scanning: Consists of browsing information that is physically located close to a source of information already identified as relevant.

In the berry-picking model, a user can initiate a search using one strategy and subsequently iterate it using a different strategy. For example, Queries 1 and 2 in Figure 3.7 might be formulated as subject searches. Then, following a thought process, Query 3 might use a chaining strategy, and so forth.

A potential problem with Bates's berry-picking model is that it is based on the results of older studies in which people were manually seeking information. Although the model and corresponding strategies are intuitively appealing, Bates does not bring any evidence that the needs and behaviour of the online user are the same as those of the user of manual information-retrieval systems.

Furthermore, Bates emphasizes the importance of extracting information along the information-seeking process without considering the aggregation and the comparison of such pieces of information. Those activities are needed to answer the users' information needs.

3.5 Summary

In this chapter I have discussed the differences and similarities among the terms searching, browsing, and seeking. I have also described the often cited Information Foraging Theory and have discussed the imbalance of this theory for the information seeking process. I have discussed the traditional information seeking models from Shneiderman, Marchionini, and Hearst and I have compared their stages where appropriate. I have also described the berry-picking model and discussed how this model lacks a stage to describe how the user integrates information found during the seeking process.

In Chapter 7, I discuss the difficulties faced by users seeking information in the domain of clinical trial data and I present my model, the Integrated Information Seeking Model which leverages in the models here described and in the Information Foraging theory. Be-

fore providing such a discussion, however, it is useful to present some empirical data on the behaviour of users seeking health and biomedical information on the Internet. To this end, I devote the following chapter to a review of the results of some empirical studies on the information needs of these type of users and on the use of both general and specialized search engines to retrieve such information.

Chapter 4

Empirical Studies of Medical Information Seeking

Understanding the behaviour and needs of users searching for information on the Internet is important for the design of effective interfaces [HTHB07, ZZZR04]. This chapter is concerned with the analysis of current studies investigating the behaviour of users seeking medical information on the Internet.

Searching for information is considered to be the most important activity for most Internet users [JP01]. According to Pew Internet Survey, 80% of American Internet users surveyed in 2002, 2004, and 2006 said that they had searched for health information (not limited to clinical trials) [FF06]. Similar results are reported by Harris Interactive from surveys conducted from 1998 to 2008, that is, 71% to 81% of American Internet users said that they had searched for health information on the Internet [Tay05].

In a survey conducted in 2003, Bennet [BCKS04] found that 60% of physicians use the Internet daily or weekly to seek clinical information. A survey conducted in 2002 by the Health on the Net Foundation found that searching for clinical-trial data was the third most common type of search conducted by medical professionals and patients seeking medical information on the Internet [MPB02]. Despite the prevalence of clinical-trial searches on the Internet, research on users' behaviour in this domain is scarce.

4.1 Medical Information Seeking

Most empirical studies on medical information seeking aim at understanding the information needs and behaviour of physicians in their clinical practice. These studies have used a variety of research methods. In this section, I discuss the main findings on medical information seeking using surveys, interviews, and observations. Later in the chapter I discuss the findings of various query-log studies.

Frequency of questions: Several researchers have investigated how often physicians encounter questions in their clinical practice for which they have no immediate answer. The studies most frequently cited are those of Covell *et al.* [CUM85], Ely *et al.* [EOE+99], and Smith [Smi96], whose findings vary somewhat. Covell found that physicians ask themselves two questions for every three patients, while Ely found that they themselves ask approximately one question for every three patients. However, Smith argues that physicians underreport their questions and estimates that physicians really ask at least one question for every patient.

Seeking answers: Since questions arise so frequently in clinical practice, one may ask whether and to what extent physicians seek to answer those questions. Gruppen [Gru90] examined the reasons physicians have for seeking answers to their questions; he found that they seek answers to solve patient-care problems, to find general care information, for purposes of patient education, out of curiosity, and for research purposes. Unfortunately, most questions remain unanswered. Covell *et al.* [CUM85] found that physicians pursue only 30% of the questions that arise in clinical practice. Similarly, Ely *et al.* [EOE+99] found that only 35% of the questions are pursued. Moreover, the time physicians invest in answering a question is extremely low; they spend less than two minutes per question on average, while medical librarians, for example, devote 27 minutes per question [EOE+99].

Obstacles to seeking answers: The majority of questions physicians ask themselves do not receive an answer, either because the answer was not pursued or because no an-

swer was found [EOC+05]. The reasons physicians provide for not seeking answers to their questions are: lack of time to search [GCE00, EOG+00, CUM85], forgetfulness [GCE00, EOE+02], out-of-date textbooks [CUM85], inadequately organized personal libraries [CUM85], the lack of a feeling of urgency [GH95], and the feeling that no answer exists [EOC+05, EOG+00, GH95].

When physicians seek answers to their clinical questions, they encounter a number of obstacles. Of these, the most frequently mentioned are difficulties in navigation and searching followed by the overwhelming amount of medical information available. Comparing the findings of Bennet *et al.* and Casebeer *et al.* (see Table 4.1), it appears that the perception of obstacles increased dramatically from 2001 to 2003 [BCKS04, CBK+02].

In another study, Ely *et al.* [EOE+02] enumerates six obstacles physicians face in answering clinical questions: the considerable expenditure of time required to find the answer, the effort required to refine the initial search which is often vaguely formulated, the effort required to find an effective searching strategy, the difficulty of finding the resource that has the desired content, the difficulty of deciding when to stop searching, and the difficulty of aggregating several retrieved items of information to formulate an answer to the question. In a more recent study, Ely *et al.* [EOC+05] found that the main obstacle physicians face in answering their questions was that they targeted a resource which lacked the desired information.

Physicians' information resources: The main sources of information used by physicians are printed materials (textbooks, journals, and drug information sources) and other medical professionals (physicians, pharmacists, librarians, laboratory personnel, and social workers) [Gor01, GCE00, CFJ+00, EOE+99, BCKS04]. Smith [Smi96] attributes the preference for human experts to the psychological reassurance physicians receive from such trusted sources; he also cites Covell [CUM85] to the effect that physicians consult such experts more frequently than they admit.

According to Bennet [BCKS04], the factors that affect the choice of a particular source of information are credibility, relevance, unlimited access, speed, and ease of use. Connelly [CRCK90] provides a similar list of factors: availability, searchability, understandability, credibility, and applicability.

Information Needs: The main reasons related to clinical practice that physicians cite for using the Internet are to conduct literature searches and to access online information. Surveys conducted in 2001 and 2003 [BCKS04, CBK+02] indicate significant increases in the use of the Internet by physicians to find patient-specific information, to take continuing education courses, and to obtain updated information from professional associations. No changes were observed in the frequency of Internet use to consult colleagues and to prescribe medications for patients (see Table 4.2).

Similarly, Bennet *et al.* [BCKS04] found that the reasons physicians have to search for clinical information are to obtain up-to date-information on specific topics, new information on a medical condition, information relevant to the diagnosis or treatment of a specific patient, drug dosage information, and information on a new drug or treatment.

Covell *et al.* [CUM85] studied the frequencies with which physicians ask various categories of questions. He found that 40% of the questions were about facts (e.g., "What are the reported side effects of bromocryptine?"), 43% were about medical opinions (e.g., "How do you manage a patient with labile hypertension?"), and 17% were non-clinical questions (e.g., "How do you arrange for home care for a patient?"). A more detailed categorization showed that 31% of the questions concerned treatment of a medical condition, 25% concerned diagnosis or symptoms, 14% were requests for drug information, 11% were communications with colleagues, 8% regarded patient services, 6% sought a general review of a specific medical condition, and 5% concerned the interpretation of laboratory results.

Table 4.1. *Obstacles physicians face to find answers to their questions [BCKS04, CBK+02]. Percentages correspond to physicians surveyed.*

Obstacles	2001	2003
Navigation/searching difficulties	16%	57%
Too much information to scan	22%	49%
Specific information not available	20%	45%
Slow load times	11%	29%
Software incompatibilities	2%	19%

Table 4.2. *Internet use by physicians in 2001 and 2003. Adapted from Bennet* et al. *[BCKS04] and Casebeer* et al. *[CBK+02]. N.A. means information not available.*

Purpose	2001	2003
Literature searching	65%	72%
Accessing on-line journals	45%	65%
Searching for patient-specific information	29%	57%
Continuing medical education courses	31%	46%
Professional association updates	33%	45%
e-mail patients	N.A.	20%
Consultation with colleagues	17%	19%
Prescription/patient orders	2%	4%

Thus, our review of the literature based on surveys, interviews, and observations shows that physicians frequently encounter medical questions in their clinical practice to which they are unable to provide immediate answers. These questions, most of which are complex and multidimensional [Smi96], usually remain unanswered. Perhaps the most common reason for this failure is that physicians are too busy to devote the time required to obtain the answers. However, when they do seek answers, they are overwhelmed by the vast quantity of information they must sift through in the process, and they encounter instrumental difficulties in navigating and searching the Internet to obtain these answers. From the standpoint of human-computer interaction, it would appear that improvements in user interfaces for the specific purpose of seeking medical information on the Internet might alleviate these difficulties and thus might also reduce the demand that finding answers to clinical questions places on a physician's time and on other groups of users (such as medical students, nurses, etc.) that have been less studied in the current literature.

4.2 Query-log Analysis

Another prevalent strategy for studying the behavior of users retrieving information from the Internet is *query-log analysis*.

A query log is a file that stores information on users' actions during Internet searches. Typical metrics to consider in a query-log analysis are the number of queries per session, the number of terms in a query, the number of Boolean operators used, the number of results obtained, and the most common search terms used.

Transactions in a query log include a *query* as submitted by the user, a *time-stamp* to identify when the query was received by the server, an anonymous *user* identification assigned to a user-machine, and code representing the set of *results* in response to a query [JS06a]. The real IP addresses is not revealed; the data is encrypted by the corresponding servers to protect the privacy of the users.

Most of the published research using query-log analysis to investigate user behaviour focuses on general-purpose search engines. Although the leading search engines, namely Google with 53.5% of all searches, Yahoo! with 19.9%, and MSN with 12.9%[1], conduct their own research, they rarely publish it, and they do not release their query logs to independent investigators. Hence, published studies concern other general-purpose engines, such as America Online, Altavista, Fireball, BWIE, AlltheWeb.com, and Excite [EJ03, SMHM99, SWJS01, JS06b]. Most such studies focus on users searching for general information. Only a few studies to date have focused on health and biomedical searches, and none, to my knowledge, specifically on searches of clinical-trial data.

Below, I discuss the current scholarly literature on the behaviour of users searching for biomedical and health information on the Internet. In the course of an extensive study of the relevant literature, I identified five peer-reviewed articles reporting the results of eight query-log analyses concerned with biomedical information. The findings of these five articles form the basis for the discussion that follows. First, I describe the search engines

[1] http://searchengineland.com/nielsen-netratings-august-2007-search-share-puts-google-on-top-microsoft-

under study; secondly, I discuss the vocabulary used in these studies, including terms such as session, query, term, and user; thirdly, I provide a comparative analysis of the findings of these studies.

4.3 Descriptive Information

Eight query-log studies are reviewed in this chapter; these studies examined the behaviour of individuals using the following engines to search for biomedical or health information: Excite, AlltheWeb.com, PubMed, SHINE, MyWelch, ASKJeeves, Find-a-Doctor, and MEDLINEplus. Excite, AlltheWeb.com, and ASKJeeves are general purpose search engines designed primarily for use by the general public, while PubMed, SHINE, MyWelch, MEDLINEplus, and Find-a-Doctor are specialized health or biomedical search engines designed primarily for use by healthcare professionals. Refer to Table 4.3 for a summary of the descriptive information of the studies used in this discussion (note that the described "type of users" is only an inference, since the query-log data does not distinguish between lay and professional users).

Excite, AlltheWeb.com, and ASKJeeves: *Excite* and *AlltheWeb.com* are engines to search for general information, which were somewhat popular in mid 1990's and early 2000's respectively. Several query-log studies have been conducted on the use of these two engines [SWJS01, JS06b, SYJ+04]; however, for the purposes of this dissertation, I discuss only a study by Spink *et al.* [SYJ+04] whose focus was restricted to health and medical queries. The logs studied by these researchers date from 2001 and include some 1.2 million queries from each engine, from which each had samples of 10,000 to be analyzed. In the same journal paper, Spink *et al.* analyzed a query log from ASKJeeves. *ASKJeeves* is a search engine designed to receive general questions posed in natural language, as well as the typical keyword-based queries. They restricted the ASKJeeves log to queries seeking medical or healthcare advice posed as natural-language questions; this was accomplished by restricting the

sample to queries containing medical or health terms and the word "should", for example, "how should I cure dandruff?". The log was comprised of 800,000 queries from which only 332 queries comply with the desired features.

PubMed and MEDLINEplus: *PubMed* and *MEDLINEplus* are search engines to access the MedLine database. MedLine is the largest database of biomedical information and belongs to the U.S. National Library of Medicine (NLM). PubMed is a search engine provided by the U.S. National Institute of Health (NIH), which has been available since 1997. PubMed was launched with the goal of making scientific information available to the public; however, estimates suggest that the majority of PubMed users are healthcare professionals, medical librarians, and researchers, while only one-third are members of the general public [LM02]. The query-log study on PubMed was conducted by Herskovic *et al.* [HTHB07] using a log dated 2005 containing 2.7 million queries. The researchers used two samples, one consisting of the entire 2.7 million queries and the other of 2,708 queries selected at random. The first was used to compute general statistics, such as number of users in one day, number of queries per day, number of queries per user, and so on. The second sample was used to classify queries using the taxonomy of *transactional*, *navigational* and *informational* queries proposed by Broder [Bro02]. MEDLINEplus aims at providing information to healthcare providers and patients. The query-log analysis of the MEDLINEplus engine was conducted by Zeng *et al.* [ZKA+02]. The log they used dated from 2001 and comprised 17,077 queries, of which the researchers extracted a sample of 1,053 for analysis.

MyWelch: The Query log study on MyWelch was conducted by Zhang *et al.* [ZZZR04]. *MyWelch* is a web portal designed for users of the Johns Hopkins University Medical Library to access a diverse number of medical-information resources. The query log dated from late 2002 to early 2003 and consisted of 16,796 information-seeking actions. An action is any physical action related to the following types of interaction

with the system: customization, searching, browsing, communicating, monitoring, and service-requesting. The study was limited to analyzing the actions of 100 highly active users such as clinicians, non-clinical faculty, students, and staff in medical science.

SHINE: *SHINE* provides information on drugs and diseases. SHINE stands for Stanford Health Information Network for Education and is a search engine designed to be operated at the point of care by clinicians to satisfy their information needs. The SHINE study was conducted by Strasberg *et al.* [SHRM99] with data from January 1998 to January 1999. The size of the log was 25,389 queries from which 380 queries were selected to be analyzed. The sample was created by selecting every 52nd query from the original log.

Find-a-Doctor: *Find-a-Doctor* is a search engine designed to find a physician by last name, first name, title, department, clinical interest, language, and board certification. Find-a-Doctor is used by users of the Brigham and Women's Hospital (BWH). The study was conducted by Zeng *et al.* [ZKA$^+$02]; the log consisted of 187,994 queries collected in 2000, from which a sample of 11,182 queries was selected, consisting of queries in which terms had been entered in the field "clinical interest."

4.4 Definition of names

Conducting a comparative analysis of query-log studies is problematic, since there is little consistency in the terms used in these studies. To assist in this discussion I describe the corresponding definitions that apply to the query log analyses described here.

Session: The Excite and AlltheWeb studies define a session as all consecutive queries from the same user without time restriction [JS06a]. The PubMed study defines a session as all consecutive queries from the same user, which use search terms that are semantically close to each other, without time restriction. For example, a user submitting

	Year of dataset	Log Size	Sample Size	Assumed type of queries	Targeted users
Excite	2001	1.2 million queries	10,000 queries	medical and health queries	General public
AlltheWeb.com	2001	1.2 million queries	10,257 queries	medical and health queries	General public
ASKJeeves	1999	800,000 queries	332 queries	medical and health queries (advice-seeking)	General public
PubMed	2005	2.7 million queries	2.7 million queries	biomedical	General public, health professionals, medical librarians, and researchers
MEDLINEplus	2001	17,077 queries	16,743 queries	medical and health queries	General public and health professionals,
MyWelch	2002-2003	16,796 actions	16,796 actions	medical and health queries	Clinicians, non-clinical faculty, students, and staff in medical science
SHINE	98-99	25,389 queries	380 queries	medical queries	Primary care physycians
Find-a-Doctor	2000	187,994 queries	11,182 queries	medical and health queries	General public, excluding health professionals

Table 4.3. *Summary of Descriptive Information of Query Log Studies*

consecutively the queries *degenerative join disease*, *degenerative arthritis*, and *diabetes* is considered as having one session in the Excite study and is considered as having two sessions in the PubMed study. The MyWelch study defines a session as a set of human-computer interactions within a given time interval.

Query: The Excite and AlltheWeb studies define a query simply as a list of terms entered to the search engine. For the PubMed study, a query is a question submitted by a human user and not by a crawler. Knowing when a query was submitted by a crawler or by a human user is difficult, the PubMed study assumed that a human user would not submit 50 or more consecutive queries. Thus the authors selected only queries from users submitting less than 50 consecutive queries in an attempt to eliminate queries from crawlers.

Term: The Excite and AlltheWeb studies define a search term simply as any unbroken string of alpha-numeric characters. In the PubMed study a term is defined as a string of characters separated from other such strings by punctuation or spaces, or a string

	Excite	AlltheWeb.com	PubMed	SHINE	ASKJeeves	MEDLINEplus	MyWelch	Find-a-Doctor
Q1. What are the features of biomedical and health searches? (# of queries, # of terms, #of results)	X	X	X			X		X
Q2. What are the biomedical and health topics searched for?			X	X	X	X		X
Q3. What are the types of medical and health queries searched? (informational or navigational)			X					
Q4. What search strategies are used in medical and health searches	X	X	X			X	X	
Q5. What is the success of medical and health searches ?			X		X	X		X

Table 4.4. *Research Questions Empirically Investigated*

contained between quotation marks or square or curly brackets. For example, the size of the query *"degenerative joint disease"* for the PubMed study is one and three for the Excite study. For the MEDLINEplus and Find-a-Doctor query logs studies a search term is defined as one or more words; thus, for example *bones of the hand* is considered one term consisting of four words.

Word: Since a term in the MEDLINEplus and Find-a-Doctor query logs studies can consist of more than one word, the authors also provide a definition of a word as a continuous string of alphabetic characters.

For the purposes of this discussion, and for the sake of simplicity, I discard the terminological distinction used in MEDLINEplus and Find-a-Doctor, using *term* in the contexts in which those search engines use *word*.

4.5 Discussion

An analysis of these studies identified the research questions with which they are concerned. These questions are itemized in Table 4.4 and discussed below.

Q1. What are the features of medical and health searches on the Internet? Three main aspects have been investigated which are the number of queries per user, the number of terms per query, and the use of Boolean operators.

- The average number of queries entered per user in Excite and AlltheWeb.com studies is 2.2, in the PubMed study it is 4.3, and in the other studies it is not reported.
- The average number of terms per query in the Excite and AlltheWeb.com studies is 2.3 terms. The average number of terms per query in the Find-a-Doctor study is 1.5 terms, and 63% of all queries were composed of a single word; less than 10% of all queries used more than 3 words. The average number of terms per query in the MEDLINEplus study is 2, and 72% of all queries were composed of one or two terms; less than 12% of all queries used more than 3 words. The PubMed study does not report the average but the median of terms per query, which is 3.
- The number of queries using Boolean operators in PubMed is 11.2%; in the other studies the usage of Boolean operators is not reported.

In summary, the results show that users make only a few queries per session, use very short queries, and may use Boolean operators infrequently.

Q2. What are the medical or health terms and topics searched for on the Internet? The PubMed study reports that the most frequent types of queries (45%) concern *chemicals/drugs* and *diseases*. The next 38% of all queries is distributed almost evenly among categories such as *biological sciences*; *anatomy*; *organisms*; and *analytical, diagnostic, and therapeutic techniques and equipment*. The most frequent attributes (or tags) used in submitted queries are author's name, PubMed ID, date of publication, name of journal, page, volume, and title. The most common search terms are cancer, cell, review, disease, protein, and human.

The SHINE study classified queries into 33 categories. *Drugs* and *infectious diseases* together accounted for 25% of all queries, while the other 75% was widely dispersed among the remaining 31 categories, such as gastroenterology, neurology, cardiology, oncology, endocrinology, or pulmonology.

The ASKJeeves study found that 35.6% of all queries concerned *general medical or health* questions, followed by questions about *human relationships* and *weight* with about 20% each. *Reproductive health/puberty* counted for 13.8% of all queries and *pregnancy/baby* counted for 8.5% of all queries.

The MEDLINEplus study reported that the top seven most common terms used in submitted queries corresponded to *diseases*, with the exception of the drug Lipitor. The seven most searched terms were diabetes, Lipitor, fibromyalgia, lupus, breast cancer, cancer, and depression.

The Find-a-Doctor study reported that the top seven most common terms used in the queries were cardiology, internal medicine, neurology, oncology, gynecology, and psychiatry.

Q3. What are the types of medical and health queries searched? Only the PubMed study considered the "type" of query posed in a search. The study classified queries as informational and navigational using the names from [Bro02]. *Informational* queries aim at finding specific information which can exist on more than one document. *Navigational* queries aim at finding a particular document. The results showed that three quarters of queries in PubMed were informational (75%) and one quarter were navigational (25%).

Q4. What search strategies are used in these searches? There is relatively little published information on the strategies used while searching for medical or health information. The PubMed study observes that users who refine their queries will either add or remove terms with the same probability. The Excite and AlltheWeb.com study observes that few users reformulate their queries but does not give the specific

number. Also, it notes that few users use synonymous terms in their queries. The MyWelch study classified the user actions stored in the log file into six categories of human-computer interactions: *starting* (i.e., identifying sources of interest) represented 14.42% of all actions, *searching* (i.e., submitting a query) 5.66%, *browsing* (e.g., reading table of contents and headlines) 6.19%, *monitoring* (i.e., visiting favorite Web pages) 47.38%, *extracting* (i.e., selecting useful pages) 2.34%, *service requesting* (e.g., checking library accounts, requesting document delivery) 4.95%, and *communicating* (i.e., communicating with colleagues or librarians) 7.41%.

Q5. What is the rate of success of medical and health searches? The PubMed study reported that from a sample of 2,708 queries, 16.1% did not produce any results and that the average number of documents retrieved in response to a query was 14,050.

The ASKJeeves analysis reveals that many users have difficulties expressing their information needs. They often lack understanding of the limitations of a search engine; the study cites examples of queries submitted as if talking to a human expert, such as "Help Me Jeeves ...", "May I ...", or "Please ...". The study also reveals that users lack the appropriate technical vocabulary to formulate effective medical or health queries. For example, the term blood sugar, while common in the vernacular, seldom occurs in scientific literature.

The MEDLINEplus and Find-a-Doctor studies mapped the terms used in a query to the Unified Medical Language System (UMLS). UMLS comprises a set of medical vocabularies such as MeSH, ICD-9, and SNOMED. This mapping revealed a significant proportion of mismatches between the queries entered by the users and the vocabularies recognized by the search engine. The authors identify these mismatches as the reason why many searches fail.

The Find-a-Doctor study revealed that 51% of all queries did not retrieve results although the information existed in the database. This study also classified the mismatches into three types and found that 35% of all mismatches are *lexical*, 58% are

Type	Subtype	Example:	
		Query term	UMLS Name
Lexical	Spelling error	Rhumetology	Rheumatology
	Morphology	Addictions	Addiction
	Concatenation	Anti phospholipid	Antiphospholipid
	Sequence	Headache migraine	Migraine headache
Semantic	Abbreviations and Acronyms	ADHD	Attention Deficit Disorder
	Synonym	Bone spurs	External extoses
	Redundancy	Cardiac arrythmia	Arrithmia
	Generalization/Specialization	Diabetic leg ulcers	Leg ulcer
	Other Semantic relationship	Retinal surgery	Retinal disease
Other	Valid term, but not in UMLS	Genomics	
	Non-medical term, not in UMLS	Mass general Hospital	
	Unclear meaning of consumer term	Scc	

Table 4.5. *Classification of term mismatches [ZKA+02]*

semantic and 12% correspond to *other* reasons. In Table 4.5 these mismatches are subdivided and examples are provided.

The Excite and AlltheWeb.com studies do not report on the rate of success of health and medical searches. However, these logs reveal that users hardly look beyond the second page of the list of retrieved results and seldom navigate to websites other than the first 10 to 20 listed.

4.6 Summary

The empirical studies discussed in this chapter indicate that although physicians pose many questions derived from their daily clinical practice, only a reduced number of questions posed are further investigated by them. The reasons why physicians do not seek an answer vary; the lack of time and inadequate resources are often cited as main reasons for not seeking answers. When seeking answers, physicians have to deal with difficulties searching, navigating and comparing the findings of an overwhelming amount of information. Among the sources of information used by physicians, the human expert has been found to be as important as printed materials. The information needs identified in the current literature are mainly related to patient care and education, such as information on medical

conditions, diagnosis, treatments, and interpretation of laboratory results.

From the query-log studies discussed in this chapter, we know the behaviour of *users of medical and health information* on the Internet: users formulate short queries (except PubMed users to some extent), write few queries in one session, review only the first few results, and rarely use advanced search capabilities such as Boolean operators. These findings are consistent with the behaviour of users searching for general information [JS06a, SMHM99, SWJS01].

The results indicate that users of the Internet searching for medical or health information are mainly interested in *medical conditions or diseases* and in *therapies*. When they reformulate their queries in an attempt to retrieve better results, they are equally inclined to add or remove terms. Most of the queries aim at finding specific information rather than particular documents.

My analysis of query-log studies also revealed two problems users encounter in seeking medical information. The first is that they sometimes do not understand the technical limitations of search engines as compared to human interlocutors, and thus have difficulty formulating their searches effectively. The second is that they frequently struggle with formulating their queries, since they lack the technical vocabulary used in the scientific documents they seek to retrieve. As one might expect, this second problem is specific to lay users as opposed to medical professionals. Thus, it turned up primarily in the study of Find-a-Doctor, a search engine used almost exclusively by lay users.

This finding brings us to the question of the types of users of medical information (cf. the Targeted Users column in Table 4.3). It is evident that such information is sought both by medical professionals and by members of the general public, such as persons who suffer from an illness or have a friend or relative who suffers from one. The information needs of these two categories of users may differ, as may their level of competency in fulfilling those needs. These differences have not been investigated in the published literature. Although such an investigation is not a major focus of this dissertation, the findings of the exploratory study discussed in Chapter 6 shed some light on this issue.

Part II

Exploratory Studies

Chapter 5

Research Design

In this dissertation I investigate the behaviour of the user searching for clinical-trial information on the Internet in order to provide design principles to improve user interfaces that support the information-seeking process. This goal can be reformulated as research questions: **Q1:** What are the information needs of users searching for clinical-trial information on the Web? and **Q2:** How can we improve current Web-search interfaces for seeking clinical-trial information? Due to the exploratory nature of my research questions, I follow a qualitative approach which consists of the collection and analysis of qualitative data. Qualitative research is used for exploratory studies where little is known regarding a phenomenon [Cre04, SSI07].

In this chapter, I describe the methods I used to investigate these research questions. The first stage of the investigation is exploratory and is intended to build an explanation of the behaviour of users seeking clinical-trial data. This explanation provided the answer to research Question 1. Once I explain user behaviour, I extrapolate from it the requirements that an interface for users of clinical-trial information should satisfy. I use this set of design principles to answer research Question 2. Next, I move on to a confirmatory stage of my investigations. In the confirmatory stage, I design two visualization tools to support the user in the information-seeking process and have users test them to evaluate their effectiveness. These investigations confirmed the constructed explanation and design principles (cf. Figure 5.1).

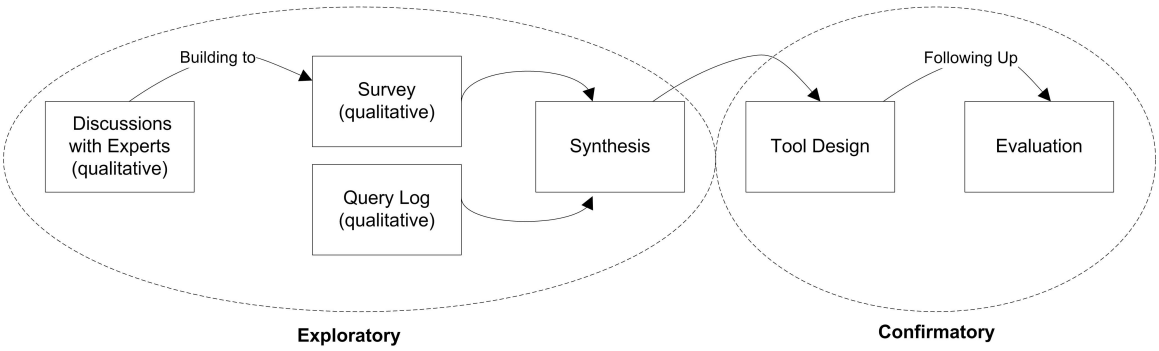

Figure 5.1. *Research Design*

5.1 Exploratory stage

The existing literature is lacking in studies specifically devoted to the needs of Internet users seeking information on clinical trials. As such, I undertook some fresh research in this area. In the exploratory stage, I elaborated Question 1 into a series of subsidiary questions concerning the behaviour of users seeking clinical-trial information on the Internet (**Q1**):

Q1.a: For what reasons, or to support what tasks, do users seek clinical-trial information?

Q1.b: What kinds of questions do users attempt to answer?

Q1.c: How do users seek information (i.e., tools used, frequency of searches, terms used)?

Q1.d: What search strategies do users employ to answer their questions?

Q1.e: What is the success rate of users?

Q1.f: To what extent are users satisfied with the information-seeking tools they employ?

To answer these research questions, I used three data-collection methods. First, I conducted a series of moderated discussions with two medical researchers who were experts in the field of clinical-trial research. Second, I designed an online questionnaire, which was made available to users of clinical trials. Third, I carried out a qualitative analysis of queries executed on a widely used biomedical search engine. The data-collection methods are discussed in the following sections. I used these three data-collection instruments in

order to compensate for the weaknesses of each and to validate my findings. The strategy of using more than one method is called "triangulation" [Cre04, SSI07]. The subsidiary research questions and the research strategies used are depicted in Table 5.1. The findings of the three exploratory studies are discussed in Chapter 6.

5.1.1 Moderated Discussions with Experts

I conducted multiple moderated discussions with two biomedical researchers from UCSF who were experts in clinical trial data. These had the purpose of identifying users roles, tasks and tool requirements. This communication took the form of teleconferencing sessions complemented by e-mail interactions. The teleconferencing sessions took place on average once a month for a period of 18 months.

The need for a tool to search and compare collections of clinical trials emerged clearly during my interviews with these investigators, who expressed their need for an efficient, intuitive tool to help them explore and understand previously published clinical trials.

5.1.2 Online Questionnaire

After moderated discussions with the experts, I wanted to validate my findings with a larger sample of users of clinical-trial data. To this end, I created an online questionnaire. Six hundred invitations to participate in the questionnaire were sent via email to webmasters at universities, research institutions, and other organizations in English-speaking countries, which either conducted clinical trials or offered courses about them. The study was conducted from 1 August to 6 November 2008 and elicited a total of 53 responses. The questionnaire consisted of a combination of multiple-choice, close-ended, and open-ended questions (nine questions in total), designed to obtain a profile of the users and to determine their needs, actions, and desires with regard to information seeking in the clinical-trial domain (cf. in Table 6.2 the list of questions asked).

Question	Discussions w/Experts	Online Quest.	Query Log Analysis
Q1.a:For what reasons, or to support what tasks do users seek clinical-trial information on the Internet?	*X*	*X*	
Q1.b:What kinds of questions do users of clinical-trial information attempt to answer?	*X*	*X*	
Q1.c:How do users seek clinical-trial information (tools used, frequency of searches, terms used)?	*X*	*X*	
Q1.d:What search strategies do users of clinical-trial information employ to answer their questions?			*X*
Q1.e:What is the success rate of users of clinical-trial information?	*X*	*X*	
Q1.f:To what extent are users of clinical-trial information satisfied with the information-seeking tools they employ?	*X*	*X*	

Table 5.1. *Exploratory Questions and Research Methods*

5.1.3 Query Log Analysis

I undertook a query-log analysis to examine that behaviour of users seeking medical information on the Web and to complement the moderated discussions and survey results with a more objective research instrument. I chose a qualitative analysis for the query-log because it is appropriate to the exploratory nature of my research questions, and I selected PubMed, a leading search engine in biomedicine, as the database.

5.2 Synthesis of findings of exploratory stage

Once I had an understanding of the tasks undertaken and questions posed by users of clinical-trial information, I elaborated a set of design principles so as to answer research Question 2: How can we improve current Web interfaces for seeking clinical-trial information? This set of principles is discussed in Chapter 7.

5.3 Confirmatory Stage

The confirmatory stage of my research was intended to confirm the findings of the exploratory stage and the resultant set of user requirements. In this second stage, I designed and tested two visualization tools aimed at supporting the user by reducing the cognitive

effort involved in seeking clinical-trial information.

5.3.1 Iterative Tool Design

I created several visualization prototypes to confirm my understanding of the needs and tasks of users of clinical-trial data. I experimented with parallel coordinates, barcharts, scatterplots, tag clouds, tables, multiple views, small multiples, and dynamic queries. The design process involved a series of prototypes with varying degrees of functionality, ranging from static drawings to Java and Flex implementations. These prototypes were examined and refined in collaboration with my two expert users. Two of the prototypes were ultimately selected to be developed by a team of programmers associated with the National Center for Biomedical Ontologies (NCBO), resulting in the tools I call CTSearch and CTeXplorer; these are discussed in Chapters 8 and 9.

5.3.2 User Studies

A preliminary evaluation of the tools CTSearch and CTeXplorer was undertaken in user studies with target users and realistic tasks. The tasks were designed to reflect the information needs identified in the exploratory stage of the research. The results of these studies were then used to refine the user requirements and the design of the tools.

5.4 Summary

Briefly recapitulating, in the exploratory stage of the research for this dissertation, three studies were designed to gain an understanding of the unique information needs of users searching for clinical-trial information on the Web, thereby answering questions **Q1.a** through **Q1.f** (cf. Table 5.1). The discussions with experts provided an initial understanding of the information needs, which was then confirmed and refined by means of an Internet questionnaire and a query-log analysis. In the confirmatory stage of the research, visual-

ization tools were developed and tested to validate the users' needs and requirements. The exploratory studies briefly described here are discussed in more detail in Chapter 6 and synthesized in Chapter 7; the confirmatory studies are discussed in Chapters 8 and 9.

Chapter 6

Information Needs of the Clinical Trials User

In this chapter, I investigate the needs of users seeking clinical-trial information on the Web (**Q1**). I describe three studies conducted in the exploratory stage of my research: a series of moderated discussions with expert users of clinical-trial data, a survey of experts and lay users of such data, and a query-log analysis of a leading database of clinical trials. The data collection methods and the results of this research are presented in the following sections.

6.1 Moderated Discussions with Experts

The first of the three exploratory studies took the form of a series of discussions with two clinical-trial experts located at the University of California (San Francisco). The first expert was a senior researcher on clinical-trial informatics with 14 years of experience in the field; the second worked for seven years in clinical-trial data management and has been working for the past four years in clinical research informatics. I moderated these discussions. Their purpose was to investigate the tasks involved in seeking clinical-trial information, to analyze the processes involved in carrying out those tasks using current technology, to identify the problems expert users encounter with current technology, and to establish a set of requirements for an improved human-computer interface for exploring clinical-trial databases.

The discussions were conducted as follows. The two expert users met at the University of California (San Francisco) and connected via a teleconference call with myself and a senior HCI researcher and designer (my supervisor), who met at the University of Victoria. The discussions occurred monthly from May, 2007, to November, 2008. I conducted them following a pre-arranged agenda established by consensus of the participants. The discussions proceeded in a democratic manner, in which decisions were always taken by consensus. As well as moderating the discussions, I took notes and often created minutes to keep track of the design decisions and concerns expressed by the participants.

The discussions centred on identifying the needs of medical researchers seeking clinical-trial information. They followed a semi-structured format, in which I prepared a few questions in advance and then asked more questions as the discussion progressed. During these discussions, I acquired the background needed to identify the information needs of the users and to propose solutions to the problems they typically encountered in seeking to satisfy those needs. We documented the identified needs as a set of tasks and use cases, which are described later in this section and in Chapter 7.

I designed prototype interfaces for a subset of these use cases, which we identified as high-priority. Those prototypes were evaluated by the experts to guarantee that I was properly interpreting their needs.

The moderated discussions with expert users of clinical-trial data, served as the requirement-gathering stage of my research. To aid the reader's understanding of this important process, I provide an example of the agenda and minutes from one of these discussions in Appendix A, which occurred during a teleconference meeting in August, 2007. During those meetings we were exploring various ideas and proposals, some of which came to fruition while others did not. The discussions were a valuable learning experience for all involved. They provided me with the background I needed to understand the problem I was investigating. They also helped the experts to clarify their initially vague conceptions of their information needs and the features of the software tools that would satisfy those needs.

As expected, it was challenging at the beginning of the project for medical experts and

computer experts to arrive at a shared understanding of the problem under investigation, because of the different terminologies used: *medical terminology* versus *computing terminology*. Additionally, the user's mental model of a computational problem can differ from that of the designer of a user interface. For example, during my first interactions with the expert users, they provided me with some drawings they had made to represent an overview of what they thought an effective interface might look like. These drawings used fictitious data and resembled a *parallel-coordinates* model; based on this, I selected a parallel-coordinates visualization technique [Ins85]. Only when we began to work with realistic data did it become apparent that such a technique was inappropriate.

6.1.1 Gathering Requirements

Allen [All03] points out that the process of software development in a research context differs from that followed in an industrial context. In an industrial setting, a software application is developed after a rigorous study to determine that it is needed and that the investment of time and money required to develop it is justifiable. By way of contrast, in a research setting, such as that in which the work reported in this dissertation was undertaken, software tools are created to explore ideas, to identify requirements, and to detect design problems.

My research followed an iterative and incremental model, in which prototype tools went through several progressive stages of development. I created these prototypes to exemplify and refine my proposed solutions and to assess my understanding of the users' needs; thus, the prototypes were used to communicate with the users. At each stage, the expert users evaluated my proposals and we determined their potential usefulness. This exploratory stage, known in the software-engineering literature as that of requirements gathering, is the initial stage in the process of developing a software tool.

Nuseibeh and Easterbrook [NE00] analyze the requirements-gathering stage in terms of its four component phases: objective setting, acquisition of background knowledge, identification of stakeholders, and collection of stakeholder requirements. They define these

phases as follows:

Objective setting: in this phase the high-level goals of the intended system are defined.

Acquisition of background knowledge: in this phase, the tool designer acquires background on the problem domain, on the problem itself, and on the needs the proposed software is intended to satisfy.

Identification of stakeholders: in this stage the stakeholders (e.g., clients and targeted users) are identified and their goals are prioritized according to importance.

Collection of stakeholder requirements: In this stage the tasks to be accomplished by the proposed software are identified.

Next, I discuss my findings from the moderated discussion in terms of those four components.

6.1.2 Findings: User Roles and Tasks

Objective setting: The goal of the targeted system is to enable the targeted users, that is, medical professionals, to find and interactively compare a set of clinical trials relevant to their questions or needs.

Background knowledge acquisition: From the discussions with experts, I became familiar with the use of the technical medical vocabulary used in clinical-trial literature (see the definitions provided in Chapter 2). I learned that medical investigators search and study multiple clinical trials related to a topic of interest, with particular attention to the trials' designs and results. Comparing similarities and differences across trials is part of the process of understanding and critically appraising the relevant trials. The researchers extract basic data, such as eligibility criteria, interventions, and outcomes, and use the data to create tables in word processors or spreadsheets. They manually draw charts or make annotations on the tables to compare the trial information.

Knowledge organization: From the moderated discussions, we identified three types of expert users of clinical trials: the *clinician*, the *trial designer*, and the *systematic reviewer*. The first is a practitioner of medicine who provides clinical care to patients, the second is a laboratory researcher who designs new medical experiments, and the third is a reviewer who appraises the quality of clinical trials and summarizes past research.

Stakeholder requirements collection: The general tasks in which each type of user engages were described by the two experts as follows:

Clinician: The clinician needs support to broaden his or her knowledge of an illness and to do so as efficiently as possible. She is looking for the symptoms associated to a disease. She is looking for safe, effective, evidence-based interventions (e.g., drug, dosage, or timing) to use in her practice. The clinician wants to know the characteristics of the sample population that participated in the experiments, as well as of the control and intervention groups into which the sample was divided. On some occasions, she has a patient who might be a potential participant in a clinical trial and seeks information about recruitment for upcoming trials or trials in progress.

Trial designer: A trial designer is interested in testing a drug for a specific disease or condition. She wants to investigate previous tests of this drug or similar drugs, whether used alone or in combination with others. She also wants to know the characteristics of the population that was tested and in which country or continent the tests were conducted. The designer is looking for gaps in previous testing (e.g., combinations of drugs that have not been tested). This information allows the designer to design a new experiment, perhaps picking a different combination of drugs or re-running a previously tested drug combination with a different population. The trial designer is also interested in understanding the risks involved in testing these drugs, as identified in previous trials. This

Table 6.1. *Tasks of Clinical Trials Professional Users*

	Clinician	Trial Designer	Systematic Reviewer
To broaden knowledge of a medical condition for clinical care	X		
To find the drug or treatment of choice to treat a medical condition	X		
To find symptoms associated with a medical condition	X		
To find adverse outputs of a given drug	X	X	
To find trials for a potential participant	X		
To find trials for a systematic review			X
To see how specific drugs have been tested	X	X	X
To design a new trial		X	

information helps her to ensure that the proposed trial will have an acceptable level of risk.

Systematic reviewer: The systematic reviewer needs support to perform critical appraisals of clinical trials and to analyze their similarities and differences, with a view to identifying trials that are similar enough to pool them together and apply statistical inferences. In addition to understanding characteristics of the population, interventions, and outcomes, the systematic reviewer needs to compare and understand the methodological features of the experiments.

The discussions with the two expert users of clinical-trial data established that existing computer interfaces do not provide efficient support for the tasks they wish to accomplish. In particular, they expressed a need for an interface that would enable them to quickly compare sets of clinical trials in terms of specific variables, and they wished this information to be displayed in a compact and intuitive graphical representation. From the three types of medical professionals who use clinical trials, the clinician, the trial designer, and the systematic reviewer; I coded and summarized their general tasks into eight typical tasks in which one or more of these users seek to extract a specific data set from the clinical trials and to compare these data across trials. The codes were *broaden knowledge*, *treatment of choice*, *experiment features*, *potential participants*, *design a trial*, *outputs*, and *systematic review*. A list of the summarized tasks can be seen in Table 6.1. Those eight tasks were validated by the two experts through email iterations.

6.2 Online Questionnaire Study

As mentioned in Chapter 5, I conducted an online questionnaire of a larger sample of users of clinical-trial data to complement the testimony of the experts I met with. Thus, the questionnaire was targeted at medical professionals and it was worded using the vocabulary I acquired from the moderated meetings. The survey questions concerned the following categories: (1) characteristics of the population, (2) frequency to search for clinical-trial information, (3) the type of information sought, (4) the tools used for searching clinical-trial information, (5) user satisfaction when searching for clinical-trial information, and (6) unmet information needs. The actual list of questions asked and corresponding types are shown in Table 6.2. Four questions were multiple choice, three questions were open-ended, and three questions were combined.[1] Before implementing the questionnaire, I pilot tested the instrument with the two experts and refined the questions as per their feedback. The average estimated time to respond to the questionnaire was 15 minutes. The survey was hosted at SurveyMonkey[2] and was available from August to November 2008. Respondents were recruited via webmasters of research institutions and universities involved in clinical trial research identified by searching the Internet. A total of 600 invitations were sent out and a total of fifty-three responses were received. Appendix B provides an example of the invitation to participate and Appendix C provides the consent form. Responses from a broad population were received between 1 August and 6 November, 2008.

6.2.1 Characteristics of the Population

Questions one and two from Table 6.2 were used to identify the characteristics of the population. Question one offered the following options: Clinician/physician, Clinical trial designer, Systematic reviewer, Medical student and Patient (or family of a patient). Users were asked to specify a different role if needed. Four additional roles emerged from the participants' responses: clinical trial coordinator/manager, pharmacist, nurse and medical

[1] A combined question is a multiple choice question with an additional field to enter an open-ended response.

[2] http://www.surveymonkey.com/

Table 6.2. *Online Questionnaire*

Question	Type	Category
1. Choose the role that best applies to you: a) Trial designer b) Systematic reviewer c) Clinician/phycisian d) Medical Student e) Patient or Family of a patient f) Other (please, describe)	combined	Characteristics of the population
2. How long have you been in this role? a) 2 years or less b) 3 to 5 years c) 6-9 years d)10 years or more	close-ended	Characteristics of the population
3a. For the scenarios that apply, how frequently do you search in this scenario? a) To broaden knowledge of a medical condition for clinical care b) To find the drug or treatment of choice to treat a medical condition c) To find symptoms associated with a medical condition d) To find adverse outputs of a given drug e) To find trials for a potential participant f) To find trials for a systematic review g) To see how specific drugs have been tested h) To design a new trial i) Other scenarios or tasks not listed here	close-ended	Reasons to search for clinical-trial information
3b. For the scenarios that apply, which resource is your first choice while searching in this scenario? (If you use other Internet resources not listed here, please specify) a) To broaden knowledge of a medical condition for clinical care ... i) Other scenarios or tasks not listed here	close-ended	Tools used for searching clinical-trial information
3c. Have you been successful in searching in this scenario? a) To broaden knowledge of a medical condition for clinical care ... i) Other scenarios or tasks not listed here	close-ended	User satisfaction when searching for clinical-trial information
4. What kind of fields do you usually use when searching for clinical trials? Other (please specify) a) Conditions b) Interventions c) Outcomes d) Eligibility criteria e) Results f) Country of recruitment	combined	Type of information sought
5. Are there any additional scenarios you would like to search using information related to clinical trials?	open-ended	Unmet Information Needs
6. Please provide a sample of a search that cannot perform easily with Internet resources?	open-ended	Unmet Information Needs
7. What additional information or features would you like to see in the available Internet resources?	open-ended	Unmet Information Needs
8. Do you find it frustrating to search for clinical trials information via the Internet? Please specify why	combined	User satisfaction when searching for clinical-trial information

Table 6.3. *Years of Experience of participants*

	Response count
2 years or less	6
3-5 years	11
6-9 years	9
10 or more years	26
	52

librarian. The distribution of participants is as follows: 17 clinicians/physicians, 12 clinical trial coordinators/managers, 10 patients or family members of a patient, five clinical trial designers, four nurses, two medical librarians, two pharmacists, and one systematic reviewer. Forty-three of the participants were medical professionals, while 10 were patients or family members of patients. The responses included an unexpectedly large number of lay users (patients or family of a patient) since the questionnaire was targeted to medical professionals.

Question two asked the participants how long they had been in the corresponding role as users of clinical-trial data. Most participants reported that they had been in their respective role for more than three years, while only six had been in their role for two years or less; one user skipped this question (see Table 6.3).

6.2.2 Findings: Frequency of Search Scenarios

To investigate the frequency that users search for clinical-trial information, participants were asked in Question 3a how frequently they perform each of a list of the predefined tasks. The predefined tasks were taken from the findings of the moderated meetings and listed in Table 6.1. The respondents could choose for each task the frequency of one or two times a month, three to nine times a month, and 10 or more times a month.

The respondents sought clinical-trial information with different frequencies. However, the large majorities of both *patients or relatives of patients* and *Medical professionals* most frequent reason to search for clinical trials was “to broaden knowledge of a medical condition for clinical care”. The second most frequent reason cited by the patients and family

Table 6.4. *Frequency of Search Scenarios for Patients or Family of Patients*

	1 or 2 times a month	3 to 9 times a month	10 or more a month	Response count
1. To broaden knowledge of a medical condition for clinical care	**4/10**	3/10	3/10	**10/10**
2. To find the drug or treatment of choice to treat a medical condition	**5/10**	3/10	0	8/10
3. To find symptoms associated with a medical condition	**3/10**	**3/10**	1/10	7/10
4. To find adverse outputs of a given drug	**5/10**	2/10	0	7/10
5. To find trials for a potential participant	**7/10**	2/10	0	9/10
6. To find trials for a systematic review	**5/10**	1/10	0	6/10
7. To see how specific drugs have been tested	**4/10**	3/10	0	7/10
8. To design a new trial	**4/10**	0	0	4/10
9. Other scenarios or tasks	**3/10**	0	2/10	5/10

Table 6.5. *Frequency of Search Scenarios for Medical Professionals*

	1 or 2 times a month	3 to 9 times a month	10 or more a month	Response count
1. To broaden knowledge of a medical condition for clinical care	13/43	**18/43**	10/43	**41/43**
2. To find the drug or treatment of choice to treat a medical condition	**21/43**	19/43	1/43	**41/43**
3. To find symptoms associated with a medical condition	**21/43**	16/43	2/43	39/43
4. To find adverse outputs of a given drug	**23/43**	12/43	1/43	36/43
5. To find trials for a potential participant	**19/43**	8/43	5/43	32/43
6. To find trials for a systematic review	**17/43**	10/43	3/43	30/43
7. To see how specific drugs have been tested	**21/43**	7/43	1/43	29/43
8. To design a new trial	**20/43**	2/43	1/43	23/43
9. Other scenarios or tasks	4/43	**7/43**	4/43	15/43

members was “to find trials for a potential participant”, while that of the professionals was “to find the drug or treatment of choice to treat a medical condition”.

Tables 6.4 and 6.5 detail the frequency with which users conduct searches for each of the tasks cited above. As indicated in the tables in bold font, the most common frequency across all tasks is one to two times per month. Few tasks occur more than 10 times per month.

6.2.3 Findings: Data Components Used

The questionnaire included a section (cf. Question 4 in Table 6.2) on data components used in seeking clinical-trial information. Respondents were provided with a list of six data com-

ponents and asked to check the ones they used: eligibility criteria, outcomes, interventions, conditions, country of recruitment, and results. The predefined list of data components was extracted from the findings of the moderated meetings. The most frequently used categories, and the only ones cited by both professional and lay users, were "conditions" and "eligibility criteria" (cf. Table 6.6). Note that whereas a majority of medical professionals searched for "interventions", patients or relatives of patients never did this.

6.2.4 Findings: Tools for Searching Clinical Trials

Participants were asked which tool was their first choice when searching for clinical-trial information (cf. Question 3b in Table 6.2). Respondents were provided with a list of seven options to choose from: Google, PubMed,[3] ClinicalTrials.gov,[4] journal database, Cochrane Collaboration,[5] and Current Controlled Trials (ISRCTN).[6] The results indicate that a variety of tools or search engines are used to seek clinical-trial information. The preferred tool depends on the task; Tables 6.7 and 6.8 provide a detailed description of tools and user preferences. Google and PubMed were most frequently used by both medical professionals and lay users when they wished to broaden their knowledge of a medical condition for clinical care. Google was also their first choice i) To find the treatment of choice for a medical condition, ii) To find symptoms associated with a medical condition, and iii) To find adverse outputs of a given drug. PubMed was most frequently used by all respondents to investigate how specific drugs have been tested.

Medical professionals most frequently used ClinicalTrials.gov to find trials for a potential trial participant. ClinicalTrials.gov, PubMed, and Journal database were more frequently used by medical professionals than by patients and family members, which is expected since medical professionals know the importance of using reliable sources of information and might know that not all retrieved results from Google are from trusted

[3] http://www.ncbi.nlm.nih.gov/pubmed/
[4] http://clinicaltrials.gov/
[5] http://www.cochrane.org/
[6] http://www.controlled-trials.com/

Table 6.6. *Fields Used when Searching*

	Patients/ family	Medical professionals
Conditions	**6/10**	**35/43**
Eligibility criteria	**7/10**	**26/43**
Interventions	0	25/43
Outcomes	3/10	18/43
Results	3/10	15/43
Country of recruitment	**6/10**	6/43

Table 6.7. *Tools Used by Patients or their Families when Searching*

	Google	Clinical Trials.gov	PubMed	Journal Database	Other
1. To broaden knowledge of a medical condition for clinical care	**3/10**	1/10	**3/10**	0	2/10
2. To find the drug or treatment of choice to treat a medical condition	**3/10**	0	1/10	1/10	2/10
3. To find symptoms associated with a medical condition	**4/10**	1/10	1/10	0	1/10
4. To find adverse outputs of a given drug	**4/10**	0	1/10	0	2/10
5. To find trials for a potential participant	2/10	3/10	0	1/10	**4/10**
6. To find trials for a systematic review	2/10	0	1/10	0	**3/10**
7. To see how specific drugs have been tested	2/10	1/10	**3/10**	1/10	1/10
8. To design a new trial	1/10	1/10	1/10	1/10	1/10
9. Other scenarios or tasks	1/10	0	1/10	0	**2/10**

sources. Additional tools used by patients or relatives of patients were WebMD,[7] MSN,[8] and Copernic.com.[9] Other tools cited as frequently used by medical professionals were UpToDate,[10] WebMD,[11] and PDR.net.[12] One medical professional commented that she rarely used Google to search for clinical-trial information.

6.2.5 Findings: User Satisfaction

Respondents were asked about their perceived success when seeking clinical-trial information (cf. Question 3c in Table 6.2). Both groups of users were inclined to perceive themselves as successful. They felt very successful when searching for clinical-trial information i) To broaden their knowledge of a medical condition for clinical care, and ii) To

[7] http://www.webmd.com/
[8] http://www.msn.com/
[9] http://mamma.com
[10] http://www.uptodate.com
[11] http://www.webmd.com/
[12] http://pdr.net

Table 6.8. *Tools Used by Medical Professionals when Searching*

	Google	Clinical Trials.gov	PubMed	Journal Database	Other
1. To broaden knowledge of a medical condition for clinical care	**16/43**	1/43	**16/43**	6/43	2/43
2. To find the drug or treatment of choice to treat a medical condition	**14/43**	2/43	12/43	3/43	9/43
3. To find symptoms associated with a medical condition	**21/43**	0	7/43	3/43	7/43
4. To find adverse outputs of a given drug	**18/43**	0	8/43	1/43	10/43
5. To find trials for a potential participant	3/43	**20/43**	2/43	1/43	9/43
6. To find trials for a systematic review	2/43	8/43	8/43	4/43	**9/43**
7. To see how specific drugs have been tested	8/43	6/43	**12/43**	2/43	1/43
8. To design a new trial	2/43	6/43	**7/43**	2/43	**7/43**
9. Other scenarios or tasks	**6/43**	2/43	2/43	2/43	3/43

Table 6.9. *Patients' Perceived Search Success*

	Very successful	Somewhat successful	Neutral	Somewhat unsuccessful	Very unsuccessful
1. To broaden knowledge of a medical condition for clinical care	**4/10**	2/10	3/10	1/10	0
2. To find the drug or treatment of choice to treat a medical condition	2/10	**3/10**	**3/10**	0/10	0/10
3. To find symptoms associated with a medical condition	**3/10**	2/10	2/10	0/10	0/10
4. To find adverse outputs of a given drug	**4/10**	1/10	2/10	0/10	0/10
5. To find trials for a potential participant	0	**4/10**	**4/10**	1/10	0
6. To find trials for a systematic review	1/10	0	**5/10**	0	1/10
7. To see how specific drugs have been tested	2/10	**3/10**	**3/10**	0	0
8. To design a new trial	0	1/10	**3/10**	0	0
9. Other scenarios or tasks	**2/10**	0	**2/10**	0	0

Table 6.10. *Professionals' Perceived Search Success*

	Very successful	Somewhat successful	Neutral	Somewhat unsuccessful	Very unsuccesful
1. To broaden knowledge of a medical condition for clinical care	**21/43**	19/43	1/43	0/43	0/43
2. To find the drug or treatment of choice to treat a medical condition	15/43	**17/43**	5/43	2/43	1/43
3. To find symptoms associated with a medical condition	13/43	**18/43**	7/43	1/43	0/43
4. To find adverse outputs of a given drug	**16/43**	14/43	7/43	0	0
5. To find trials for a potential participant	5/43	**21/43**	6/43	2/43	0
6. To find trials for a systematic review	3/43	**19/43**	7/43	1/43	1/43
7. To see how specific drugs have been tested	5/43	**14/43**	9/43	2/43	0
8. To design a new trial	3/43	**10/43**	9/43	0	3/43
9. Other scenarios or tasks	**7/43**	5/43	2/43	0	1/43

find adverse outputs of a given drug. Interestingly, only one-third of medical professionals felt successful when seeking information for the purpose of designing a new trial (cf. Tables 6.9 and 6.10 for a summary of the answers).

On the other hand, when users were asked about their frustration levels when searching for clinical-trial information (cf. Question 8 in Table 6.2) almost half of medical professionals found it frustrating, whereas only one-fifth of lay users reported experiencing such frustration (cf. Table 6.11).

Additionally, Question 8 had an open-ended section, which asked the users to explain their response to the close-ended section of Question 8. Note that in the following discussion of findings I provide examples only for the participants which granted me permission to quote them. Medical professional responses to the open-ended section included three answers with a positive connotation such as: "ClinicalTrials.gov has served me quite well", "I have learned how to search", and "Internet is a valuable resource [...] if used appropriately and with care"; whereas, fifteen answers had a negative or a request connotation: the difficulty to determine the reliability of a clinical trial (e.g., "It's easy to find them, just hard to tell which ones are good/safe"), the need to emphasize the geographic location of open trials (e.g., "geographic aspects should be emphasized –show trial locations before trial details. This would facilitate use by consumers. Its frustrating to read about a good sounding trial to find out that its in Korea only or located halfway across the USA"), the high frequency of out-of-date information (e.g., "Web sites are not always current with enrollment information"), the large number of results from a query often saturated with irrelevant information (e.g., "Often times the fields of interest populate with many trials of which only a couple are actually applicable."), the problem of finding trials by acronyms (e.g., "Trials are often referred to by acronyms e.g., COMPASS, SWORD yet these are often not recognized by search engines as the acronym is not included in the abstract or title"), and usability issues in current user interfaces (e.g., "there is no easily searchable database").

Lay users to the open-ended section in Question 8 included two responses with a neg-

Table 6.11. *Frustration Level while Searching for Clinical-Trial Information*

	Patients/ family	Medical Professionals
Strongly disagree	1/10	6/42
Somewhat disagree	3/10	9/42
Neutral	**4**/10	7/42
Somewhat agree	2/10	**17/42**
Strongly agree	0	3/42

ative connotation and one with a positive connotation ("I have always used the internet to research my disease for anything I need to know about it. And have been successful in finding what I need to know")

6.2.6 Findings: Unmet Information Needs

To identify the unmet information needs of users of clinical trials, three open-ended questions were asked (cf. Questions 5, 6, and 7 in Table 6.2). The responses were interpreted and coded [Cre04] as in the following example. A participant submitted a search for "Treatment of lung cancer after failure on prior chemotherapy". This search was interpreted as a complex query because it was including information from more than one facet (or category), summarized as search for "intervention X [AND] condition Y [AND] previous intervention Z", and coded as a "complex query".

Respondents were asked what other tasks they wished to perform that were not included in the questionnaire (Question 5 in Table 6.2). Medical professionals' answers varied and included: searching post-marketing studies, trial design, consent forms, sponsor, city, institution, ethnicity, all diseases associated to a medical speciality (e.g., "New/Impending Trials Anesthesiology Related"), phase, eligibility criteria, rare medical conditions (e.g., "a source for 'grouped' rare or orphan medical conditions-versus just going to all Internet sources and taking so much time."), biomarkers,[13] contract research organization (CRO),[14] epidemiology, and disorders related to a specific physiological function (e.g., "ability to

[13] Biomarkers are measurable substances contained in blood and other body fluids, which are used to indicate the state of a medical condition.

[14] CRO is a organization contracted to conduct clinical trials with a sponsor

search for say, all related vision disorders without having to specify ”macular degeneration” and ”diabetic retinopathy” (i.e., have hierarchical search capabilities as found in PubMed)”). Lay users answers to this question included searching for future availability of trials in new locations and success rates of specified therapies.

Respondents were also asked to give examples of queries that could not be performed easily with currently available Internet resources (Question 6 in Table 6.2). Medical professionals indicated that it was difficult to find treatment information related to multiple conditions that occur simultaneously and/or to patients previously treated with more than one treatment (e.g. “Breast cancer, recurrent, ER/PR positive, HER-2/neu positive, failed 1st line hormonal treatment and previously treated with Herceptin”). Other examples of queries that cannot readily be performed with existing technology included seeking information regarding rare medical conditions; generating a list of drugs and outcome studies; seeking information on prevention strategies for a given condition; information on impending, unregistered, ongoing, or closed trials (e.g. number of participants, response rate, consent forms, contact information, unusual adverse outcomes, results, etc); and the level of commitment required from a participant (time per visit, number of visits, etc.). The heterogeneity within the data in the description of trials was also indicated as a difficulty in performing queries.

Participants were asked for additional information or features that they would like to see in the available Internet resources (Question 7 in Table 6.2). Medical professionals expressed the need for finding protocols that fit specific guidelines, search trials for radius of miles from within a given geographical location, and a centralized and open repository of trials easy to access. Lay users expressed the desire to have pharmaceutical information associated with a drug tested.

With the responses of Questions 5, 6, and 7 I refined the reasons users have to search for clinical trials as follows:

- To broaden knowledge of a medical condition for clinical care

- To find the drug or treatment of choice to treat a medical condition
- To find trials testing diseases associated to a medical specialty
- To find trials testing diseases associated to a specific physiological function
- To find symptoms associated with a medical condition
- To find adverse outputs of a given drug
- To find trials for a potential participant
- To find trials for a systematic review
- To see how specific drugs have been tested
- To design a new trial
- To find biomarkers

6.3 Query-Log Analysis

The interviews and survey provided an understanding of the ways in which users perceive their own needs and intentions while seeking clinical-trial information. However, I also needed objective data on their actual behaviour during the information-seeking process. Thus, I undertook a query-log analysis to obtain such data. I had one goal for the log analysis. I wanted to know what search strategies they use to find the information they seek.

The definitions adopted here are the following: a *session* is defined as all consecutive queries from the same user without time restriction, a *query* is list of terms submitted to the search engine by the user, and a *term* is defined as a string of characters separated from others by punctuation or spaces.

6.3.1 Findings: Search Strategies

The query log consisted of the usage recorded for a portion of a day in PubMed in 2005 and contained three fields, namely, an encrypted code associated with the user, a timestamp, and the actual query. The query log contained 500,000 queries in total. The query log was provided by Hersovic as a sample of the query he used in his query log analysis of PubMed [HTHB07].

The log analysis revealed that users were submitting sequences of similar queries. Those similar queries were most probably refinements of the initial query. To explore this possibility, a sample of user sessions was extracted. I wanted not only to explore how many times users refined a query, but also, and more importantly, what strategies and patterns they used in the refinement. I questioned whether they added new words to the query and, if so, how many words were added, whether they deleted words from the query, and characterize the nature of the refinements. A sample of 100 user sessions was randomly selected from the PubMed log file. These sessions involved users who executed only a single query along with those who executed several. The sample was coded [Cre04] by a second researcher aside from myself. We were looking for consistent user behaviour to classify the type of refinement the user carried out. Examples of codes were such as *misspelling*, *stop-words*, *pivotal terms*, and *added author*.

Table 6.12. *Number of Refinements per User Session*

Num. of Refinements	0	1	2	3	4	5	6	7	8	9	10	11	12
Frequency	20	7	13	20	7	3	6	2	0	3	3	2	1
Percentage (%)	23	8	14.9	23	8	3.4	6.9	2.3	0	3.4	3.4	2.3	1.1

Table 6.13. *Number of Terms per Query*

Terms	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Frequency	2	73	137	82	37	24	10	5	0	1	2	4	1	3	1	0
Percentage (%)	.5	19.1	35.9	21.5	9.7	6.3	2.6	1.3	0	.3	.5	1	.3	.8	.3	1

Users most frequently performed two or three refinements, as seen in Table 6.12. The average number of query refinements carried out was 3.18. The average query length was 2.86, this is consistent with the results of Herskovic *et al.* [HTHB07] (see Table 6.13).

Table 6.14. *Sample user session*

Timestamp	Query
12906	PMR
12976	Progressive Muskelrelaxation
13011	Progressive Muscle Relaxation
13142	PMR, Jacobsen

The strategies identified in the query-log analysis for search refinement were as follows: Users performed two or three refinements per session; retained a single, pivotal term throughout most of their refinements; and frequently switched between subject searches and author searches (i.e. they queried the name of one or more authors); made spelling corrections; used commas or quotation marks to distinguish concepts; and frequently included "stop words"[15].

In the example session in Table 6.14, the user begins with an acronym "PMR", then expands the acronym but misspells the term "Muscle Relaxation", then corrects the misspelling, and finally returns to the acronym and adds the author name "Jacobson". In another user session, the user used commas in the submitted query *cord, stem cell, bone regeneration* to separate terms. PubMed, however, does not recognize commas as a mechanism for delimiting multi-word terms. It appears that the user intended "stem cell" and "bone regeneration" as single concepts. In another session, the user submits the query *"eeg" "children" "working memory"* using quotations to separate the terms.

In the example session in Table 6.15, the user's query refinements begin by pivoting on the term "prostate cancer"; then show an interest in the author "Thor Straten", a researcher in cancer and vaccination and other therapies; and then returns to pivot on the initial query. It appears that after examining publications of "T. Straten" the user has now a wider vocabulary to try in combination with "prostate cancer".

[15] Stop words are words frequently used, such as "for","of","in","the", etc., which are typically removed by the search engine. One example of the use of stop words is the query: "journal of the american college of cardiology the adult white congenital heart disease" with the stop words "of" and "the".

Table 6.15. *Sample User Session*

Timestamp	Query
7421	prostate cancer and immunotherapy
7434	prostate cancer and immunotherapy and clinical trials
7523	prostate cancer and immunotherapy and clinical trials and t cells
7784	thor straten
9836	immunotherapy and prostate cancer and clinical trials
16113	peptide and prostate cancer and clinical trials
16229	peptide vaccination and prostate cancer and clinical trials

6.4 Limitations of the exploratory studies

The research studies conducted in this dissertation have strengths and weaknesses. The results of the moderated discussions might be challenged on the grounds of the limited number of experts participating. I relied on the expertise of only two persons. However, our interactions continued over a considerable period (about 18 months) and included email correspondence in addition to the monthly meetings. Moreover, I consider that the quality of the experts compensated for their small quantity; their experience with the problems of clinical-trial registration and knowledge management was extensive, and the analysis we performed together was thoughtful and detailed.

The online questionnaire was targeted at medical professionals working in the clinical-trial domain, as this is my primary user group. Thus, the language in which the questions were formulated assumed a degree of familiarity with the kinds of tasks that professionals would perform. Unexpectedly, we received a large number of responses from patients and members of patients' families. This group of respondents may have misinterpreted some of the questions. For example, four patients or relatives of patients declared that they sought clinical-trial information "to design a new trial", a purpose that is clearly beyond the capacity of this category of users (see Table 6.4). This makes one suspect they may have interpreted the wording of the tasks in a different context than was intended. Nevertheless, the findings provide some preliminary insight into the information needs of lay users who are looking for clinical-trial information to broaden knowledge of a medical condition for clinical care and to find trials for a potential participant. Future studies are needed to explore their needs further.

Finally, the query-log analysis only looked at a subset of 100 user sessions. I encountered a significant degree of saturation in this small subset during the coding process. However, by examining a larger set of user sessions, I may be able to observe other search-refinement patterns. I attempted to mitigate the disadvantages of each instrument by triangulating the results of all three, that is, of the moderated discussions, the questionnaire, and the query-log analysis. The results appear consistent across all of these sources.

6.5 Discussion of findings

Searching the Internet for clinical-trial information and understanding the relationship between the retrieved documents is an important, but cognitively complex process for users. Many of these users wish to improve their understanding of a disease or to learn the effects of a treatment. Others seek information on optimal dosage of a drug, on the advantages of a given preventive measure, on which treatments need further testing, on the effectiveness of a treatment on different populations, and so on.

The results from the three exploratory studies help to answer research Question 1: What are the unique information needs of users searching for clinical-trial information on the Web? The moderated meetings with the two experts provided an initial understanding of those needs, which was confirmed and refined by the on-line questionnaire and the query-log analysis. The main finding of the interviews was that searching and comparing clinical trials is difficult and time-consuming. The experts identified a series of tasks, such as broadening one's knowledge of a medical condition or identifying a preferred treatment, that are typically carried out by three categories of professional users of clinical-trial information: systematic reviewers, trial designers, and clinicians (see Table 6.1). The on-line questionnaire provides results for a broader population consisting of clinicians, patients, trial coordinators, nurses, and so on. It reinforces the findings of the moderated discussions with regard to some of the typical tasks performed, while also identifying some additional tasks. The third study, the query-log analysis, complements the first and second

by providing objective data about the search strategies employed by users of clinical-trial information.

The three studies focus on specific data components of clinical-trial information, such as interventions, conditions, outcomes, results, and characteristics of the population being studied. The findings of the interviews and, to a lesser extent, the on-line questionnaire suggest that users have difficulty satisfying their information needs, especially when their queries involve more that one data component, for example, "How [intervention X] works on [condition or disease Y]?" and the answer is contained in multiple documents.

The three exploratory studies provide much insight into the needs of users of clinical-trial information, the tasks they perform, the data components involved in these tasks, and the difficulties they encounter using existing technology. With this knowledge in hand, we are ready to move on to propose enhancements to current information-seeking interfaces that would support the needs and tasks of these users.

Chapter 7

Towards Improving the Seeking Experience of Users of Clinical Trials

In my investigations I found that seeking clinical-trial information and making sense of the corresponding findings is a challenging activity. Usually, users start with a vague idea of how to find useful information and formulate an initial query, which, in most cases, they proceed to refine as they better understand the vocabulary used in the clinical-trials domain. Typically, results appear as a large, text-based list of documents. Unfortunately, due to the large number of relevant documents that may satisfy the query, it may be difficult for users to gain an overall understanding of the results. Furthermore, to find the answers to their questions, users spend considerable time aggregating selected documents and comparing trials across different data components (facets), which is a complex task on account of the heterogeneity of values and metrics within the retrieved lists of trials.

In this chapter, I undertake research Question 2: *How can we improve current Web-search interfaces for clinical-trial information?* I start by proposing a new model of information seeking, which borrows some features of previous models discussed in Chapter 3 but expands upon them to incorporate important aspects of the information-seeking process that they omit. Then I discuss the challenges faced by users seeking clinical-trial information. Finally, I define a set of design principles and use cases to design more effective user interfaces to support information seeking in the clinical-trial domain.

7.1 Integrated Information-Seeking Model

Having analyzed various information-seeking models alongside the information-foraging theory in Chapter 3, I now present a new model, the Integrated Information-Seeking Model (cf. Figure 7.1), that builds upon that theoretical basis while integrating the findings of my exploratory studies, which are discussed in Chapter 6. In contrast to the models that use either browsing or searching exclusively, my model allows users to meet their information needs by searching, browsing, and by iterating between the two [BYRN99]. In addition, the integrated model considers the process of amalgamating the result set into a single corpus of documents to meet the larger users' need–not that of finding relevant information, but of answering the questions that brought them to seek for information in the first place.

To search, users must recall the terms to submit a query [SBC98, Mar06, BYRN99]. From here, they can formulate the query and submit it to a selected search engine (the choice of such engine might change at any iteration of the process). As a result of submitting a query, users receive a collection of retrieved documents, which they have to evaluate in order to determine what portions of the retrieved results are useful for their information needs. Thus, they extract relevant information from the retrieved results. They can highlight and add comments to printed documents, or copy and paste pieces of relevant information into a new electronic document. Users can either reformulate the current query by using new terms, or they can initiate a new query based on an evolution of the information need [Bat89].

In contrast users can browse by following an appealing information scent, navigating to wherever the scent leads, and proceeding to extract pieces of relevant information [Pir07]. They can alternate between browsing and extracting relevant information from the retrieved results. Besides iterating on searching and browsing, users aggregate their findings to compare and contrast the results, and to satisfy their information needs.

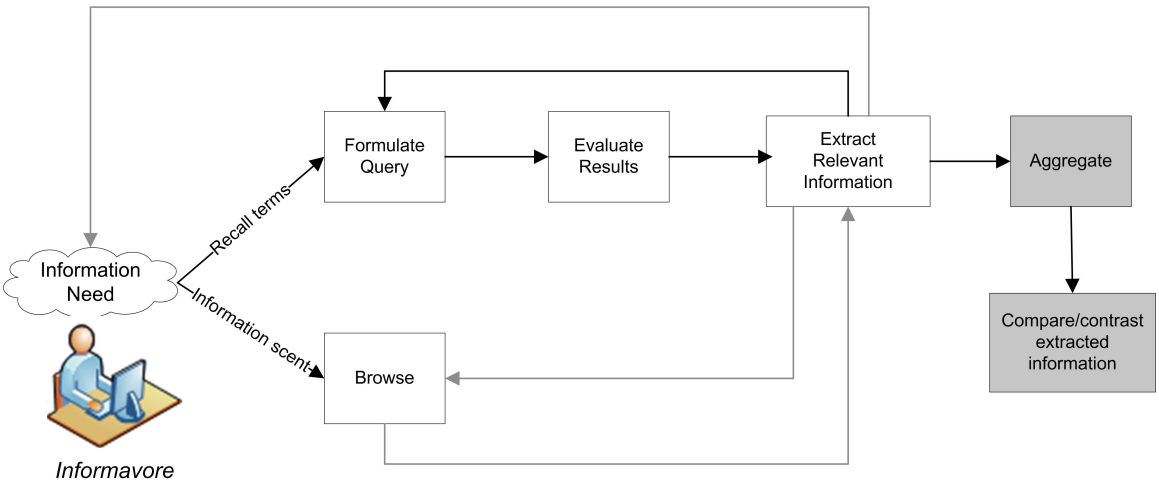

Figure 7.1. *Integrated Information-Seeking Model. Additional elements to previous models are identified in grey.*

7.2 Information-Seeking Challenges

Information seeking in the clinical-trial domain has specific challenges from the human-computer interaction (HCI) point of view, because of the cognitive load imposed on the user. In this section, I identify those challenges based on my literature review (cf. Chapters 3 and 4) and the findings of my exploratory studies (cf. Chapter 6). I discuss the challenges in terms of my Integrated Information-Seeking Model.

Reformulation : This step takes place after the initial query is submitted and the corresponding results have been browsed and evaluated (cf. A in Figure 7.2). Users iterate and reformulate their queries until they are satisfied with the results. In each iteration they add, remove, or change terms as they acquire and assimilate more vocabulary from the retrieved trial documents during the process. They employ different strategies to refine their queries, such as pivoting and changing the type of search, for example, from a subject search to an author search. The typical problem is that to iterate, users must choose *new* keywords to use in a subsequent query, which would lead to better results. When they are not familiar with the underlying vocabulary, it is difficult for them to decide which terms to try. Frequently, terms have multiple

synonyms, and medical terms are hard to spell and remember. Additionally, users may lose the context of their original search when reformulating a query.

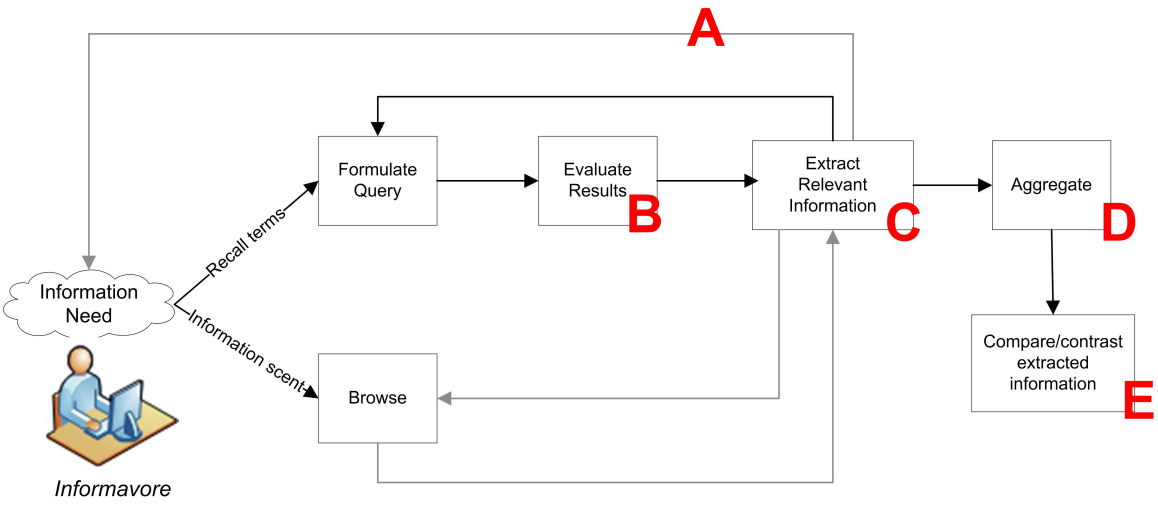

Figure 7.2. *Integrated Information Seeking Model*

Evaluate results: When users receive a list of results, a difficult task lies ahead of them. They have to review manually each promising entry and decide which are useful for their purposes (cf. B in Figure 7.2). Usually, the problem at this step is that there is not enough information to clearly distinguish which results are useful. Results are typically listed in a static representation ordered according to unrevealed variables.

Extract relevant information: Users of clinical trials are looking for specific data within the trials. For example, they are interested in medical conditions studied, treatments tested, or results obtained. Thus, they extract specific information from the trials based on their particular interests (cf. C in Figure 7.2). This may be 1) text-based information, either in paragraphs or in the form of lists of medical terms; 2) information regarding continuous variables, such as the age of participants; 3) categorical information, such as the names of drugs tested; 4) time-dependent information, such as outcomes measured at specific points of time; and 5) information about important dates, such as the start and end dates for recruiting participants. The problem at this

stage is that users have to read each trial to find and extract the data elements sought.

Aggregate: Aggregating information takes place after relevant information from the retrieved results has been extracted (cf. D in Figure 7.2). Such aggregation takes the form of artifacts, such as new documents, annotations, or tables of printed trials. These artifacts are organized by categories or data components, such as eligibility criteria, country of recruitment, and interventions. The problem at this stage is the heterogeneity of trial designs and results. Thus, different authors use different variables and metrics in their experiments, all of which may or may not appear in their published reports. For example, since the requirements to participate in a trial are defined according to the discretion of the trial designer, the number of variables used as eligibility criteria can vary considerably between trials.

Compare/contrast selected results: To find the answers to their questions, users frequently need to combine information within the retrieved documents (cf. E in Figure 7.2). To this end, they have to find the relationship between categories of data. The problem at this stage is again the *heterogeneity* of clinical trials. For example, to compare and contrast the interventions used in a set of related trials can be difficult, because different brand names or generic terms may be used to refer to the same drug; to exacerbate the problem, some trials do not reveal the name of the drug tested but instead use a pseudonym, such as "drug CP-675,206".

7.3 Design Principles

To improve the user experience in seeking clinical-trial information, I propose to extend the typical user interface with a search box to include the following requirements (cf. Table 7.1):

1. Query reformulation: Additional terms related to the submitted query should be provided to users to support formulation of subsequent queries. Feedback should be

Table 7.1. *Requirements for Designing Interfaces for Clinical Trials Users*

	Literature review	Interviews w/Experts	Online Questionnaire	Query Log
Query reformulation	X			X
Clustering of results	X	X		
Clustering comparison		X	X	X
Summarization of results	X	X		
Dynamic queries	X	X		
Multiple selection of terms		X	X	X

provided to the user regarding the context and history of the queries submitted.

2. Clustering of results: The retrieved results should be grouped by predefined categories or facets, such as conditions, interventions, and outcomes. Users should be able to take in the clustered results at a single glance.

3. Linking clusters: The relationship between subsets of results clustered according to different categories or facets should be represented when needed. Users should be able to select and deselect terms to explore such relationships.

4. Summarization of results: The information should be presented in a condensed form to fit an average computer display. Multiple levels of granularity should be supported. The initial summaries should provide an overview of the result set, while further summaries should narrow its scope and provide more detailed information.

5. Dynamic queries: A browsing mechanism should be supported. Users should be provided with interactive features, such as check boxes or sliders, to enable them to explore the data in a dynamic way.

6. Selection (and de-selection) of multiple terms: Users need to see the relationships among various terms (which may correspond to multiple categories) that might be selected in reformulating queries.

7.4 Use cases

During the moderated discussions and in collaboration with the two experts, we derived ten use cases, the use cases were first drafted by one of the experts and I described them using Larman's [Lar01] notation. I present the use cases below in terms of a name, a brief summary of the user's purpose and the output of the target system in response to the user's action, the types of users involved, and the priority. The order of priority or importance is from "primary"(the highest) through "secondary" to "optional" (the lowest). The use cases are summarized in Table 7.2.

1. **Use Case:** Search for related trials.

   Summary: The user wants to find a collection of trials related to a specific question. The system displays an overview of the retrieved collection of trials in a form capable of being displayed on a laptop computer.

   Actors: Medical professionals, patient/family

   Priority level: Primary

2. **Use Case:** Identify all treatments tested for a given medical condition.

   Summary: The user wants to identify all the treatments tested for one or more medical conditions of interest. The systems displays in a compacted format the corresponding treatments. The user can easily identify the treatment of choice.

   Actors: Medical professionals, patient/family

   Priority level: Primary

3. **Use Case:** Identify all medical conditions tested for a given treatment.

   Summary: The user wants to identify all the medical conditions tested for one or more treatments of interest. The system displays in a compacted format the corresponding medical conditions. The user can easily identify the medical condition most frequently tested for such treatment.

   Actors: Medical professionals, patient/family

   Priority level: Primary

4. **Use Case:** Identify and compare eligibility criteria.

   Summary: The user wants to identify the characteristics of the populations in a given set of trials and to compare these characteristics across trials. The system displays eligibility criteria (for example qualifying conditions, past interventions, age, gender, etc.).

   Actors: Medical professionals

   Priority level: Primary

5. **Use Case:** Compare interventions.

   Summary: The user wants to compare the interventions used in a given set of trials. The system displays the drugs, dosage, and timing used.

   Actors: Medical professionals, patient/family

   Priority level: Primary

6. **Use Case:** Identify and compare efficacy and safety outcomes.

   Summary: The user wants to identify the efficacy and safety outcomes in a given set of trials and to compare these outcomes across trials. The system displays the outcomes and specifies the intervals at which they were tested.

   Actors: Medical professionals, patient/family

   Priority level: Primary

7. **Use Case:** Identify and compare trials by geographical location.

   Summary: The user wants to identify trials by geographical location and to compare locations across trials. The system displays the geographic distribution of trials.

   Actors: Medical professionals

   Priority level: Secondary

8. **Use Case:** Identify and compare sample distribution.

   Summary: The user wants to know how many patients were recruited and how they were distributed into intervention and control groups and to compare this information across trials.

Table 7.2. *Use Cases for Users of Clinical Trials*

	Priority Level	Medical Professionals	Patient/ Family
1. Search for related trials	Primary	X	X
2. Identify all treatments tested for a given medical condition	Primary	X	X
3. Identify all medical conditions tested for a given treatment	Primary	X	X
4. Identify and compare eligibility criteria across trials	Primary	X	
5.Compare interventions	Primary	X	X
6. Identify and compare efficacy and safety outcomes	Primary	X	X
7. Identify and compare trials by geographical location	Secondary	X	
8. Identify and compare sample distribution	Secondary	X	
9. Identify and compare methodological features	Secondary	X	
10. Identify and compare trials by chronological date	Optional	X	

Actors: Medical professionals

Priority level: Secondary

9. **Use Case:** Identify and compare methodological features.

   Summary: The user wants to identify the methodological features of trials and to compare these features across trials. The system displays information concerning the blinding method, the name of the primary investigator, and the statistical methods used.

   Actors: Medical professionals

   Priority level: Secondary

10. **Use Case:** Identify and compare trials by chronological date.

    Summary: The user wants to identify and compare start and end dates across trials. The system provides a view of the chronological distribution of trials.

    Actors: Medical professionals

    Priority level: Optional

7.5 Summary

In this chapter I have provided a model I call the Integrated Information-Seeking model that leverages the concepts, theories and models described in Chapter 3. This model differs

from the traditional models in that it represents the interaction between extracting information and browsing. Moreover, it adds two stages to the information-seeking process: the aggregation and comparison-and-contrast stages. Both are important steps in fulfilling the information need. I have used the Integrated Information-Seeking Model to summarize the challenges that the information-seeking process presents to users of clinical trials and I have defined a set of design principles and use cases to guide the design of more effective interfaces to support users in seeking this type of information. In Chapters 8 and 9 I describe two visualization tools which I designed and tested to validate my findings and design principles, namely CTSearch and CTeXplorer. The purpose of these tools is to provide cognitive support to users at three stages of the information-seeking process: that of reformulating their initial queries, that of evaluating the retrieved results, and that of comparing and contrasting selected trials. As discussed previously, these are complex tasks. I approach the information-seeking problem by providing visualizations aimed at two levels of abstraction. At the beginning of the information-seeking process, a representation of results at a high level of abstraction is desirable to provide an overview of the retrieved results. As the process continues, users are interested in a detailed comparison of a narrower set of relevant trials, and representations at a lower level of abstraction are desirable. The following chapters describe CTSearch and CTeXplorer in greater detail.

Part III

Confirmatory Studies

Chapter 8

CTSearch: A Tool for Seeking Clinical-Trial Information

Based on my exploratory studies, I identified the requirements that an effective user interface for *professionals* and *lay users* seeking clinical-trial information should meet, as described in the preceding chapter. To validate these requirements, I designed two visualization tools, CTSearch and CTeXplorer. In this chapter, I discuss the design of CTSearch, give an example of its use, and give a description of a preliminary user study and its findings. I designed CTSearch to cover the use cases "1. Search for related trials", "2. Identify all treatments tested for a given medical condition", and "3. Identify all medical conditions tested for a given treatment" (cf. Table 7.2 on page 84). A discussion on CTeXplorer is presented in the next chapter. For the reader interested in existing visualization approaches to support users seeking information in electronic environments, I present a review of related work in Appendix E.

8.1 Targeted Database

To design CTSearch, I decided to use an existing database: ClinicalTrials.gov, which includes a large number of clinical-trial protocols. ClinicalTrials.gov is the world's largest repository of clinical-trial protocols. It was established by the National Institutes of Health (NIH) in response to Section 113 of the US Food and Drug Administration (FDA) Act,

which provides for the registration of all clinical trials conducted under FDA regulations. ClinicalTrials.gov was launched in February 2000, and its main goal is to provide accurate information on clinical trials with human subjects [ZIT+07]. As of 2009, ClinicalTrials.gov holds more than 70,000 trials and receives 50,000 visits per day.[1]

ClinicalTrials.gov stores trials in a semi-structured format. That is, the entry for each trial consists of a set of fields or data components such as main title, summary, eligibility criteria, outcomes, or start date. Most of the data components are filled out with a text-based description, while a few are tagged with specific medical terms from the Unified Medical Language System (UMLS).

In ClinicalTrials.gov, when users search for specific information, the results are returned as lists where the user can only access details of a resource one at a time. Thus, it is left to the user to mentally combine the information so that they can compare particular aspects of the query results. The typical search box is depicted in Figure 8.1, and the advanced search interface is depicted in Figure 8.2. The latter shows how the user can enter a query using the annotated information in the clinical trial description through an advanced search screen. The retrieved results are always represented as depicted in Figure 8.3 revealing how it is difficult to match the individual results to the specific query terms submitted without a significant amount of manual scanning through the results. The current interface, and those of other sites similar to ClinicalTrials.gov, unfortunately do not promote further iteration on the preliminary query. To iterate, a user is forced to come up with new terms for a new query. Furthermore, the context of their original results are now lost. Thus, ClinicalTrials.gov is a good candidate to be extended using the design principles introduced in Chapter 7.

Another difficulty is that the annotations added to the data may often have no uniformity; this is because different authors may use different terms to describe the same concept (in other words, synonyms). For example, documents related to liver cancer may be tagged with any combination of terms such as “Liver carcinoma”, “HCC”, “Hepatocellular carci-

[1] http://clinicaltrials.gov/ct2/info/about

Figure 8.1. *Basic search screen for ClinicalTrials.gov*

Figure 8.2. *Advanced search screen for ClinicalTrials.gov*

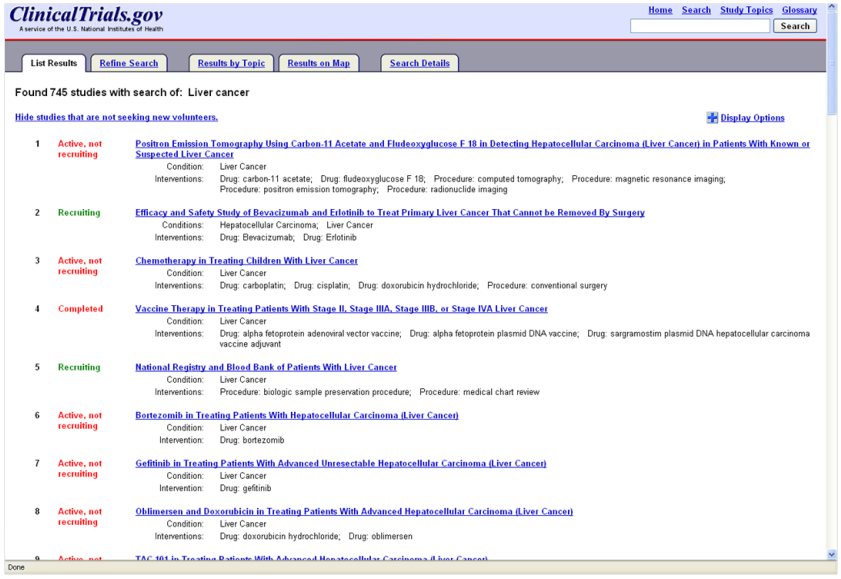

Figure 8.3. *Search results for ClinicalTrials.gov*

noma", or "Liver metastases". Additionally, sometimes, the descriptions and annotations on the data are ambiguous and incomplete (e.g., the real name of a drug tested may not be mentioned by the description or the document's annotations). Again, the set of results as shown in Figure 8.3 does not address this issue. For example, if a user enters "Liver cancer" as a search term in the condition field, 541 results are returned, on the other hand if they use "Hepatocellular carcinoma" they get 384 results. There is no way for the user to know how these result sets intersect or how they may relate to one another. It is these challenges that CTSearch addresses: support for iteratively exploring a semi-structured information space, and feedback on overlapping results from varied but related queries.

8.2 Tool Design

The aim of CTSearch is to support medical professionals and lay users in finding, retrieving, and evaluating meaningful information from a large collection of clinical-trial protocols. CTSearch's design takes advantage of fact that the documents in ClinicalTrials.gov are *semi-structured*, which means they are described by tags such as conditions, interven-

tions and outcomes. Tags are metadata assigned which documents and other sources of information [GL06]. There are a number of visualization approaches that could be applied to information seeking interfaces and each approach has its advantages and disadvantages as I discuss in Appendix E. After considering the tradeoffs among the different options, I chose the tag cloud metaphor because they afford the visualization of the tags within the ClinicalTrials.gov dataset and because the tag cloud metaphor could be extended to support my design principles as described in Chapter 7. CTSearch's design adds multiple tag clouds to the typical search box interface and permits multiple selection and deselection of terms in one or more tag cloud. Tag clouds are compact visual representations of keywords, or tags. They typically consist of single keywords; thus, a tag cloud is a set of keywords displayed in loose association in the shape of a cloud. Font size, weight, and color may be used to represent tag characteristics such as frequency and age.

In CTSearch, tag clouds were used to provide an overview of retrieved documents, where each tag cloud corresponds to one facet (data component) of the trials. The facets were chosen based on the data most frequently used as identified in the exploratory studies discussed in Chapter 6. To implement the design ideas, a computer programmer, S. Falconer, was hired by the NCBO. I supervised Mr. Falconer during the implementation of CTSearch and gave him continuous feedback on each iteration of its development. The design, implementation, and user evaluation of CTSearch was first discussed by Hernandez, Falconer and Storey in [HFS+08b] and summarized in [HFS+08a].

In Figure 8.4 a depiction of CTSearch is presented. The tool includes the typical search box (at A) and the query results summarized and grouped by meaningful categories, represented as tag clouds (at B). Each tag cloud corresponds to one dimension (or facet) of a clinical trial, for example, conditions, interventions, study design, and phases. Each tag cloud shows a view of the prevalent terms for each facet from the retrieved results. Larger fonts are used for more frequently occurring terms in the result set. The traditional list of documents is displayed on the right side of the screen (at C). One or more terms can be selected from each tag cloud to refine the search. When the user selects multiple tags, the

query will consist of the intersection or conjunction of the selected tags and the terms used in the original query. The user receives feedback regarding the number of results and the query formulated (at D). Finally, the tag clouds are collapsible as shown in Figure 8.4 (the study design and phase tag clouds are collapsed).

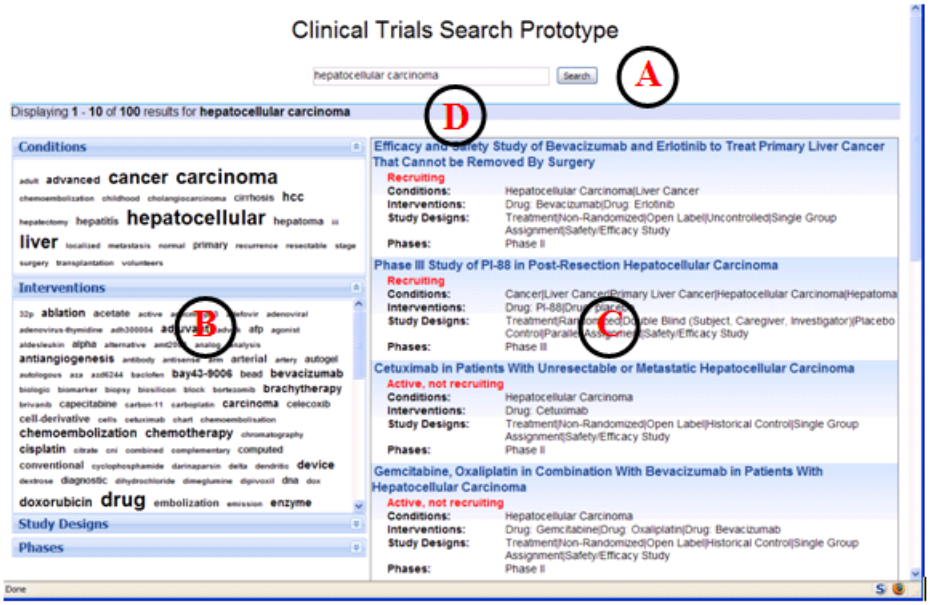

Figure 8.4. *CTSearch Interface*

8.3 Scenario

In order to demonstrate how CTSearch can be used I present the following scenario. Suppose that a trial designer is interested in designing a new trial of drugs to treat cancer of the liver (hepatocelullar carcinoma). She is interested in identifying the interventions tested for this type of cancer (this is use case 2 in Chapter 7). Using CTSearch, she follows the following steps:

Step 1: The trial designer submits a query "hepatocellular carcinoma" in the search box (cf. A in Figure 8.4). From the 100 retrieved results displayed in the traditional

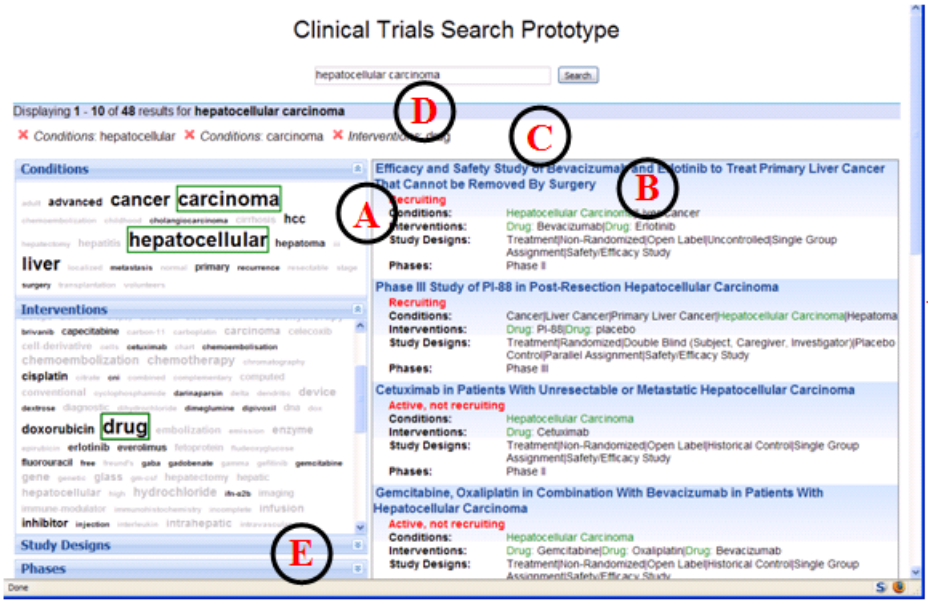

Figure 8.5. *Additional features of the results view*

list (at C) she cannot distinguish between trials in which liver cancer was the target disease and those in which it is mentioned only incidentally.

Step 2: To filter out the latter, she selects the terms "hepatocellular" and "carcinoma" from the tag cloud for diseases or medical conditions (cf. A in Figure 8.5). The number of trials is reduced to 48.

Step 3: The trial designer is only interested in drugs tested rather than any other type of therapy; thus, she selects "drug" from the tag cloud Interventions. As a consequence, all the non-related tags are grayed out across all tag clouds (cf. A and E in Figure 8.5). The tags "hepatocellular", "carcinoma", and "drug" are also highlighted by appearing in green boxes in the tag clouds and in green-coloured font in the list of results.

Step 4: She discovers the most frequent drug names used in trials that tested for liver cancer are Doxorubicin, Cisplatin, Erlotinib, and so forth.

Step 5: She receives feedback on the terms selected from the tag clouds (cf. C in Fig-

ure 8.5). The number of results and the number of filtered documents are both updated (cf. D in Figure 8.5).

8.4 User study

I designed CTSearch to be used by clinical-trial professionals and by the general public to help them overcome the difficulties they face when seeking clinical-trial data. I included the general public because it emerged as a need from my online questionnaire discussed in Chapter 6. Additionally, the general public is one of the target populations of ClinicalTrials.gov, both for recruitment into trials and as users of clinical-trial information. I conducted a preliminary user study to confirm the usefulness of CTSearch for searching and browsing clinical-trial data and formulating and reformulating queries regarding such data. The user study consisted of two parts: 1) individual sessions in which users performed tasks using two different tools (CTSearch and Baseline) for searching clinical-trial data and 2) a questionnaire asking about their experiences using the two tools and requesting input on how the tools could be improved.

Before undertaking the user study, I tested my protocol with two participants to make sure the tasks were understandable; I do not report the results of those pilot evaluations, since the participants were experts in computer applications and were outside my target population.

After obtaining ethics approval, I recruited nine participants based on the following criteria: that they were professional or lay users of clinical-trial information, and that they were 19 years of age or older. No other salient characteristics were required; in particular, the recruits could be of any gender, race, or social class or position. As it turned out, no professional users were forthcoming, and the test population consisted of nine lay users, ranging in age from 25 to 60, from various occupations, and with various educational backgrounds.

Individual sessions: The participants engaged in individual sessions using CTSearch and

Baseline. The Baseline tool is a simplified version of CTSearch that provides the same look and feel, but includes only a traditional list of results without the panes of tag clouds representing aspects of clinical trials. Thus, the only difference between the two tools is the availability of the linked tag clouds. A collection of 900 trials extracted from ClinicalTrials.gov was used as the dataset to test CTSearch. A usability laboratory was used to video-record the participants' faces and voices as they used the tools and to digitally record their interaction with the tools. I conducted individual preliminary training sessions with the participants. These began with a brief introduction to the tools and to the collection of 900 trials; I then asked the participants to perform seven randomly pre-selected search tasks with each of the tools. The order of use for each tool was randomized. The seven training tasks were defined from the information needs discussed in Chapter 6 and from the use cases defined in Chapter 7 as follows:

1. Name three popular drugs that were used in trials to treat lung cancer.
2. Find two drugs tested in few clinical trials for lung cancer.
3. Find the trials that have treated depression in patients with lung cancer. How many and what type of interventions were used?
4. Find the trials that tested Tamoxifen for breast cancer. How many trials are there? What study design features are more common in these trials?
5. Find the three medical conditions more commonly treated with the drug Docetaxel.
6. Find the most frequent medical conditions associated with Fibromyalgia considered in clinical trials.
7. Give one example of a type of cancer that is not treated with chemotherapy in clinical trials. How would you double check the validity of your example?

Following the training portion of the individual sessions, we moved on to the user study proper. The participants were asked to perform a self-selected, open-ended

task using each of the two tools, CTSearch and Baseline, in succession. They were encouraged to "think aloud" while performing the task and to take written notes on paper. The open-ended task was assigned because it is close to the real scenario of a user with a personal interest seeking clinical-trial information on the Internet. The task was described as follows:

1. Try two or more queries of your own interest about medical research and please describe what you learn doing these queries.

Questionnaire: After completing the tasks, each participant filled out a questionnaire. Nine closed-ended questions were asked to identify the participants' levels of motivation, satisfaction, and confidence using CTSearch and Baseline on a seven-level Likert scale, ranging from "strongly disagree" to "strongly agree". Some of these questions were, for example, 1) "I found the CTSearch very useful for finding interesting information", 2) "I found my experience with CTSearch very satisfying", and 3) "I would imagine that most people would need to use CTSearch for long periods of time, in order to understand a set of related documents". I designed the closed-ended part of the questionnaire based on the recommendations contained in the "Guidance on usability" ISO9241-11.[2] Seven open-ended questions were used to identify the usefulness of CTSearch, such as, 1) "What type of queries would you like to formulate using CTSearch?" 2) "What type of tasks did you find difficult to do and why?" and 3) "Did you use the linking between tag clouds to interpret and understand the data?".

8.5 Findings

The findings of this user study are derived from the information recorded during the tasks in the individual user sessions and from the responses to the questionnaire. I played-back the

[2] http://www.iso.org/

taped sessions and took notes. I organized my notes looking for emerging themes [GC00, Cre04]. I complemented my observations with the participants responses to the post-study questionnaire. The results are organized according to usefulness of the tool, self-reported user satisfaction, and user suggestions for improvements:

CTSearch usefulness: During the individual user sessions with CTSearch, I observed that the participants refined their queries by selecting terms in the tag clouds while doing the seven training tasks and the open-ended task. The number of times a participant used this feature ranged from two to 12. Although three users refined their queries only a few times (four times or less), they were able to accomplish some tasks by hovering on terms and by scrolling the tag clouds. The users correctly inferred that the size of the font represented the frequency with which a term appeared in the results. Participants appeared to carry out the assigned task confidently, based on "skimming" the tag clouds and by observing and comparing the font size of terms. Interestingly, users did not dismiss the traditional list of results as a source of information and frequently referred to the list to *double check* their impressions and possible answers to the training questions. For example, if the user searched for "tamoxifen" and then selected "breast cancer" from the conditions tag cloud, she would read the conditions field from the list of results to confirm that each document would contain "breast cancer".

Three users were observed to experience a mild level of frustration while using this tool. In all of these cases, the users could not verbalize the strategy needed to solve the difficult task and decided to move on to the next task.

In the user study, users were asked to formulate two or more queries of their own interest using CTSearch. One out of nine users refused to elaborate free queries arguing that she did not have an interesting query to formulate. One user posed 10 queries and was asked to stop due to time constraints. One user performed five queries, another performed four, and a third performed two queries each, while three

users performed one query each. Some interesting comments were: "I find drug names I am not familiar with", "Wow, melatonin is used for premature birth brain damage! Wow, this is interesting! This is the most interesting thing I learned this week", and "Actually, I want to look for a couple more" (i.e., perform more queries).

Baseline usefulness: Five users vocalized that they experienced mild to high frustration during the training tasks. In some cases, the users refused to finish the tasks, although they were able to describe the steps needed to accomplish them. Three out of nine users assumed an order of relevancy in the list of results that would accommodate their needs. For example, for task #5 "find three medical conditions more commonly treated with Docetaxel", if the user searched for the drug Docetaxel then he/she thought the results were given in order of relevancy for conditions. This false assumption meant that the user would look at the first few results and would extract the medical conditions to answer the assigned task.

Five of nine users refused to perform open-ended queries using the Baseline tool, stating that they did not have a query of interest to formulate. However, this behavior may be a result of the user becoming tired after performing the training tasks. We attempted to mitigate this by alternating the order of the tools for each user. Of the four users who were willing to elaborate open-ended queries using Baseline, three performed one or two queries, and one formulated seven queries.

User Satisfaction: I posed several direct questions to our participants in the post-task questionnaire to assess their levels of satisfaction with CTSearch and Baseline. The results from these questions are summarized in Table 8.1. They show that users consistently favoured CTSearch over the Baseline tool in all respects. For example, all nine users disagreed with the statement "I experienced frustration while using CTSearch", while only two users disagreed when the same statement was applied to Baseline.

I also asked the participants if they found the synchronization between the tag clouds

Table 8.1. *Answers to the Closed-ended Questions*

	Strongly Disagree	Somewhat Disagree	Lightly Disagree	Neutral	Lightly Agree	Somewhat Agree	Strongly Agree
I found the CTSearch very useful for finding interesting information	0	0	0	0	0	3	6
I found the Baseline very useful for finding interesting information	1	3	0	1	1	1	1
I found my experience using CTSearch very satisfying	0	0	0	1	0	4	4
I found my experience using Baseline very satisfying	3	1	2	1	0	1	0
I imagine that most people would need to use CTSearch for long periods of time	5	2	1	0	2	0	0
I imagine that most people would need to use Baseline for long periods of time	0	1	1	1	0	1	5
I experienced frustration while using CTSearch	4	2	3	0	0	0	0
I experienced frustration while using Baseline	1	0	1	1	0	4	2

helpful for interpreting and understanding the data. All but one indicated that they found it very useful; for example, one user commented, “Selecting other tags associated with your keyword search helps with either finding more relevant trials or narrows your search to a more specific trial”. One participant also commented that she found the clouds useful for understanding the frequency with which a term appeared in the trial set. Seven of the nine users also indicated they were confident about the reliability of the results found through the tag-cloud queries.

In general, comments on the CTSearch tool were quite favourable; for example, when asked if they found the tool difficult to use, one user answered, “No difficulties”, while another wrote, “I would like to use this tool for all my Internet searches”. Most of the users (seven of the nine) also stated that the linking between tag clouds helped them to interpret and understand the data. These are some of the comments users provided: 1) “Yes, I found looking at the words related to topic helped me refine my search. Keywords related to topic narrowed my focus on topics”; 2) “Yes,

by selecting other tags associated with your keyword search helps with either more relevant trials or narrows your search to a more specific trial"; and 3) "GREAT! It got you where you wanted to go". Overall, there were more negative comments on the Baseline tool than on CTSearch, e.g., "I found it very busy with information everywhere", "I lost more time", and "I needed to check a large number of possibilities in order to answer the question".

Suggestions for improvement: The questionnaire included several questions designed to obtain suggestions for improvement to CTSearch. I also obtained such suggestions from comments made by the participants during the user study.

One particular theme that appeared during the survey and study tasks related to displaying results of a "negative" search. Participants had a difficult time understanding what it meant when CTSearch grayed out tags unrelated to the terms included in the search. Participants also expressed dissatisfaction at being unable to formulate a logical NOT in their queries.

One user indicated that he would like to know the specific frequency with which a given tag appears in the trials, rather than being provided with only a rough indication via the font size. Another suggested that it would be preferable for the non-relevant tags to be removed from the display, rather than just "grayed-out".

Another area in which three users found room for improvement was with regard to the list of results. First, they wondered whether there was any meaning to the order of results. (The answer is that the results appeared in order of relevancy.) Secondly, they wished to be provided with the option of choosing the variables (for example, relevancy, date of publication, or alphabetical order) used to rank the result list. The addition of spelling correction and phrase search was recommended by two users each. One user suggested adding the ability to compare results of trials. Finally, several users suggested making more use of colored and bold fonts, for example, to relate search terms to keywords in the results.

8.6 Limitations

There are a number of limitations of the user study conducted to evaluate CTSearch: first, I designed CTSearch for medical professionals and lay users seeking clinical-trial information. However, in this study, I recruited only lay users. These users proved unfamiliar with advanced queries or with the use of tag clouds. I wanted to discover and fix potential problems these users might have in using such a tool before testing it in further studies with medical experts. Since these further studies remain to be conducted, the evaluation of CTSearch remains incomplete. Second, regarding the tasks used in the training sessions, I acknowledge that they were biased towards the CTSearch features, and thus may have put the Baseline tool at a disadvantage with respect to perceived user satisfaction; for example, CTSearch summarizes query results into conditions and interventions and the seven tasks were related to information contained in those facets. Third, it is also possible that the participants might have guessed that the CTSearch tool with tag clouds was the interface of interest to me as an HCI researcher, which may have inclined them to favour CTSearch over the Baseline tool in order to please the researcher. Finally, in the open-ended task, it emerged that the limitations of the database (ClinicalTrials.gov) may have affected user satisfaction. For example, since the results of the trials were omitted from the set of protocols, the users' attempting to answer such questions as "what is the efficacy of drug X" were frustrated.

8.7 Discussion

My observations of how the participants used CTSearch and the Baseline tool indicate that they had more difficulties using Baseline than CTSearch. These observations are consistent with the users' perceived satisfaction as revealed in the post-study questionnaire. The participants learned to use the CTSearch tool and understood how to use its tag clouds to iterate on their queries as discussed in section 8.4. The study also provided insights on

how the tool might be improved for browsing semi-structured clinical-trial databases. The findings of the user study suggest that multiple tag clouds are indeed useful for searching and exploring collections of clinical trials.

Chapter 9

CTeXplorer: A Tool Beyond Information Seeking

Comparing collections of related clinical trials is an important but complex and time-consuming activity for professional users, as I discussed in Chapter 6. The comparison of trials is an activity that traditionally has required the manual extraction and manipulation of data from journal articles, which complicates cross-trial comparison. To validate my findings on the information needs of *professional users* beyond seeking for trials, I designed and ran preliminary evaluations on a visualization tool named CTeXplorer. I designed CTeXplorer to approach the use cases "4. Identify and compare populations across trials", "5. Identify and compare Interventions across trials", and "6. Identify and compare efficacy and safety outcomes across trials" (cf. Table 7.2). In this chapter, I describe CTeXplorer, a tool I designed to meet the needs of medical professionals, who wish to compare in detail the designs of multiple clinical trials based on their eligibility criteria, interventions, or study outcomes. I start by describing the design of CTeXplorer and then discuss a user study I conducted to evaluate the tool's usefulness.

9.1 Tool Design

To design CTeXplorer I asked the expert users from UCSF to provide me with a sample of clinical trials. They responded with a sample of twelve trials on the use of antiretrovirals in

the prevention of mother-to-child transmission of HIV. The experts provided the sample in the form of a spreadsheet with data extracted from journal articles on these trials that were published between 1994 and 2006. The data referred to the following trial design elements: eligibility criteria, interventions, outcomes, and sample size.

To help implement the design ideas, a computer programmer, C. Bennet was hired by the CHISEL Lab at the University of Victoria and to help finish the implementation a second programmer, C. Callendar was hired by the NCBO. I supervised Mr. Bennet and Mr. Callendar during their contribution to the implementation of CTeXplorer and gave them continuous feedback on each iteration of its development. The design, implementation, and user evaluation of CTeXplorer was first discussed in [HCSS08] and summarized in [CHSS07].

The approach selected for the design of CTeXplorer employs multiple linked views and dynamic queries. Used together, these two visualization techniques support queries which filter trials and update the views to reflect the filtering.

The technique of *multiple views* displays one or more different graphical representations of the same data on the screen and supports the coordinated exploration of information. The graphical representation used can include basic charts, such as bar charts and scatterplots, or more sophisticated representations, such as treemaps [Shn08], or TableLens [RC94]. In a simple multiple-view system, the number and type of views is fixed. In more advanced interfaces, the system permits the representation of varied numbers and types of views. An influential example of a multiple-view system is the SnapTogether visualization tool developed by Chris North [NS00].

The technique of *dynamic queries* is based on the use of widgets such as sliders and check boxes to perform queries on the underlying dataset. The user receives feedback immediately as the widgets are operated, with the input and output being integrated in a single interface. Examples of dynamic queries are the Attribute Explorer [TSWB94], the LifeLines [PMS+98], and the FilmFinder [AS94].

The CTeXplorer interface presents three linked views: Eligibility Criteria View, Inter-

ventions View, and Outcomes View (cf. Figure 9.1). The views are ordered from top to bottom and left to right in a chronological sequence that follows the course of the trial. Users can see at a glance relative trial sizes, which drugs were administered in which trials and during which trial period (i.e., pregnancy, labor, and postpartum in the mother and shortly after birth in the infant), and the time points at which maternal and infant safety and infant HIV infection were assessed. At the user's discretion, the main interface can also include one static view, the Geographical View, which represents the trials by geographical location Each view can be expanded to full-window size to allow users to focus on specific data.

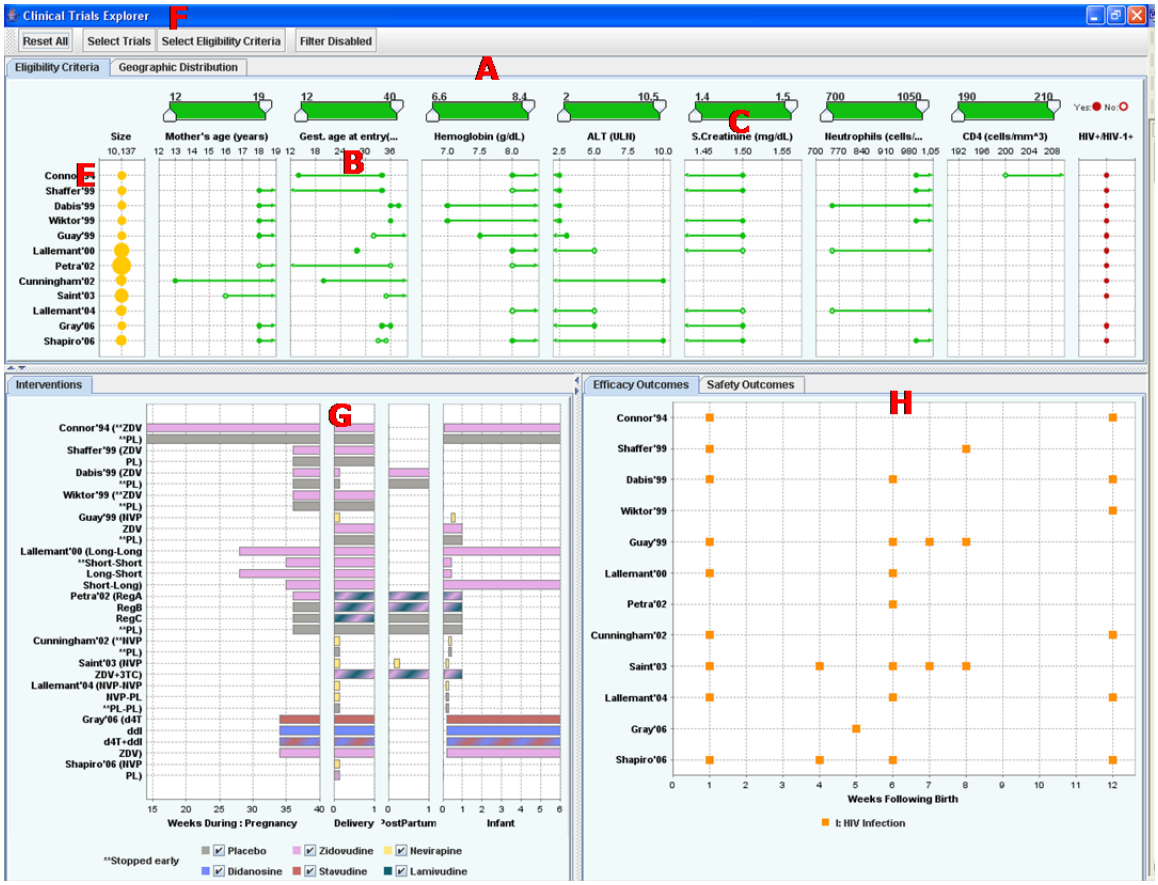

Figure 9.1. *CTeXplorer Interface*

The Eligibility criteria view (cf. A in Figure 9.1) displays eight criteria per trial by default, which constitute a partial list of the available set of eligibility criteria variables. In the *Eligibility Criteria View*, each trial is shown on a single row and the eligibility criteria are displayed as columns. Eligibility criteria are represented according to two types of variables: continuous and dichotomous.

Continuous variables are represented by glyphs to accommodate precise comparative operators (e.g., "greater than", "greater than or equal to", or "less than"). The glyphs allow the representation of values when an explicit upper or lower limit was not specified.

A glyph is composed of one or two anchors (either empty or filled), a line, and an optional arrowhead to indicate direction in the case of open-ended value ranges. When the lower and upper limits of a variable are specified a line is used with either filled or empty circle anchors at each end. An empty circle anchor is used for "less than" or "greater than" operators and a filled circle indicates "less than or equal to" or "greater than or equal to". For example, the gestational age for the trial "Connor '94" (cf. B in Figure 9.1) is represented by a glyph indicating a range from "equal to 14 weeks" to "equal to 34 weeks" of pregnancy.

Using bi-sliders, the user can filter trials based on continuous values. Only those trials that have data values overlapping within the range of the bi-slider will be displayed and other trials will be filtered out (cf. C in Figure 9.1).

Dichotomous variables are displayed in the *dichotomous* Eligibility Criteria View side-by-side with continuous variables. I use filled circles to indicate "Yes" (i.e., criterion must be satisfied for inclusion in the trial), empty circles to indicate "No" (i.e., criterion must not be satisfied for inclusion in the trial), and no circle to indicate "Not specified" (cf. D in Figure 9.1).

In the left-upper part of the Eligibility Criteria View, there are four buttons: "Reset All", "Select Trials", "Select Eligibility Criteria", and "Filter Enable/Disable". "Reset All"

reinitializes the state of the application to display all trials again and is commonly used after querying the trials (see F in Figure 9.1). "Select Trials" is used to choose trials for comparison; by default, all the trials are listed in order of publication date of the article in which its results were reported. "Select Eligibility Criteria" adds or removes selected criteria from the view. "Filtering Enable/Disable" switches between graying out trials that do not satisfy the query or removing them from the list.

The size of the population in each trial is represented by a circle of proportional size, and the aggregated total of participants in all the trials is displayed at the top of the column. The number of participants in each trial can be displayed by hovering over the circles (see E in Figure 9.1)

In the **Interventions view**, CTeXplorer uses the horizontal axis to represent time, so that treatment is covered from beginning to end. This axis is not linear, but consists of segments that cover pregnancy in weeks, labor from onset to delivery, and both mother and child postpartum periods in weeks (cf. G in Figure 9.1). Interventions are a complex aspect of clinical trials, and it is challenging to represent them comparatively. In CTeXplorer, horizontal bars are used to represent treatments used in the trials according to the period of intervention (i.e., during pregnancy, during labor, postpartum mother, and postpartum infant). Hovering the cursor over the bar reveals the exact start and stop times for the treatments.

Coloured bars show drugs applied to each trial group. A legend at the bottom of the chart indicates the colours used to represent each drug. For example, the trial group "Connor ZDV" uses the drug Zidovudine, and the group "Connor PL" uses a placebo (see G in Figure 9.1). There are cases where two drugs are used, for example, the trial group "Gray d4T+ddl" used two drugs, Didanosine and Stavudine, simultaneously. To indicate that multiple drugs were given, bars are coloured with two or more colours.

The user can select and deselect drugs administered using the check boxes in the legend at the bottom of the chart. This filtering may alter the number of trials in the visualization, in which case all related views are updated simultaneously. Thus, if we are interested in

discerning which trials used Nevirapine or Lamivudine, we select those two drugs and deselect all the others.

CTeXplorer uses two asterisks to indicate trial groups whose treatment was stopped early. The representation of multiple groups per trial, especially when stopped groups are included, was a challenging design problem. In the interface shown in G of Figure 9.1, the horizontal bars represent the treatment period of the original trial design.

The **Outcomes View** uses a scatter plot to show clusters of outcomes at the given points in time. Coloured glyphs are used to easily distinguish between different outcome types. The Outcomes View is aligned horizontally with the Interventions View. The glyphs are described in a legend at the bottom of the view. Outcomes are divided in two panes, "Efficacy Outcomes" and "Safety Outcomes", and the user can toggle between the two (see H in Figure 9.1). As with the Interventions View, the user can select and deselect outcome types using check boxes in the legend area. This filtering will update all other views. For example, we may be interested in finding the trials where the safety outcome "Mother Anemia" was evaluated.

The **Geographic View** is intended to display the location of the trials. In CTeXplorer, the location of each trial is identified on a world map (see Figure 9.2). By selecting tabs, the user can toggle between the Eligibility Criteria View and the Geographic View, each of which is displayed across the top of the main window.

9.2 Scenario

In order to demonstrate how CTeXplorer can be used I present the following example. A trial designer is interested in designing a new trial of drugs to treat mother-to-child transmission (MTCT) of HIV during pregnancy and childbirth. She is interested in finding trials where Lamivudine was tested individually or in combination with other drugs. She is also interested in testing whether patients given these drugs starting in the early stages of pregnancy were less likely to transmit HIV to their offspring. She is interested in finding gaps in

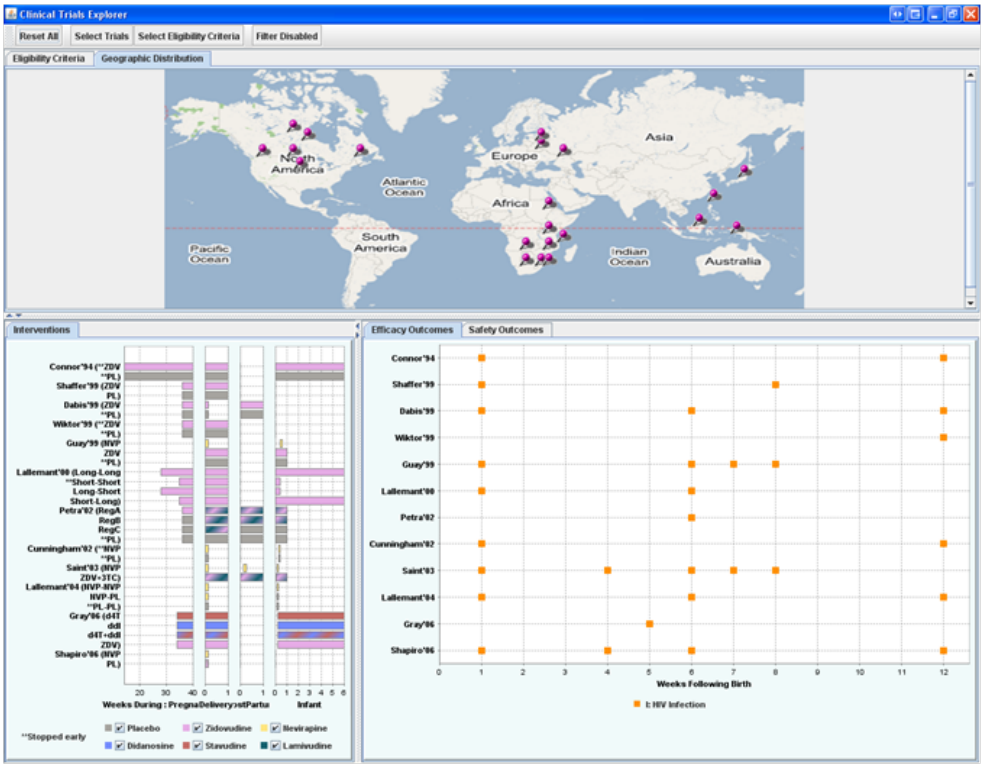

Figure 9.2. *Geographical View Across the Top*

previous trials, untested hypotheses, or drug combinations that have yet to be tested. Using CTeXplorer, she follows the following steps:

Step 1: The trial designer selects only Lamivudine in the Interventions View and deselects all other treatments. She clicks the "Filter Disable button" to filter out the trials that do not match her query. She finds that there have been only two trials were Lamivudine was used, namely Petra'02 and Saint'03. She also notices that in both trials, Lamivudine was administered in combination with Zidovudine during delivery and postpartum and to both the mother and the infant (cf. A in Figure 9.3).

Step 2: She discovers that there was little overlap in gestational age between the two trials (cf. B in Figure 9.3). Petra'02 recruited participants with less than 36 weeks of pregnancy whereas Saint'03 recruited participants with more than 35 weeks of preg-

nancy. For her purposes, the trial Saint'03 included participants that she considers as being in a late stage of pregnancy.

Step 3: She notices that in Petra'02, the one trial involving Lamivudine used in an early stage of pregnancy, the incidence of infants infected with HIV was tested on only one occasion, six weeks after delivery (cf. C Figure 9.3)

Step 4: She decides to search for all trials that recruited participants in the early stages of pregnancy. She clicks the "Reset All" button and moves the right slider for gestational age to 25 weeks. Four trials included patients in early stages of pregnancy although they were not restricted to recruit only patients in this stage. Those trials are: Connor'94, Shaffer'99, Petra'02, and Cunningham'02 (cf. Figure 9.4). Additionally she notices that the trial Lallemant'04 did not report the gestational age of the participants.

Step 5: She finds that only three out of four trials treated mothers during pregnancy. Those trials are Connor'94, Shaffer'99, and Petra'02; and the drug tested was coincidentally Zidovudine (cf. Interventions View in Figure 9.4). She decides to design her trial to recruit patients in early stages of pregnancy and to combine Lamivudine with Zidovudine to treat the participants during pregnancy, during delivery, and postpartum.

9.3 User Study

I designed CTeXplorer to be used by researchers and clinicians who are interested in clinical trials and have minimal experience with computer applications. I conducted a user study to confirm my understanding of the information needs of such users and to confirm the usefulness of the tool. The user study consisted of three parts: individual sessions in which users performed predefined tasks with CTeXplorer, a post-study questionnaire, and a focus-group session.

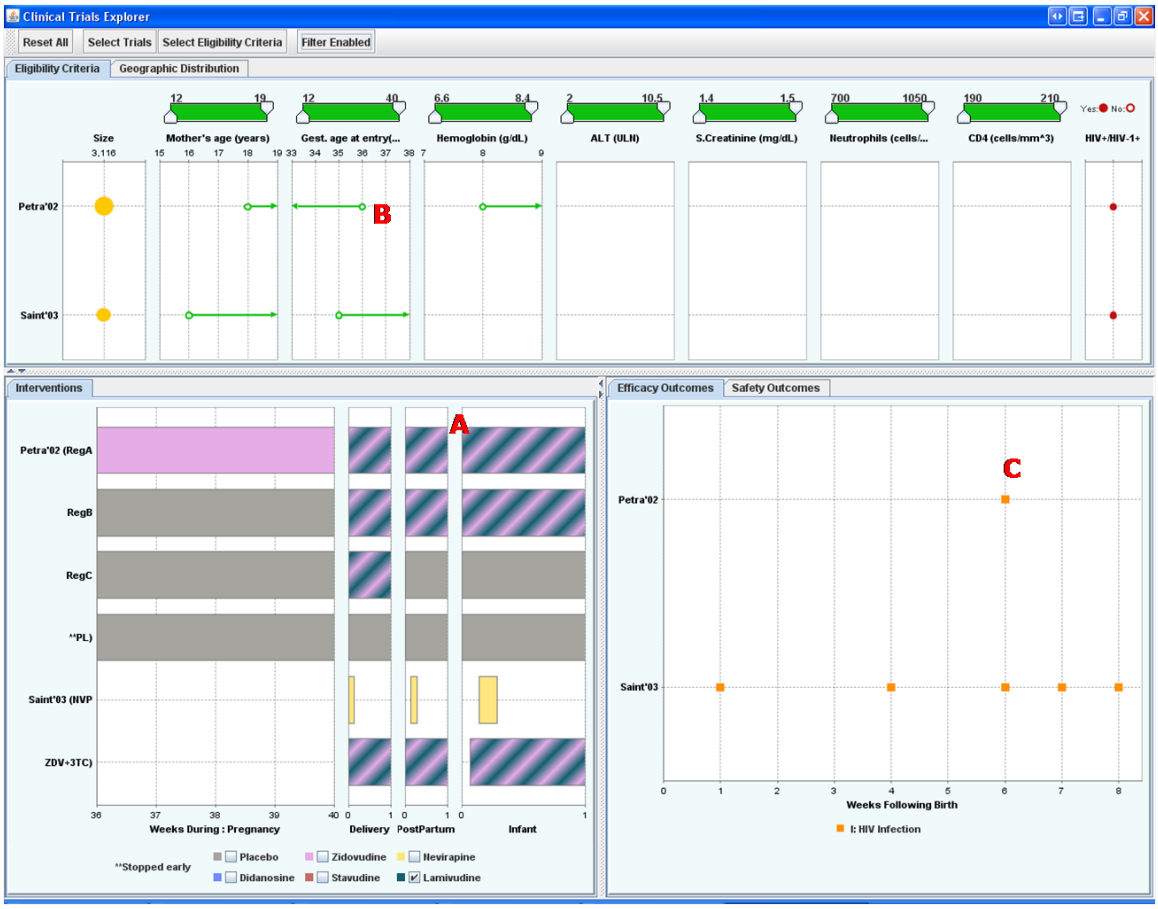

Figure 9.3. *Trials Testing Lamivudine*

Before proceeding with the study, I tested my protocol with three participants to make sure the tasks and questions were understandable. Additionally, to validate my design, I described CTeXplorer to students in an Information Visualization class and received anonymous feedback from nine of these students. Also, I discussed the design of CTeXplorer with a research team highly trained in HCI (CHISEL Lab at University of Victoria). Those informal evaluations are not mentioned in detail because none of those users were experts in clinical trials, and thus were outside my target population. I used the feedback received in those informal evaluations to refine CTeXplorer. It is this refined version of the tool that has been described above and that I tested in the user study.

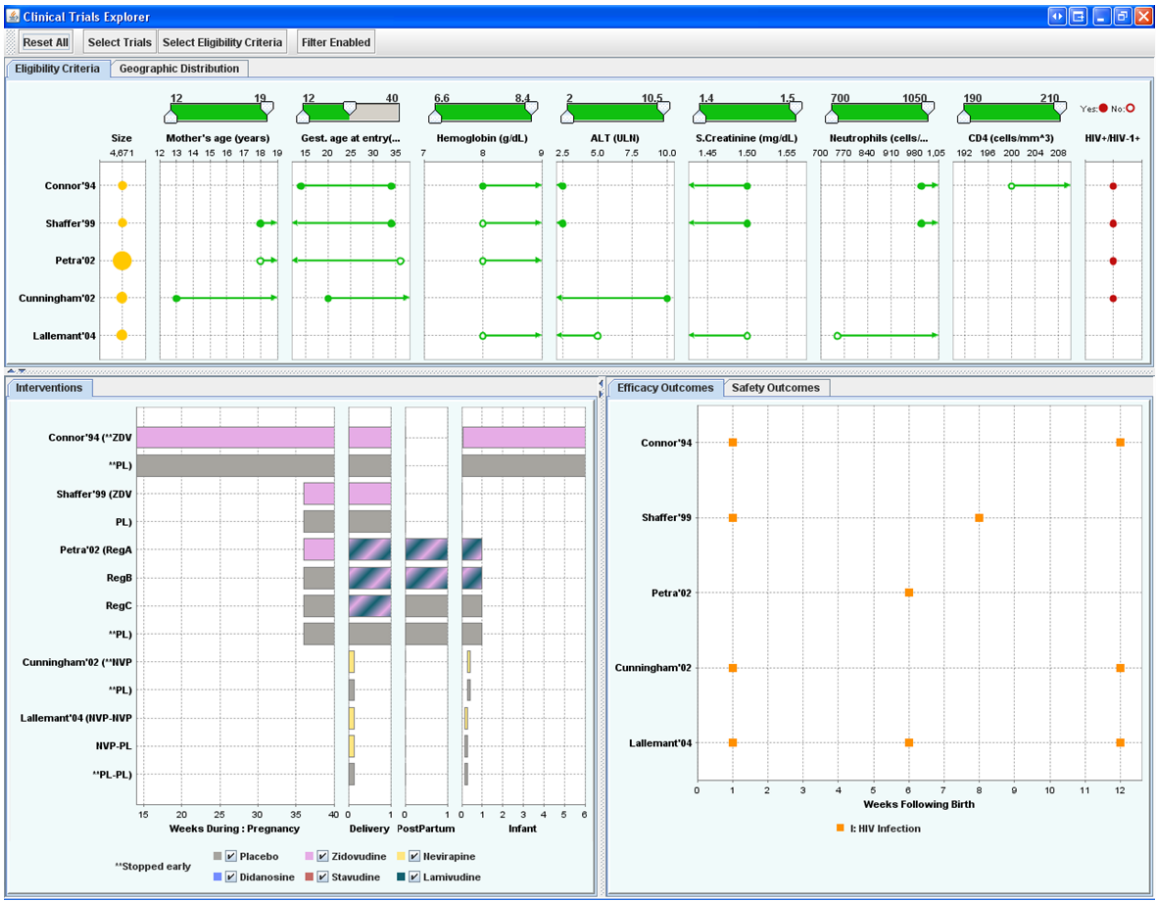

Figure 9.4. *Trials Including Participant in Early Stages of Pregnancy*

After obtaining the approval of the Human Research Board Ethics at the University of Victoria, I recruited participants based on the following criteria: that they were professionals in the clinical-trials field and that they were 19 years of age or older. No other salient characteristics were required; in particular, the recruits could be of any gender, race, or social class or position. In the end, three participants were recruited for the study: 1) an HIV trial designer, 2) a systematic reviewer, and 3) a Cochrane editor.[1]

Individual sessions: The participants engaged in individual sessions using CTeXplorer. Portable usability-lab equipment was used to video-record their faces and voices and

[1] http://www.cochrane.org/

to record their use of the tool. A senior interviewer was present and took handwritten notes during the sessions. Participants were given a brief orientation on the functions of the tool and a brief description of the collections of trials. They were then asked to carry out a set of 13 tasks (defined in terms of our use cases) and were encouraged to "think aloud" while performing these tasks. Some of these tasks were 1) What is the most common pregnancy stage of participants? and 2)What is the age range of the population under study? See Appendix D for the complete list of tasks used in the study. The tasks were designed so that they would require a user to search for specific clinical-trial information, to compare data across trials, to find relations among trials, and to find gaps in the data. Users needed in average 45 minutes to complete the tasks.

Questionnaires: After completing the tasks, each participant filled out a questionnaire. Six close-ended questions were asked to identify the levels of motivation, satisfaction, and confidence of the participant using CTeXplorer on a Likert scale with seven levels, ranging from "strongly disagree" to "strongly agree". Some of these questions were, for example, 1) Do you feel motivated to use the tool? and 2) I found the organization of data appropriate for the domain. I designed the close-ended part of the questionnaire based on the recommendations contained in the "Guidance on usability" ISO9241-11.[2] Eight open-ended questions were used to identify insights that the CTeXplorer visualizations provided to the user, such as, 1) Could you describe a new trial that may fill some gap in the existing trials? and 2) Did the tool allow you to detect similarities and differences among the MTCT trials? I designed the open-ended part of the questionnaire to explore whether the tool is useful for specific tasks. See Appendix D for the complete questionnaire used in the study. The questionnaire was set up using OVO Studio logger software [21].

Focus Group: The users also participated in a focus group. This session was led by a senior researcher to eliminate any bias that I might introduce. Open-ended questions

[2] http://www.iso.org/

were asked to determine whether the participants' information needs were satisfied by CTeXplorer. Examples of questions asked included: 1) Describe how you currently explore information on clinical trials 2) Did you find the tasks representative of the kinds of tasks you would normally do when exploring clinical-trial data? and 3) Describe your experience of understanding the trials that were presented through the tool. See Appendix D for the complete list of questions asked in the focus group session. There was a lively discussion that provided many insights on the current design and possible future modifications.

9.4 Findings

The findings of this user study were derived from an coding our observations, the recorded information on the tasks performed during the individual sessions, the answers from the questionnaire, and the input received during the session with the focus group. I looked for emerging themes, which is a typical analytical technique used in qualitative studies [GC00, Cre04].

The findings were as follows:

Eligibility Criteria View. Users explained that it is not enough to know the criterion used to select the population; but also that they need to know and compare the values of baseline features. One user suggested displaying just one criterion at a time with the option to cycle through all of them by clicking "next" rather than presenting a list for selecting and deselecting criteria.

Interventions View. It was suggested by one user that another view is needed to represent the total number of participants being administered each drug (e.g., How many people took Zidovudine?). Two users required that the size of the study groups in each trial should also be displayed.

Outcomes View. All three users felt that the Outcomes View should display not only the

variables tested in the trials but also the values obtained for each of the variables.

User Satisfaction. The questionnaire revealed that the three users were motivated to use CTeXplorer (e.g., "I am very impressed and see practical applications", "This tool can revolutionize how we look things"). They said they felt confident when using CTeXplorer and felt the system was reasonably easy to learn and use (e.g., "I felt surprisingly comfortable").

Providing Insight. The three users stated that CTeXplorer supports the comparison of trials, making it feasible to design new trials based on perceived gaps in the data revealed by the tool. CTeXplorer also supports perception of relationships among data in different views: as one user observed, "I could see that the Dabis trial had few interventions but many outcomes". The three users thought that the organization of data was appropriate for the domain.

Additional Requirements. The user study revealed some requirements that had not emerged from the interviews with the two experts. The users asked for links to the raw data (spreadsheet or journal paper) to verify first impressions of the trials; the ability to export and import data from other applications (e.g., Excel); and a mechanism for ranking trials on the basis of reliability or credibility. Since the three professional users typically work in collaboration with other researchers, they required a means of sharing their understanding of the trials with their collaborators. They also requested a way to annotate, highlight, comment, or tag the data in the views. Users felt they should be able to group trials according to similarities in specified parameters. They requested that the target system should afford more flexibility to the user in deciding what information to display at a given time. They required that trial results be visualized together with the elements of trial design. Finally, they asked for the ability to save views.

During the individual sessions, it was challenging to get the participants to think aloud, and they had to be gently reminded to talk to themselves while performing their tasks.

Only one user followed this request, and the recording of her speech proved very revealing. The other two participants were too absorbed in the tasks and in experimenting with CTeXplorer to verbalize what they were doing. In these sessions, two participants completed the 13 tasks correctly. The third participant completed only 10 of the 13 tasks, seven of them correctly and three incorrectly. The interpretation of these results may be quite controversial. I designed CTeXplorer to support understanding of related trials and this goal was achieved by two of three participants. Why the third participant did not achieve a clear understanding of the displayed data was not clear. Since this participant was the one who complied with the suggestion to use the "think aloud" method, the audio recordings revealed that the user worked hard to understand the questions and come up with sensible answers. One reason for the lower performance may be that the participant was rushed and anxious about a pending appointment. Another reason might be that the screen size of some of the view elements was small; the participant's body language suggested difficulties in viewing data on the screen.

9.5 Limitations

The user evaluation of CTeXplorer was aimed at confirming user requirements and to test the usefulness of CTeXplorer. I acknowledge that the validity of my findings can be questioned because of the small number of participants in the user study (only three users). Typical HCI studies have at least five users. My sample is below the traditional number of participants for this type of study. However, by combining different strategies in this qualitative study (individual sessions, questionnaire, and focus group) I achieved a high degree of understanding of the information needs of professional users of clinical-trial information, who are the target users of the tool. I believe that the close agreement between the results of the user study and those of my interviews with the two expert users lends credibility to my findings.

A problem with the design of CTeXplorer is that it is based on one domain of clinical

trials, that is MTCT trials. The feasibility of acquiring trials in other domains is rather low. Manual extraction of key information is time consuming and requires highly trained staff on medical sciences. This challenge can be solved by using existing clinical trials repositories where trial design and methodological features are already captured such is the case so repositories as ClinicalTrials.gov.

9.6 Discussion

The user study results show that CTeXplorer is a good start for providing support for clinical-trial experts who are performing detailed evaluation and comparison of related but heterogeneous trials. As a result of this exploratory case study, I gained an understanding of the value of visualization in supporting comparisons of clinical trials.

One challenge for conducting this study was the difficulty of recruiting participants from the pool of professional users of clinical trials, since these are highly paid individuals with tight schedules. This challenge could be mitigated in future studies by making CTeXplorer available online, including recent enhancements based on the requirements revealed in the user study, and by promoting the importance of this work in the medical field.

Chapter 10

Lessons Learned

This dissertation addresses the problem of providing a better experience to Internet users seeking medical information derived from clinical trials. To this end, I have investigated in depth the information needs and wishes of users of such information, I have identified a set of requirements and use cases to design more effective interfaces to support the information-seeking process, I have designed two visualization tools to meet those requirements, and I have tested the tools with users performing realistic tasks. The empirical findings of this research can be used by designers of user interfaces to produce more effective information-seeking tools. In this chapter, I discuss the contributions of my research, I consider its limitations, and I propose future directions for research.

10.1 Research Questions and Answers

In this dissertation I posited two research questions which I answered as follows:

Q1: What are the information needs of users searching for clinical-trial information on the Web?

Users need to find a set of relevant trials to answer their questions, to group clinical-trial information according to specific categories such as conditions, interventions, outcomes, and eligibility criteria, compare information within each group and among groups, and explore the summarized information in multiple ways.

Q2: How can we improve current Web-search interfaces for clinical-trial information?

In Chapter 7 I have proposed to add the following features to the typical search box: provide the user with additional terms to facilitate query reformulation, display results clustered by category, link clusters to explore relationships between clusters and among clusters, display the results in a compact format, support a browsing mechanism, and support the selection of multiple terms and multiple categories. The pertinence of my proposals were confirmed in Chapters 9 and 8 with the design and evaluation of two visualization tools for seeking and comparing clinical trials.

10.2 Contributions

The investigations and findings reported in this dissertation contribute in several ways to the literature on the topic of seeking medical information.

An Integrated Information-Seeking Model: After reviewing the relevant scholarly literature on Information Seeking, I proposed a refined model of the information-seeking process that integrates and expands upon previous models. The integrated model reflects the iteration between searching and browsing activities as part of the seeking process. Additionally, I defined the stages of aggregating and comparing/contrasting results which are crucial stages to finding the answers to the users' information needs. This model reflects the information-seeking process followed by users of clinical trials and can be applied to the general information seeking users. The Integrated Information-Seeking Model is described in Chapter 3.

Behaviour of user's searching on the Internet for medical information: I analyzed previous query-log studies on health and medical information. An extensive search of the query-log literature turned up eight studies on the behaviour of Internet users searching for health and medical information. To the extent possible, I compared the research questions addressed by those eight studies and their respective findings. The results suggest that, 1) in general, Internet users searching for health and medical

information behave similarly to users searching for general information; 2) they are frequently interested in finding information on interventions and on medical conditions; and 3) these users have difficulties with the use of technical medical terms in formulating searches.

Information needs of clinical-trials' users: I developed and used multiple research instruments to gain an in-depth understanding of the questions posed and tasks performed by users of clinical trials during the information-seeking process. In the exploratory stage of my research these instruments consisted of moderated discussions with experts, an on-line questionnaire, and a qualitative query-log analysis. In the confirmatory stage of my research I designed and tested two prototypes CTSearch and CTeXplorer for the dynamic exploration of clinical-trial data. CTSearch to be used both by professionals wishing to inform their practice with up-to-date, evidence-based information and by patients wishing to access up-to-date medical information. CTeXplorer for dynamic comparison of clinical-trial data aimed at professionals assessing scientific information.

A set of design principles for clinical-trials Web interfaces: Based upon the knowledge gained in my exploratory studies, I proposed a set of design principles to inform the design of more effective interfaces to support the Internet user seeking clinical-trial information.

Tag clouds for exploring semi-structured documents: In this research, I tested the usefulness of multiple tag clouds for exploring semi-structured documents. This is an approach that can be applicable to domains other than that investigated herein. One example is the implementation of CTSearch by C. Treude to explore Jazz work items as depicted in Figure 10.1.

A call to action: In this research I have explored and documented a problem that has not been specifically addressed in the literature, that is "How to design better user interfaces to support the needs of clinical-trial users."

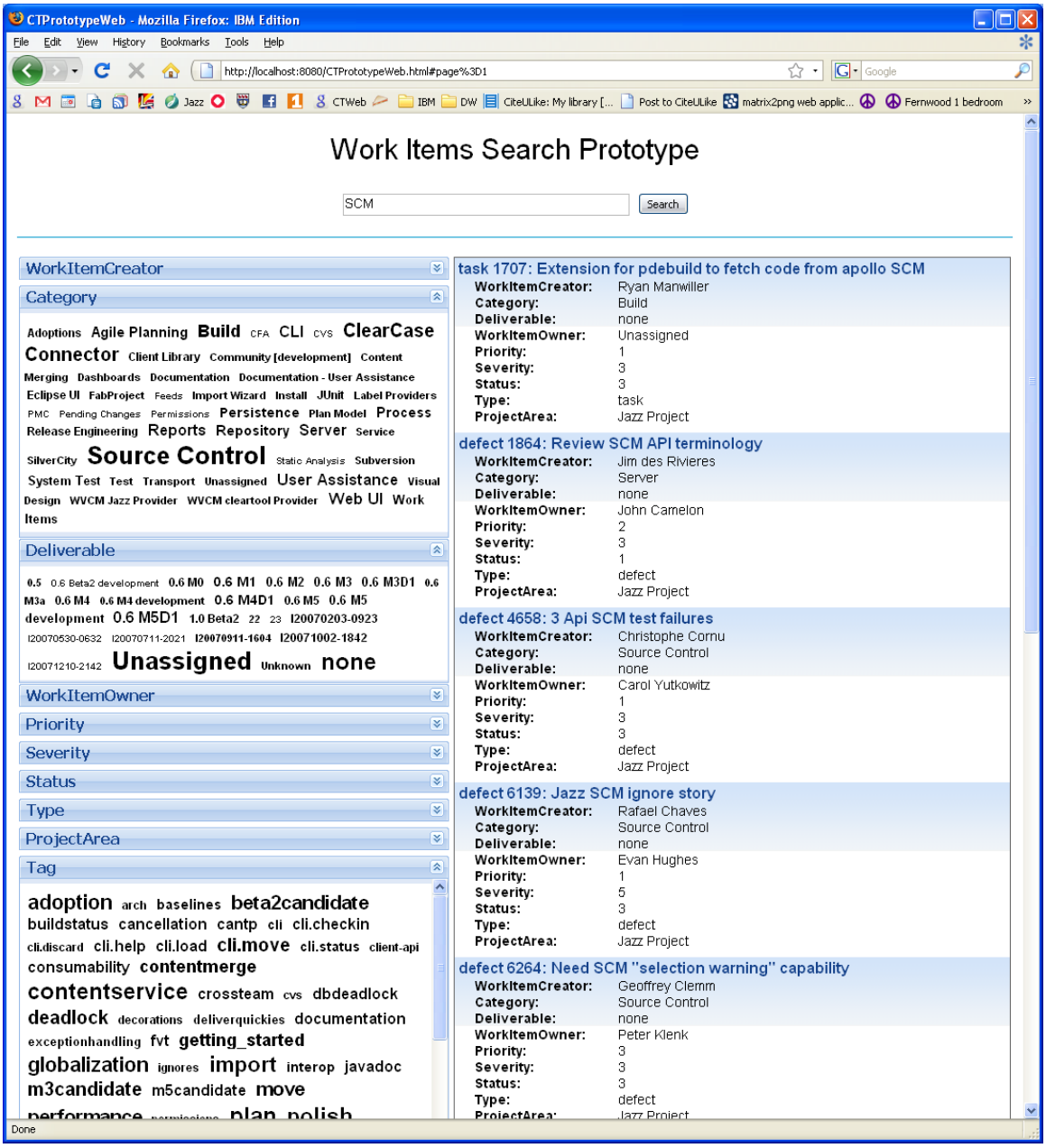

Figure 10.1. *Multiple Tag Clouds for Jazz Work Items*

10.3 Limitations

My investigations adopted a qualitative approach because of the exploratory nature of my research questions. I chose data-collection methods that would allow me to identify the users' needs and the tasks they wish to accomplish, and to determine how users think a visualization approach might help them achieve their goals. However, any research design is subject to certain limitations. The limitations of my investigations are discussed below.

Limited number of participants: The number of participants was limited, especially in the interviews and in the evaluation of CTeXplorer. This is because these investigations required the participation of professional users, who are difficult to access.

Limited types of participants: My user studies on the usefulness of the two tools were also limited with regard to the types of participants: CTSearch was tested only with lay users and CTeXplorer only with experts.

Lack of observational studies: Generally speaking, my exploratory investigations relied on users' reports of their own behaviour rather than on a researcher's objective observations of that behaviour *in situ*. Although the query-log analysis partially corrected for any bias that this element of subjectivity may have introduced into the results, further observational studies would be required to control for it thoroughly.

Lack of integration between tools: The two visualization tools described in this dissertation, namely, CTSearch and CTeXplorer, are not functionally integrated. These tools were designed to address different needs and to answer different questions that arose in the course of the research. They used different datasets: CTSearch used a semi-structured dataset and was tested with a large number of trials from ClinicalTrials.gov, which includes trials treating a diversity of medical conditions. CTeXplorer used a highly structured set of trials and was tested only with a reduced set of trials, specifically, trials concerned with the prevention of mother-to-child transmission of HIV.

Limited number of visualizations tested: I designed one tool to address the problem of seeking information and one to address the problem of comparing a pre-selected set of clinical trials. In the initial phase of the design process for both tools, I made several attempts using parallel coordinates and node-based graphs; however, the preliminary feedback I received in the discussions with the two experts indicated that those approaches were unlikely to result in success. To arrive at a tool that would constitute a substantial improvement over existing approaches would require a larger number of alternative visualizations and a comparative study of their effectiveness, both against one another and against currently available instruments.

Limited use cases implemented: In the design of my visualizations I considered only use cases that were critical or whose priority level was "primary". Of the ten use cases I defined, CTSearch accommodated only three and CTeXplorer three others.

10.4 Future Work

After reflecting on the work I have done for this dissertation and the findings obtained, I realize that there is much work that could be done to enhance the value and impact of my investigations.

Alternative visualizations: I consider that my description of the problem in terms of requirements and use cases is sufficiently detailed to permit other designers to propose alternative visualizations. Development of such visualization techniques and a comparative study of them is desirable. Some of these alternatives might include the following:

- Using term-recognition techniques to add semantic content to the terms and thus provide more informative interfaces to the user; for example, an interface might provide definitions of terms, synonyms, superclasses, and subclasses.

- Designing visualizations that integrate multiple levels of granularity; for example, CTeXplorer might be integrated into CTSearch.
- Explore other tag clouds representations and different ordering in the list of query results.

Explore collaborative needs: Research is not an isolated activity; thus, a promising field of investigation is the collaborative needs of users of clinical trials. Further studies are required to explore the needs of teams of users seeking clinical-trial information and to extend the requirements and use cases to include the collaborative aspects of research. Some ways of exploring this possibility are the following:

- Making the tools available on-line. Ideally CTSearch and CTeXplorer could be integrated with current search interfaces to repositories of current clinical trials, such as ClinicalTrials.gov and the New Zealand Registry.
- Recruiting a team of participants and keeping a log of their user interactions, with a view to conducting a longitudinal study on the use of visualizations by a team of researchers.

Extend my tool designs to other datasets: To integrate CTeXplorer with CTSearch, it is necessary to design future versions of CTeXplorer with data extracted from ClinicalTrials.gov. Such data needs to be preprocessed to allow the detailed comparisons that are supported by CTeXplorer. Furthermore, extending my current designs to use other semi-structured repositories such as PubMed/MedLine would enable them to serve a larger population of users.

Expand upon my query-log analysis: To strengthen my qualitative query-log analysis, a new qualitative study could be conducted on a different dataset, for example, using a larger sample or a log from another source, such as ClincalTrials.gov.

Extend the number of use cases implemented: In my proposed visualizations I only addressed "critical" use cases and did not consider secondary ones. Thus, the latter

remain to be examined.

10.5 Conclusions

Users of clinical-trial information are currently struggling to satisfy their information needs. Much can be done to improve existing information-seeking interfaces that offer a single box to enter keywords. Improvements must be based on the users' needs and wants rather than on the imagination and good intentions of human-computer interaction designers. In this research I have proposed a set of features, derived from exploratory studies with real users, that should be considered in the design of better user interfaces to support the information-seeking process. It is my hope that this work will provide the basis for the development and evaluation of effective interfaces to reduce the effort users must invest to find and make sense of clinical-trial information.

References

[ACM] ACM SIGCHI Curricula for Human-Computer Interaction.

[All03] Mary Margaret Allen. Empirical Evaluation of a Visualization Tool for Knowledge Engineering. Master's thesis, University of Victoria, Victoria, Canada, 2003.

[AS94] Christopher Ahlberg and Ben Shneiderman. Visual Information Seeking: Tight Coupling of Dynamic Query Filters with Starfield Displays. In *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '94)*, pages 313–317, Boston, MA, USA, 1994. ACM.

[Bat89] M.J. Bates. The Design of Browsing and Berrypicking Techniques for the On-line Search Interface. *Online Review*, 13(5):407–424, 1989.

[BCKS04] N.L. Bennett, L.L. Casebeer, R.E. Kristofco, and S.M. Strasser. Physicians' Internet Information-Seeking Behaviors. *Journal of Continuing Education in the Health Professions*, 24(1), 2004.

[BM06] Christopher H. Brooks and Nancy Montanez. Improved Aannotation of the Blogosphere Via Autotagging and Hierarchical Clustering. In *Proceedings of the 15th International Conference on World Wide Web (WWW '06)*, pages 625–632, Edinburgh, Scotland, 2006. ACM.

[Bro02] Andrei Broder. A Taxonomy of Web Search. *SIGIR Forum*, 36(2):3–10, 2002.

[BYRN99] Ricardo Baeza-Yates and Berthier Ribeiro-Neto. *Modern Information Retrieval*. Addison Wesley, May 1999.

[BZ05] K. Bielenberg and M. Zacher. Groups in Social Software: Utilizing Tagging to Integrate Individual Contexts for Social Navigation. Master of science in digital media, University of Bremen, Bremen, Germany, 2005.

[CBK+02] L. Casebeer, N. Bennett, R. Kristofco, A. Carillo, and R. Centor. Physician Internet Medical Information Seeking and On-line Ccontinuing Education Use Patterns. *Journal of Continuing Education in the Health Professions*, 22(1):33–42, 2002.

[CFJ+00] K.W. Cogdill, C.P. Friedman, C.G. Jenkins, B.E. Mays, and M.C. Sharp. Information Needs and Information Seeking in Community Medical Education. *Academic Medicine*, 75(5):484–86, 2000.

[CHSS07] Simona Carini, Maria-Elena Hernandez, Margaret-Anne Storey, and Ida Sim. An Interactive Tool for Visualizing Design Heterogeneity in Clinical Trials. *XV Cochrane Colloquium*, 2007.

[CKPT92] Douglass R. Cutting, David R. Karger, Jan O. Pedersen, and John W. Tukey. Scatter/Gather: A Cluster-Based Approach to Browsing Large Document Collections. In *Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '92)*, pages 318–329, Copenhagen, Denmark, 1992. ACM.

[CMS99] Stuart K. Card, Jock D. Mackinlay, and Ben Shneiderman, editors. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1999.

[CRCK90] DP Connelly, EC Rich, SP Curley, and JT Kelly. Knowledge Resource Preferences of Family Physicians. *The Journal of Family Practice*, 30(3):353, 1990.

[Cre04] John W. Creswell. *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*. Prentice Hall, second edition, July 2004.

[CUM85] D.G. Covell, G.C. Uman, and P.R. Manning. Information Needs in Office Practice: Are They Being Met? *Annals of Internal Medicine*, 103(4):596, 1985.

[DFA+98] Alan J. Dix, Janet E. Finlay, Gregory D. Abowd, Russell Beale, and Janet E. Finley. *Human-Computer Interaction*. Prentice Hall, second edition, January 1998.

[Dun97] P.M. Dunn. James Lind (1716-94) of Edinburgh and the Treatment of Scurvy. *British Medical Journal*, 76(1):64, 1997.

[DVE06] R. De Vries and C. Elliott. Why Disclosure? *Journal of General Internal Medicine*, 21(9):1003–1004, 2006.

[EH05] Brian Everitt and David Howell, editors. *Clinical Trials and Intervention Studies*. ISBN: 0-470-86080-4. John Wiley and Sons, LTD., June 2005.

[EJ03] Caroline M. Eastman and Bernard J. Jansen. Coverage, Relevance, and Ranking: The Impact of Query Operators on Web Search Engine Results. *ACM Transactions on Information Systems*, 21(4):383–411, 2003.

[EOC+05] J.W. Ely, J.A. Osheroff, M.L. Chambliss, M.H. Ebell, and M.E. Rosenbaum. Answering Physicians' Clinical Questions: Obstacles and Potential Solutions, 2005.

[EOE+99] John W. Ely, Jerome A. Osheroff, Mark H. Ebell, George R. Bergus, Barcey T. Levy, Lee M. Chambliss, and Eric R. Evans. Analysis of Questions Asked by Family Doctors Regarding Patient Care. *British Medical Journal*, 319(7206):358–361, August 1999.

[EOE+02] J.W. Ely, J.A. Osheroff, M.H Ebell, M.L. Chambliss, D. C. Vinson, J. J. Stevermer, and E.A. Pifer. Obstacles to Answering Doctors' Questions about Patient Care with Evidence: Qualitative Study. *British Medical Journal*, 324(7339):710–, 2002.

[EOG+00] John W Ely, Jerome A Osheroff, Paul N Gorman, Mark H Ebell, M Lee Chambliss, Eric A Pifer, and P Zoe Stavri. A Taxonomy of Generic Clinical Questions: Classification Study. *British Medical Journal*, 321(7258):429–432, 2000.

[FF06] S. Fox and D. Fellows. Online Health Search 2006, 2006.

[FS95] Kenneth P. Fishkin and Maureen C. Stone. Enhanced dynamic queries via movable filters. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '95)*, pages 415–420, Denver, CO, USA, 1995.

[GC00] M.K. Giacomini and D.J. Cook. Users' Guides to the Medical Literature XXIII. Qualitative Rresearch in Health Care. Are the Results of the Study Valid? *Journal of the American Medical Association*, 284(3):357–362, 2000.

[GCE00] M.L. Green, M.A. Ciampi, and P.J. Ellis. Residents Medical Information Needs in Clinic: Are They Being Met? *The American Journal of Medicine*, 109(3):218–223, 2000.

[GH95] Paul N. Gorman and Mark Helfand. Information Seeking in Primary Care: How Physicians Choose Which Clinical Questions to Pursue and Which to Leave Unanswered. *Medical Decision Making*, 15(2):113–119, 1995.

[GL06] D. Gendarmi and F. Lanubile. Community-Driven Ontology Evolution Based on Folksonomies. *Lecture Notes in Computer Science*, 4277:181, 2006.

[Goo03] Kenneth W. Goodman. *Ethics and Evidence-Based Medicine: Fallibility and Responsibility in Clinical Science*. Cambridge University Press, August 2003.

[Gor01] P. Gorman. Information Needs in Primary Care: A Survey of Rural and Non-rural Primary Care Physicians. *Studies in Health Technology and Informatics*, pages 338–342, 2001.

[Gru90] LD Gruppen. Physician Information Seeking: Improving Relevance Through Research. *Bulletin of the Medical Library Association*, 78(2):165, 1990.

[HCSS08] Maria-Elena Hernandez, Simona Carini, Margaret-Anne Storey, and Ida Sim. An Interactive Tool for Visualizing Design Heterogeneity in Clinical Trials. In *Proceedings of the AMIA Symposium*, pages 298–302, November 2008.

[Hea95] Marti A. Hearst. TileBars: Visualization of Term Distribution Information in Full Text Information Access. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '95)*, pages 59–66, Denver, CO, USA, 1995.

[Hea06] Marti A. Hearst. Clustering Versus Faceted Categories for Information Exploration. *Communications of the ACM*, 49(4):59–61, 2006.

[HFS+08a] Maria-Elena Hernandez, Sean M. Falconer, Margaret-Anne Storey, Simona Carini, and Ida Sim. Multiple Synchronized Tag Clouds for Semi-Structured Documents. In *IEEE InfoVis*, Columbus, OH, USA, 2008.

[HFS+08b] Maria-Elena Hernandez, Sean M. Falconer, Margaret-Anne Storey, Simona Carini, and Ida Sim. Synchronized Tag Clouds for Exploring Semi-Structured Clinical Trial Data. In *Proceedings of the 2008 Conference of the Center for Advanced Studies on Collaborative Research (CASCON '08)*, Ontario, Canada, 2008.

[HK97] Marti A. Hearst and Chandu Karadi. Cat-a-Cone: An Interactive Interface for Specifying Searches and Viewing Retrieval Results Using a Large Category Hierarchy. In *Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '97)*, pages 246–255, Philadelphia, PA, USA, 1997. ACM.

[HMHS06] Y. Hassan-Montero and V. Herrero-Solana. Improving Tag-clouds as Visual Information Retrieval Interfaces. In *International Conference on Multidisci-*

plinary Information Sciences and Technologies (InSciT2006), pages 25–28, Merida, Spain, 2006.

[HP96] Marti A. Hearst and Jan O. Pedersen. Reexamining the Cluster Hypothesis: Scatter/Gather on Retrieval Results. In *Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '96)*, pages 76–84, Zurich, Switzerland, 1996. ACM.

[HTHB07] Jorge R. Herskovic, Len Y. Tanaka, William Hersh, and Elmer V. Bernstam. A Day in the Life of PubMed: Analysis of a Typical Day's Query Log. *Journal American Medical Informatics Association*, 14(2):212–220, 2007.

[Ins85] Alfred Inselberg. The Plane with Parallel Coordinates. *The Visual Computer*, V1(4):69–91, December 1985.

[JF97] Susanne Jul and George W. Furnas. Navigation in Electronic Worlds: CHI '97 Workshop Report. *SIGCHI Bulletin*, 29:44–49, 1997.

[Jon98] Steve Jones. Graphical Query Specification and Dynamic Result Previews for a Digital Library. In *Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology*, pages 143–151, San Francisco, California, United States, 1998. ACM.

[JP01] Bernard J. Jansen and Udo Pooch. A Review of Web Searching Studies and a Framework for Future Research. *Journal of the American Society for Information Science and Technology*, 52(3):235–246, 2001.

[JS06a] B.J. Jansen and A. Spink. How are we searching the World Wide Web? A comparison of nine search engine transaction logs. *Information Processing and Management*, 42(1):248–263, 2006.

[JS06b] B.J. Jansen and A. Spink. How are we Searching the World Wide Web? A Comparison of Nine Search Engine Transaction Logs. *Information Processing and Management*, 42(1):248–263, 2006.

[Käk05] Mika Käki. Findex: Search Result Categories Help Users When Document Ranking Fails. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*, pages 131–140, Portland, OR, USA, 2005. ACM.

[KHGW07] Byron Y-L Kuo, Thomas Hentrich, Benjamin M . Good, and Mark D. Wilkinson. Tag Clouds for Summarizing Web Search Results. In *Proceedings of the 16th International Conference on World Wide Web (WWW '07)*, pages 1203–1204, Banff, Alberta, Canada, 2007. ACM.

[KLP96] A.J. Kleiboemer, M.B. Lazear, and J.O. Pedersen. Tailoring a Retrieval System for Naive Users. In *Proceedings of the Fifth Annual Symposium on Document Analysis and Information Retrieval (SDAIR)*, Las Vegas, NV, USA, 1996.

[Lar01] C. Larman. *Applying UML and Patterns: An Introduction to Object-Oriented Analysis and Design and the Unified Process*. Prentice Hall PTR, Upper Saddle River, NJ, 2001.

[LM02] Eve-Marie Lacroix and Robert Mehnert. The US National Library of Medicine in the 21st Century: Expanding Collections, Nontraditional Formats, New audiences. *Health Information Library Journal*, 19(3):126–158, 2002.

[Mar95] Gary Marchionini. *Information Seeking in Electronic Environments*. Cambridge University Press, New York, NY, USA, March 1995.

[Mar06] Gary Marchionini. Exploratory Search: from Finding to Understanding. *Communications of the ACM*, 49(4):41–46, 2006.

[Mor06] A. Morabia. Pierre-Charles-Alexandre Louis and the Evaluation of Bloodletting. *Journal of the Royal Society of Medicine*, 99(3):158–160, 2006.

[MPB02] Vincent Baujard Melanie Provost, Matthew Perri and Celia Boyer. Survey of Health and Medical Internet Users, 2002.

[NE00] Bashar Nuseibeh and Steve Easterbrook. Requirements Engineering: A Roadmap. In *Proceedings of the Conference on the Future of Software Engineering (ICSE '00)*, pages 35–46, New York, NY, USA, 2000. ACM Press.

[NS00] Chris North and Ben Shneiderman. Snap-Together Visualization: A User Interface for Coordinating Visualizations via Relational Schemata. In *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI '00)*, pages 128–135, Palermo, Italy, 2000. ACM.

[PF00] Wanda Pratt and Lawrence Fagan. The Usefulness of Dynamically Categorizing Search Results. *Journal of the American Medical Informatics Association*, 7(6):605–617, 2000.

[Pia05] Steven Piantadosi. *Clinical trials: A Methodologic Perspective*. Wiley-Interscience, New York, NY, second edition, August 2005.

[Pir07] P. Pirolli. *Information Foraging Theory: Adaptive Interaction with Information*. Oxford University Press, USA, 2007.

[PMS+98] C. Plaisant, R. Mushlin, A. Snyder, J. Li, D. Heller, and B. Shneiderman. LifeLines: Using Visualization to Enhance Navigation and Analysis of Patient Records. In *Proceedings of the AMIA Symposium*, pages 76–80, Maryland, USA., 1998.

[Poc84] Stuart J. Pocock. *Clinical trials: A Practical Approach*. John Wiley & Sons, January 1984.

[PSHD96] Peter Pirolli, Patricia Schank, Marti Hearst, and Christine Diehl. Scatter/-Gather Browsing Communicates the Topic Structure of a Very Large Text Collection. In *Proceedings of the SIGCHI Conference on Human Factors*

in Computing Systems (CHI '96), pages 213–220, Vancouver, BC, Canada, 1996. ACM.

[Ran31] S.R. Ranganathan. *The Five Laws of Library Science*. Edward Goldston, London, UK, 1931.

[RC94] Ramana Rao and Stuart K. Card. The Table lens: Merging Graphical and Symbolic Representations in an Interactive Focus + Context Visualization for Tabular Information. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '94)*, pages 318–322, Boston, MA, USA, 1994. ACM.

[RGL87] Joel R. Remde, Louis M. Gomez, and Thomas K. Landauer. SuperBook: an Automatic Tool for Information Exploration—Hypertext? In *Proceedings of the ACM Conference on Hypertext (HYPERTEXT '87)*, pages 175–188, Chapel Hill, NC, USA, 1987. ACM.

[RLM+08] E.D. Rentz, L. Lewis, O.J. Mujica, D.B. Barr, J.G. Schier, G. Weerasekera, P. Kuklenyik, M. McGeehin, J. Osterloh, J. Wamsley, et al. Outbreak of Acute Renal Failure in Panama in 2006: A Case-Control Study. *Bulletin of the World Health Organization*, 86(10):749–756, 2008.

[RMC91] George G. Robertson, Jock D. Mackinlay, and Stuart K. Card. Cone trees: animated 3d visualizations of hierarchical information. In *CHI '91: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 189–194, New York, NY, USA, 1991. ACM.

[RMM01] H. Reiterer, G. Mußler, and T.M. Mann. A Visual Information Seeking System for Web Search. *Oberquelle H., Oppermann R., Krause J.(Hg.), Mensch & Computer*, pages 297–306, 2001.

[SBC98] Ben Shneiderman, Donald Byrd, and W. Bruce Croft. Sorting out Searching: a User-Interface Framework for Text Searches. *Communications of the ACM*, 41(4):95–98, 1998.

[Sha05] B. Shaw. Utilizing Folksonomy: Similarity Metadata from the del. icio. us System. Technical report, December 2005.

[Sha07] Nigham Shah. Ontrez Project Report. Technical report, Stanford Center for Biomedical Informatics Research, Stanford, CA, USA, 2007.

[Shi05] C. Shirky. Ontology is Overrated: Categories, Links, and Tags, 2005.

[Shn08] Ben Shneiderman. Treemaps for Space-Constrained Visualization of Hierarchies, 2008.

[SHRM99] H.R. Strasberg, P.R. Hubbs, T.C. Rindfleisch, and K.L. Melmon. Analysis of Information Needs of Users of the Stanford Health Information Network for Education. In *Proceedings of the AMIA Symposium*, page 965. American Medical Informatics Association, 1999.

[SMHM99] Craig Silverstein, Hannes Marais, Monika Henzinger, and Michael Moricz. Analysis of a Very Large Web Search Engine Query Log. *SIGIR Forum*, 33(1):6–12, 1999.

[Smi96] R. Smith. What Clinical Information Do Doctors Need? *British Medical Journal*, 313(7064):1062, 1996.

[Spe06] Donna Spencer. Four Modes of Seeking Information and How to Design for Them, March 2006.

[Spo93] Anselm Spoerri. InfoCrystal: A Visual Tool for Information Retrieval & Management. In *Proceedings of the Second International Conference on Information and Knowledge Management*, pages 11–20. ACM, 1993.

[SSI07] Forrest Shull, Janice Singer, and Dag I.K.Sjoberg. *Guide to Advanced Empirical Software Engineering*. Springer-Verlag, London, 2007.

[SWJS01] Amanda Spink, Dietmar Wolfram, B. J. Jansen, and Tefko Saracevic. Searching The Web: The Public and Their Queries. *Journal of the American Society for Information Science and Technology*, 5(3), 2001.

[SYJ+04] A. Spink, Y. Yang, J. Jansen, P. Nykanen, D.P. Lorence, S. Ozmutlu, and H.C. Ozmutlu. A Study of Medical and Health Queries to Web Search Engines. *Health Information and Libraries Journal*, 21(1):44–51, 2004.

[Tay05] Humphrey Taylor. Number of "Cyberchondriacs" U.S. Adults Who go Online for Health Information Increases to Estimated 117 Million, 2005.

[TC05] James J. Thomas and Kristin A. Cook. *Illuminating the path: The Research and Development Agenda for Visual Analytics*. National Visualization and Analytics Center, 2005.

[Tor06] Melanie Tory. Notes on Visualization Course. December 2006.

[TSWB94] Lisa Tweedie, Bob Spence, David Williams, and Ravinder Bhogal. The Attribute Explorer. In *Conference Companion on Human Factors in Computing Systems (CHI '94)*, pages 435–436, New York, NY, USA, 1994. ACM Press.

[VB96] Aravindan Veerasamy and Nicholas J. Belkin. Evaluation of a Tool for Visualization of Information Retrieval Results. In *Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '96)*, pages 85–92, Zurich, Switzerland, 1996. ACM Press.

[Vee96] Aravindan Veerasamy. Abstract Evaluation of a Tool for Visualization of Information Retrieval Results, 1996.

[VH97] Aravindan Veerasamy and Russell Heikea. Effectiveness of a Graphical Display of Retrieval Results. In *Proceedings of SIGIR97*, pages 236–245, Philadelphia, PA,USA, 1997. ACM Press.

[VHN95] Aravindan Veerasamy, Scott Hudson, and Shamkant Navathe. Querying, Navigating and Visualizing an Online Library Catalog. Technical report, Graphics, Visualization and Usability (GVU) Center at Georgia Tech, 1995.

[VW07] Thomas Vander Wal. Folksonomy Coinage and Definition. April 2007.

[War00] Colin Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2000.

[WB06] Duolao Wang and Ameet Bakhai. *Clinical Trials – A Practical Guide to Design, Analysis, and Reporting*. Remedica Publishing, March 2006.

[WKDs06] Ryen W. White, Bill Kules, Steven M. Drucker, and m.c. schraefel. Supporting Exploratory Search–Introduction. *Communications of the ACM*, 49(4):36–39, 2006.

[WTP+95] J. A. Wise, J. J. Thomas, K. Pennock, D. Lantrip, M. Pottier, A. Schur, and V. Crow. Visualizing the Non-visual: Spatial Analysis and Interaction with Information from Text Documents. In *Proceedings of the 1995 IEEE Symposium on Information Visualization (INFOVIS '95)*, page 51, Atlanta, Georgia, 1995. IEEE Computer Society.

[YS93a] Degi Young and Ben Shneiderman. A Graphical Filter/Flow Representation of Boolean Queries: A Prototype Implementation and Evaluation. *Journal of the American Society of Information Science*, 44(6):327–339, 1993.

[YS93b] Degi Young and Ben Shneiderman. A Graphical Filter/Flow Representation of Boolean Queries: A Prototype Implementation and Evaluation. *Journal of the American Society for Information Science*, 44:327–339, 1993.

[ZIT+07] D.A. Zarin, N.C. Ide, T. Tse, W.R. Harlan, J.C. West, and D.A.B. Lindberg. Issues in The Registration of Clinical Trials. *Journal of the American Medical Association*, 297(19):2112, 2007.

[ZKA+02] Q. Zeng, S. Kogan, N. Ash, RA Greenes, and AA Boxwala. Characteristics of Consumer Terminology for Health Information Retrieval. *Methods of Information in Medicine*, 41(4):289–298, 2002.

[ZZZR04] Dongming Zhang, Caroline Zambrowicz, Hong Zhou, and Nancy K. Roderer. User Information Seeking Behavior in a Medical Web Portal Environment: A Preliminary Study. *Journal of the American Society for Information Science and Technology*, 55(8):670–684, 2004.

Appendix A

Example of Notes from a Moderated Discussion

Date: August 7, 2007 Participants: Exper1, Expert2, Programmer, Researcher1, Researcher2

Agenda

1. Review of minutes of past meeting
2. Update of activities at the University of Victoria (UVic)
3. Update of activities at the University of California (San Francisco) (UCSF)
4. Action plan for August

Minutes of the meeting of May 31, 2007

1. We agreed that the target platform of CTeXplorer is Web-based. We decided that the UVic team needs to do a bit of research about the technology available and analyze the trade-offs.
2. Researcher2 is to investigate way of creating more adaptive/intelligent interfaces that can provide quick views appropriate to the identified user needs.
3. How to annotate/tag/comment the graphs is another issue that needs to be addressed for the target platform.

4. We decided to incorporate the features of CTeXplorer into the new Web-based application (we will also keep our eyes open for enhancements that can be readily included).

5. Expert1 suggested that for the scalability issue, additional visual metaphors [aside from those already used in CTeXplorer] are to be considered.

6. Expert1 and Expert2 estimate that we will have the ontology for clinical trials (OCRI) at the end of the year. We will have also trials from other domains to test our visualizations.

7. Regarding results, we will go for incorporating a view that "modestly" represents the summary of odds ratios, but this need a bit more work from Expert1 and Expert2 to confirm the minimal data they want to see in these views. Also, they will consider whether co-interventions [clinical interventions other than that which is the primary object of a trial] should be represented, and if so how.

8. Expert1 and Expert2 will also explore the needs for the ontology interface, that is, the component between the ontology of clinical trials and the suite of visualizations we are developing.

Next steps The UVic team agreed to demo the current features of CTeXplorer for the UCFS team at the next meeting.

Next Meeting It was decided that the next meeting would occur on August 17th.

Appendix B

Online Questionnaire: Request to Forward Invitation

Dear Webmaster

We are Maria-Elena Hernandez and Sean Falconer, PhD students from the Chisel Lab at the University of Victoria. We are developing visual representations of clinical trials information with the purpose to facilitate a user to find and compare this type of data. You can find examples of our visualization at http://www.thechiselgroup.org/ctexplorer. In order to improve our proposals, we are conducting a survey to identify how people search for medical experiments. Participants of this study will be asked to answer an on-line questioner which will take approximately 15 minutes. The survey is hosted in "surveymonkey" at http://www.surveymonkey.com/s.aspx?sm=If192kBwBI750_2fb6tRGpzA_3d_3d.

Would you be able and willing to participate in survey and/or post an advertisement about this study somewhere on your local mailing lists and/or newsgroups for us? We have been ethically approved by our local ethics board.

If you can do this, let us know and we'll send you the invitation message.

Kind Regards,

Maria-Elena Hernandez

PhD Student, CHISEL Lab

University of Victoria

Canada

Appendix C

Online Questionnaire: Consent Form

You are being invited to participate in an online survey: Information Needs of Clinical Trials Users that is being conducted by Maria-Elena Hernandez and Sean Falconer Ph.D. students in the department of Computer Science at the University of Victoria. This survey is part of the research Visualizations for Clinical Trials, conducted under the supervision of Dr. Margaret-Anne Storey in the CHISEL lab (http://www.thechiselgroup.org). This research is being funded by the National Center for Biomedical Ontology, under roadmap-initiative grant U54 HG004028 from the National Institutes of Health.

The main objective of this survey is to understand the information needs of clinical trials users. The knowledge gained from this survey will be used to design new user interfaces to support a user to search and browse for clinical trials information. You are being asked to participate in this survey because we need feedback of users of clinical trials information. If you have this interest, then your feedback is essential to achieve excellence in our new user interface designs. This survey will take you at least 15 minutes.

In terms of protecting your anonymity, we will not ask your name or location and we will not use methods of identifying your location when filling out the survey. You will be assigned a unique ID. The unique identifiers will be used to aggregate your participation but will in no way be used to identify you personally. Only trained researchers will be analyzing the data. These researchers are neutral third parties and are not responsible for any evaluation of your work.

All information disclosed to researchers in this project is confidential. All data will be

kept secure and protected at all times in password protected files on a secure server. Study data will be kept for three years. At the end of this time computer data files will be deleted.

Please be advised that information about you that is gathered for this research study (WILL NOT INCLUDE IDENTIFIABLE INFORMATION) will be shared with the sponsor of this research and this sponsor is located within the United States. As such, there is a possibility that information about you may be accessed without your knowledge or consent by the U.S. government in compliance with the U.S. Patriot Act.

Your participation in this research must be completely voluntary. If you do decide to participate, you may withdraw at any time without any consequences or any explanation. You may also request to withdraw previously recorded information.

For questions, please contact Maria-Elena Hernandez (maleh2@uvic.ca (250) 472-5777) or Sean Falconer (falconer.sean@gmail.com (250)472-5778). Additionally, you may verify the ethical approval of this study, or raise any concerns you might have, by contacting the Associate Vice-President, Research at the University of Victoria by telephone: 250-472-4545 or by email ethics@uvic.ca.

1. Clicking the box I agree below indicates that you understand the above conditions of participation in this survey and that you have had the opportunity to have your questions answered by the researchers.

2. IN SOME CASES THE RESEARCHERS OF THIS STUDY MAY WANT TO QUOTE ANONYMOUSLY YOUR OPINIONS, IF YOU GRANT PERMISSION TO DO THIS; PLEASE CLICK THE BOX BELOW PERMISSION TO QUOTE ME OR DO NOT QUOTE ME.

Appendix D

Evaluation of CTeXplorer

D.1 Tasks to Evaluate the Usefulness of CTeXplorer

Thanks for taking the time to participate in this study and helping us improve CTeXplorer. Once we have finished explaining how the CTeXplorer tool works, please complete the following tasks. Fill in the blank spaces with your answers. If you have any questions, please ask the researcher before proceeding. If you do not wish to complete a task, you may skip it and proceed to the next in the list. Information collected through this questionnaire will only be presented in anonymous and aggregate form.

1. Could you describe the general eligibility criteria for the participants across trials?
2. What is the most common pregnancy stage of participants?
3. What is the age range of the population under study?
4. What is the range of sample size? (i.e. identify both the largest and the smallest trials)
5. Which intervention or intervention combination was tested with the greatest number of participants?
6. What are the safety outcomes measured when using Nevirapine?
7. Which trials measure " mortality" for the mother and child?

8. Is there a combination of drugs that has not been tested in one of these clinical trials?
9. Which two trials are the most similar with respect to treatment?
10. Which trial do you find to be the most unusual trial from this collection?
11. Could you describe the main similarities and differences in the Interventions during labor for this set of trials?
12. Do trials that evaluate maternal mortality as safety outcome have an eligibility criteria regarding CD4 cells?
13. Do trials that evaluate infant anemia as outcome administer any treatment to the infant?

D.2 Post-Study Questionnaire

Thanks for taking the time to participate in this study and helping us improve CTeXplorer. Below there is a list of questions regarding the tasks you already completed. If you do not wish to answer a question, you may skip it and proceed to the next in the list. Information collected though this questionnaire will only be presented in anonymous and aggregate form.

1. Are there some important options that should be included in the tool? Which ones?
2. Do you feel that the tool enables richer understanding of the set of given clinical trials? In what sense?
3. Did the tool allowed you to detect similarities and differences in the population, treatment and/or outcomes? Please give examples.
4. Is it feasible to find gaps in the information? Which ones?
5. Could you describe a new trial that may fill some gap in the existing trials?

6. Do you feel motivated to use the tool?
    - Strongly disagree
    - Somewhat disagree
    - Lightly disagree
    - Neutral
    - Lightly agree
    - Somewhat agree
    - Strongly agree
7. My experience using the tool was very satisfying
    - Strongly disagree
    - Somewhat disagree
    - Lightly disagree
    - Neutral
    - Lightly agree
    - Somewhat agree
    - Strongly agree
8. I felt confident using the tool
    - Strongly disagree
    - Somewhat disagree
    - Lightly disagree
    - Neutral
    - Lightly agree
    - Somewhat agree

- Strongly agree

9. I found the system cumbersome to use

   - Strongly disagree
   - Somewhat disagree
   - Lightly disagree
   - Neutral
   - Lightly agree
   - Somewhat agree
   - Strongly agree

10. I would imagine that most people would need to use this tool for long periods of time, in order to come up with insights from the trials.

    - Strongly disagree
    - Somewhat disagree
    - Lightly disagree
    - Neutral
    - Lightly agree
    - Somewhat agree
    - Strongly agree

11. I found the organization of data appropriate for the domain

    - Strongly disagree
    - Somewhat disagree
    - Lightly disagree
    - Neutral

- Lightly agree
- Somewhat agree
- Strongly agree

12. How would you like us to change the layouts and organization of the data to facilitate system use/understanding?
13. How can we improve the tool to help you in getting insights about a collection of trials?
14. Any other suggestions?

D.3 Focus Group Session

Thanks for taking the time to participate in this study and helping us improve CTeXplorer. We will be using the list of questions below during the focus group session to stimulate discussion on the process you currently use to explore clinical trials and your experience from using the CTeXplorer tool. Additional related questions may be asked during the session. Information collected through this focus group session will only be presented in anonymous and aggregate form. If you wish to refrain the study or focus group, you may do so at any time.

1. Describe how you currently explore information on clinical trials
2. Did you find that the tasks were representative of the kinds of tasks you would normally do when exploring clinical trials?
3. Describe your experience from using CTeXplorer
4. Describe your experience of understanding the trials that were presented through the tool.
5. Describe your confidence on the knowledge you gained from using the tool

6. Is there other information that should be included in the tool?

7. Do you think we are using the right representation of data?

8. Do you think this visualization is applicable to other clinical trial domains?

Appendix E

Visual Information Seeking

In Chapter 3, I found that users seeking information in electronic environments usually do not begin with a clear understanding of how to fulfill their information needs, but that they develop such understanding as they iterate during the seeking process. In Chapter 4, I reviewed current empirical studies which suggest that users express their needs using very short queries and that they frequently have difficulties formulating questions in the health and medical domain. In this Appendix, I discuss existing approaches to the problem of designing visual interfaces to support users seeking information in electronic environments. The design has to enable users to specify their queries, browse, and interpret the documents retrieved by the system [RMM01]. As Hearst [BYRN99] points out, this topic has received relatively little attention in the literature on information retrieval. This Appendix describes relevant work that has been done in the field of information visualization to support the information seeking process.

E.1 Definition

Information visualization (InfoVis) is a field derived from HCI that deals with the problem of depicting large amounts of information using graphical representations or metaphors. Applying InfoVis techniques to the information-seeking problem facilitates the representation of different aspects of the documents stored in a database and allows users to choose which aspects are to be depicted [RMM01]. Information visualization is defined by Ware as

"the use of interactive visual representations of abstract data to amplify cognition" [War00], where cognition is the human information processing that occurs while learning or using knowledge [CMS99]. In addition to amplifying cognition, information visualization facilitates the comprehension of data at different levels of granularity, helps users test hypotheses, and works as a medium of communication and education [Tor06]. Information-visualization systems and techniques can amplify human cognition in six ways [CMS99, TC05]:

Increased resources: Visualization systems can increase human processing resources, for example, by directly using the human visual capabilities or by reducing the amount of human working memory needed to process large amounts of information.

Reduced search: Visualization can facilitate finding information by grouping together related information and by displaying that information more compactly on the screen.

Enhanced Recognition of Patterns: Visualization helps to identify relationships and patterns within the data. As well, visualization favors recognition over recall (recognizing information in a graphical depiction is easier that remembering it).

Perceptual Inference: Visual representation of complex data allows the user to make inferences which would be unlikely to occur in a strictly textual representation.

Perceptual Monitoring: Visualization can make use of human perceptual attention capabilities for monitoring variables.

Manipulable Medium: Visualization facilitates an interactive medium that permits exploration of information.

Thus, InfoVis systems can greatly facilitate information processing in general. The question that concerns us here is how InfoVis techniques can be applied by HCI designers to the problem of information seeking and more specifically to the domain of clinical trials.

E.2 System Applications

In this section, I describe specific InfoVis systems aimed at providing cognitive support to users seeking information. They include Boolean-based interfaces, frequency-based interfaces, graph-based interfaces, tag-based interfaces, hierarchy-based interfaces, and clustering-based interfaces.

E.2.1 Boolean-based Interfaces

The initial focus of research on the use of visualization techniques in information retrieval was to improve search interfaces for the user formulating Boolean queries. Boolean queries are supported in most search engines and are useful to accomplish complex tasks such as "Find clinical trials related to liver cancer in Los Angeles where patients were not treated with chemotherapy." Examples of visual representations to support Boolean queries are tools such as 1) VQuery [Jon98] and InfoCrystal [Spo93], based on Venn diagrams, and 2) the Filter-Flow model based on a pipeline metaphor [YS93a], 3)STARS, using a block-oriented representation, and 4)the Magic Lenses interface [FS95].

While Boolean searches are a powerful tool for locating specific information, query-log analyses reveal that few users specify boolean queries, as we saw in Chapter 4. The use of Boolean operators presents a number of challenges to the lay user. InfoVis interfaces can alleviate some of these difficulties but not all of them.

Research has shown that users find Boolean queries difficult to formulate [YS93b]. There are at least two reasons for this. The first is the lack of consistency among search engines in the representation of Boolean operators [Jon98]. The second is that user must understand the logic values associated with the operators "AND", "OR", and "NOT" and the rules of parenthesis precedence; apparently, the majority of Internet users do not know Boolean logic [Jon98, BYRN99, YS93b].

There is some evidence to suggest that users are more successful at formulating Boolean queries using the graphical interfaces discussed in this section than doing so using textual

commands [Jon98, YS93a]. One may thus infer that the InfoVis interfaces are more intuitive and do not require the same degree of understanding of the logic underlying Boolean searches.

However, the use of Boolean operators can still present problems to the lay user, whatever the nature of the interface. These problems include:

- Users may receive too many results as a consequence of giving too many terms and "OR" (union) operators or they may receive too few or zero results as a consequence of using too many terms and "AND" (intersection) operators; this problem often arises because the user is not familiar with the vocabulary used in the targeted documents [BYRN99].

- Documents where synonyms or alternative terms are used could be missed. For example, given the query "graphical searching", documents that contain the term "visual seeking" instead could be missed [BYRN99].

- Typically, a Boolean interface does not provide any feedback regarding the number of times a term appears in each document (with exception of the InfoCrystal interface). Retrieved documents that rarely use the search terms are categorized as having the same importance as those in which the terms occur only a few times, or even just once [BYRN99].

- In the case of the clinical-trial domain, if users submit a query such as "liver cancer" and receive a list of results, they do not know whether the term was used in the documents to refer to the disease that was studied in the trials or whether it occurred, irrelevantly to the user, in some other context, such as a discussion of methodology, the description of the outcomes measured, etc.

E.2.2 Frequency-based Interfaces

A number of frequency-based interfaces have been designed to support the user in the evaluation stage of the information-seeking process. These typically represent query results in a more informative way than text-based list. Examples of are TileBars [Hea95] and Tkinq [VHN95] which represent the frequency of appearance of the submitted terms in the retrieved set of documents. For example given the query "breast cancer treatments", those tool will represent the frequency of the terms "breast", "cancer", and "treatment" in the retrieved results. Empirical evaluations of Tkinq suggest that users can make more accurate decisions about which documents to pursue using this frequency-based depicting of retrieved results versus using a traditional text-based list of results [VH97, Vee96, VB96]. On the other hand, no evaluation is reported in the effectiveness of TileBars.

The **disadvantages** of using this frequency-based interfaces to present results in a summarized format are as follows:

- There is no information on terms that are related or similar to the ones submitted in the query. For example, TileBars and Tkinq may effectively display documents that contain "liver cancer" but will not list those that contain "hepatocellular carcinoma" or "HCC", although they refer to the same medical condition.
- Frequency-based interfaces such as TileBars are useful when searching for a few specific terms but do not seem scalable to more that two or three terms.
- The use of shading to represent frequency can be misleading. For example, in TileBars one long article that uses the term "tamoxifen" 200 times would give the impression of being less relevant than a short paper that uses the same term 200 times [BYRN99].

Another type of frequency-based interfaces are tag clouds which are described in the following section.

E.2.3 Tag-based Interfaces

A recent visualization approach applicable to information-seeking is the use of tag clouds. Tag clouds were created to display and help the user navigate sets of socially generated tags and are useful to identify emerging themes. Tag Clouds are visual representations of keywords, or tags (cf Figure E.1). Tags are metadata assigned to describe text-based documents and other sources of information [GL06]. Metadata is sometimes referred to as data that describes other data. Tags typically consist of single keywords that have been previously chosen by lay users to help them retrieve the tagged information. A tag cloud is a set of keywords displayed in loose association in the shape of a cloud. Font size, weight, and color may be used to represent tag characteristics such as frequency and age. Each tag is a hyperlink to a collection of the resources tagged with that keyword. Collections of socially generated tags and the associated tagged resources are called *folksonomies* (a term coined from *folk* and *taxonomy* [VW07]). A variety of systems such as YouTube, Flicker, del.icio.us and CiteUlike support user-tagging of videos, pictures, websites, and academic documents.

Usually tag clouds display the keywords alphabetically arranged, but in some other cases the keywords are arranged by frequency, by semantic proximity, or simply at random. Hassan-Montero and Herrero-Solana [HMHS06] propose a representation of tags clustered by semantic similarity to represent themes. The clusters of semantically related keywords appears as row of terms, with the most frequently used terms appearing in the middle rows. The clustering reveals relationships that are hidden in the more typical alphabetically organized tag clouds; they also have the advantage of eliminating the duplication of terms. Brooks [BM06] also endorses clustering as a way of improving tag clouds, with the further suggestion that the clusters be organized hierarchically rather than linearly.

Bielenberg and Zacher [BZ05] propose a circular representation of tag clouds. The most frequent tags are located at the centre of the circle in the largest font size, with others appearing at distances corresponding to their frequency and in smaller font sizes. Shaw [Sha05] represents tag clouds in a graph-based depiction. Tags are nodes within the

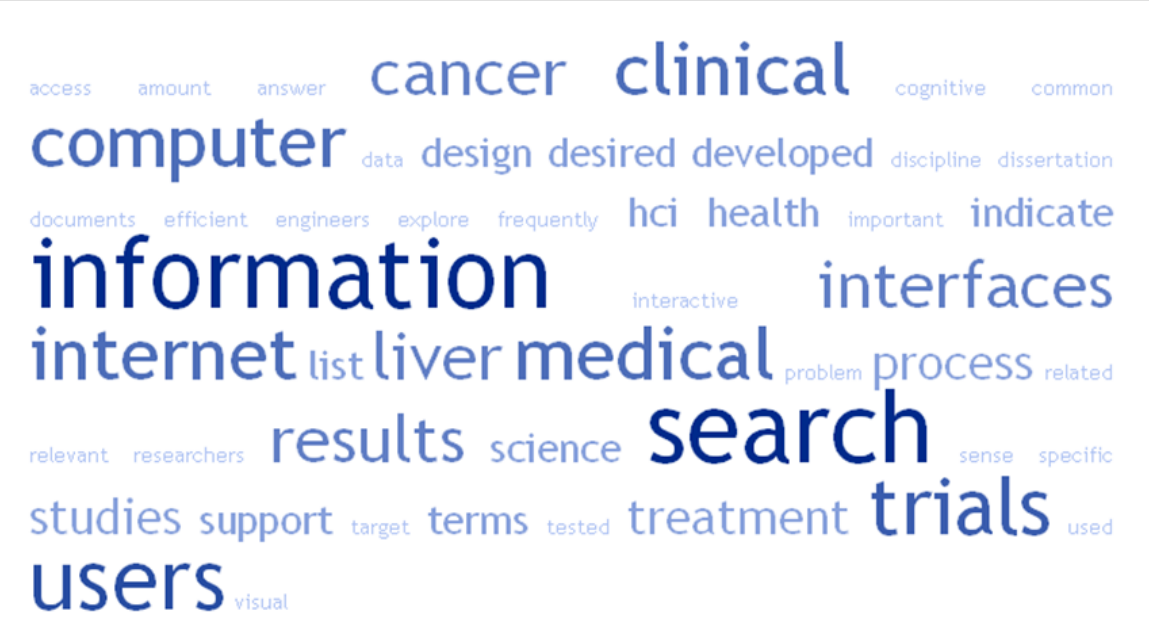

Figure E.1. *The Tag Cloud Metaphor*

graph, and edges represent semantic links between the tags.

tag clouds are useful to provide an overview of documents. Kuo *et al.* [KHGW07] designed the tool PubCloud, which uses tag clouds to summarize terms from query results using PubMed as the target database. When a user formulates a query, the system creates a tag cloud based on the frequency and age of keywords. The tool enables users to select only one tag from the cloud at a time; this selection opens a separate window with a traditional list of publications containing the selected tag. To continue the search, the user must switch back and forth between the two windows, which are depicted side by side in Figure E.2.

Tag clouds are a promising method to summarize the results of a query. As we observed in previous chapters, lay users seeking medical information are often disadvantaged by their unfamiliarity with technical vocabulary and consequently are obliged to use a time consuming process of trial and error. A user-generated tag cloud might help to shorten the information seeking process, since it will typically include technical terms, such as “hepatocellular carcinoma,” as well as the semantically related but more popular terms,

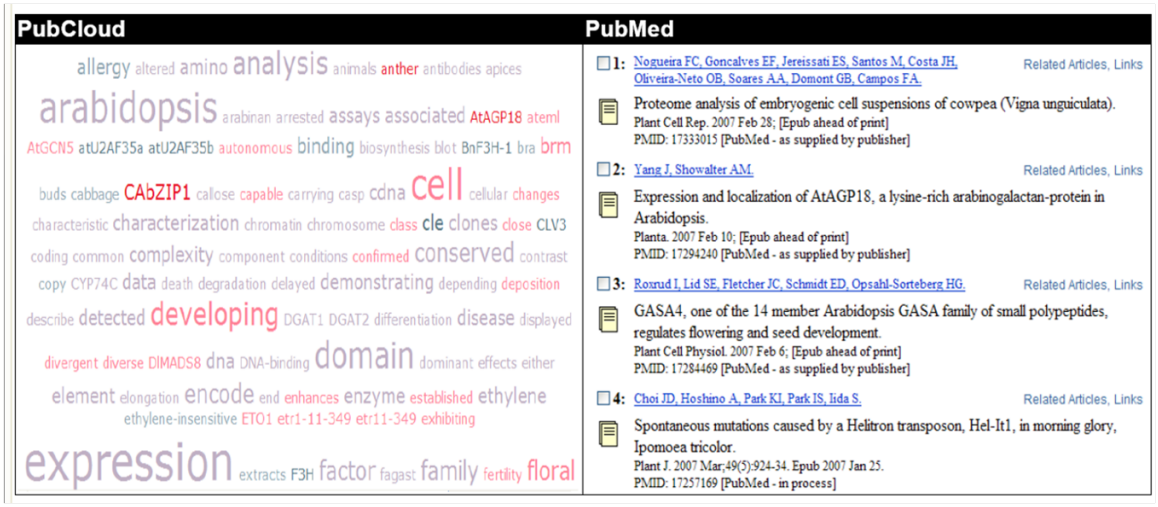

Figure E.2. *The PubCloud [KHGW07]*

such as “liver cancer.”

However, there are a number of problems for the clinical trials domain in current implementations:

- A tag cloud enables the user click on a tag from the cloud that was generated by his or her initial query to refine that query by adding an additional term. However, since the user may only add one tag at a time, refinement of any query beyond this second level is not supported.

- The tags in a cloud can only be used individually; queries based on a union of tags are not supported; in other words, the meaning of a phrase is lost. For example, querying for “hepatocellular carcinoma” (a type of liver cancer) will retrieve all the resources tagged as “hepatocellular,” “carcinoma,” or “hepatocellular carcinoma.” Consequently, “breast carcinoma” will be in the list of results even though “breast carcinoma” is not related to “hepatocellular carcinoma.”

- The list of results generated by using a tag cloud is usually organized by relevance. It does not provide any indication of the context in which query terms are used in the documents. For example, if the tag “Tamoxifen” (a cancer drug) is used in a

query, the results will include many documents that use the terms only incidentally and provide little or no information about it.

E.2.4 Hierarchy-Based Interfaces

Hierarchy-based interfaces are used to formulate and reformulate queries. They require the user to explore a hierarchical taxonomy of concepts to find terms to formulate a query. The advantage of a hierarchy-based interface is that it favours recognition over recall. In other words, it enables users to locate terms, for example, technical biomedical terms that are not part of their active vocabulary. They do not need to know how to spell the terms; they simply select them from the taxonomy. This type of search interface can be particularly useful for exploring semi-structured documents, that is, documents categorized according to a controlled vocabulary, as is the case in clinical trials. Examples of hierarchy-based interfaces are Ontrez [Sha07], Cat-a-Cone [HK97], and faceted search [Hea06].

Ontrez: Ontrez is a hierarchy-based interface which can be used to find clinical-trial information. To formulate a query Ontrez presents a hierarchical structure displaying biomedical terms incorporating the NCI Thesaurus. A user browses through and selects a term within the hierarchically organized thesaurus, and a list of associated sources of information is displayed, including *Gene Expression Omibus, ArrayExpress, PubMed,* and *ClinicalTrials.gov*. Ontrez is an ontology-based search service targeting data discovery for researchers with ontology expertise. Ontologies are hierarchical classification systems with rich semantics and rules associated with the terms they include. Ontrez extracts ontological terms from the user query and expands the query with related ontological terms (superclasses and subclasses) from controlled vocabularies. The query results are presented as lists grouped by source; for example, a search for “hepatocellular carcinoma” reports almost 700 trials just from ClinicalTrials.gov (Figure E.3).

Cat-a-Cone: Cat-a-Cone is an appealing example of a visualization approach that uses

a hierarchy-based interface to support query formulation and reformulation. Cat-a-Cone represents the hierarchy of terms as a three-dimensional tree from which multiple terms can be iteratively selected; the selection refines a query by making the tree rotate and re-adjust to facilitate the visibility of the newly selected terms (Figure E.4). Cat-a-cone is based in the cone-trees metaphor [RMC91]. Activated on demand, a window displays query results using the SuperBook [RGL87] metaphor, where each entry occupies facing pages in a virtual representation of a book. Cat-a-Cone was used to search and display results from MedLine. In Figure E.4, Cat-A-Cone uses the first page to display the title of the document and associated keywords and the second page to display the abstract.

Faceted search: Faceted search is another example of a hierarchy-based approach to support query formulation and reformulation. Faceted searches make use of a set of predefined categories. The concept of faceted classification is not new; Ranganathan proposed the model in 1931, basing it on a set of hierarchical categories, where each hierarchy represents a "facet" or aspect of the data [Ran31]. A faceted search combines browsing a taxonomy of terms with querying, using the traditional search box. Browsing and selecting terms from the taxonomy refines the search results. Studies have shown that users prefer faceted-search interfaces to a single search box interface [Hea06, Käk05].

One example of a faceted-search engine is Flamenco (see Figure E.5), which generates titles from biomedical journals using 16 facets, such as MedicalSpeciality, BiologicalScience, LifeScience, and ChemicalScience, each of which contains a list of subcategories. The facet MedicalSpeciality includes subcategories as anesthesiology, dental medicine, and dermatology. If the user searches for "cancer," the MedicalSpeciality facet shows, for example, nine results within the Medicine subcategory and two within the Oncology subcategory. Some of the results from Oncology also occur under Medicine. However, that fact is not evident to the user [HFS+08b].

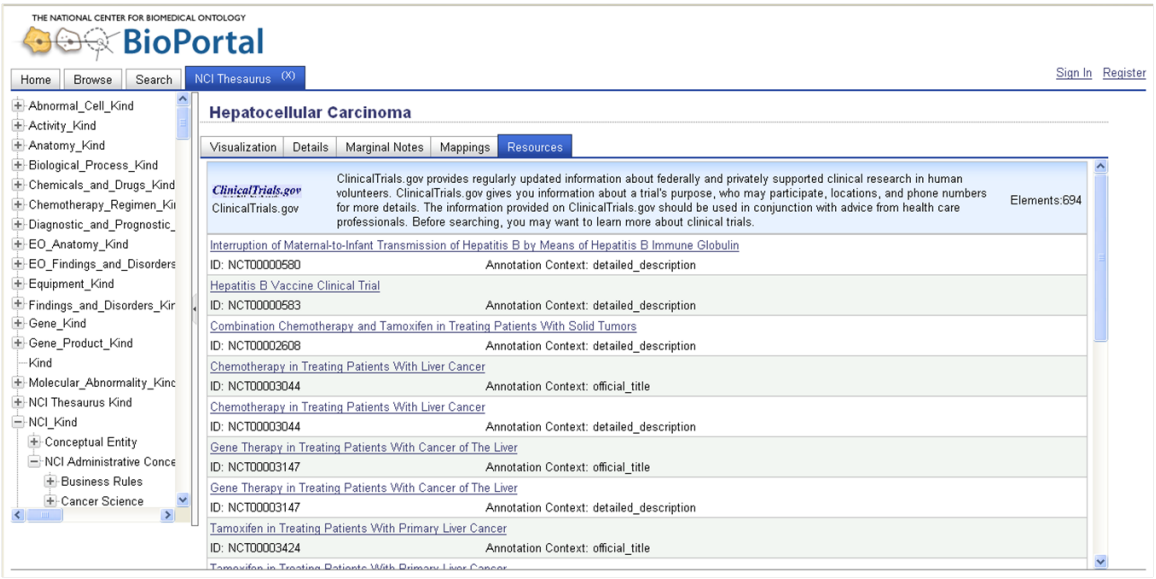

Figure E.3. *Ontrez output view [Sha07]*

The **disadvantages** that a hierarchy-based interface presents for information seeking are the following:

- To locate the terms they need, users must navigate the hierarchy. This implies some grasp of the structure of the taxonomy, which may exceed the knowledge of lay users.
- Sometimes, if users need to refine the query, they have to start again by browsing the hierarchy and choosing a new term to generate a new list of documents.
- Most of the existing hierarchy-based interfaces do not provide additional information on the semantics of the terms; for example, a user has no way of knowing the definition of a term that occurs in the taxonomy, its synonyms, or which of these synonyms is preferred in the literature. For example, a user may be searching for clinical trials on “liver cancer” but the taxonomy may use “hepatocellular carcinoma” instead. If the user does not recognize this synonym, he or she will not succeed in generating the desired query.
- Formulating queries that use more than one term at the same level of the hierarchy

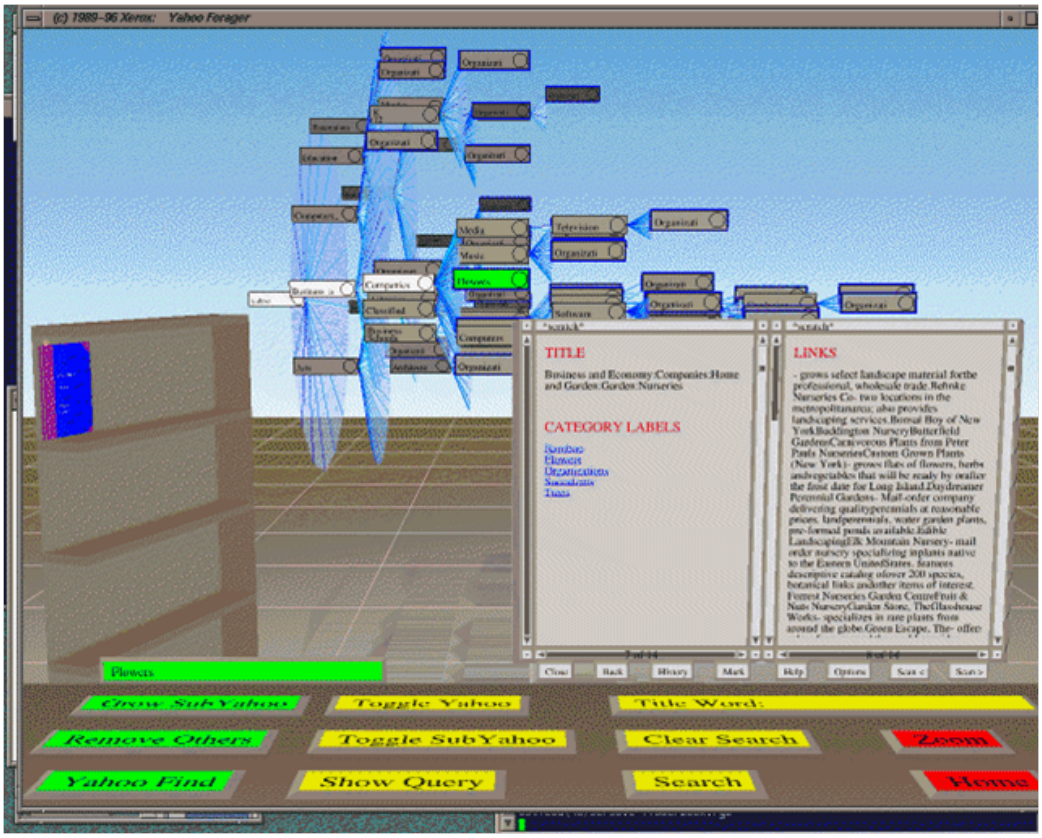

Figure E.4. *Cat-a-Cone [HK97]*

is not possible; for example, the question "what is the treatment of choice for liver cancer?" is practically impossible to answer using a hierarchy-based interface, because such a query uses information from "treatment" (or "interventions") and from "medical conditions", which are located on parallel branches of the taxonomy.

- The interfaces only allow access to highly structured documents, which are not always available.
- Although these interfaces reveal the hierarchical relationships between terms, they conceal other important relationships which are not hierarchical.

In general, hierarchy-based interfaces emphasize the taxonomical structure of the vocabulary. However, users are interested in locating documents, not in navigating taxonomies. This preference is evident, for example, in the development of the search engine

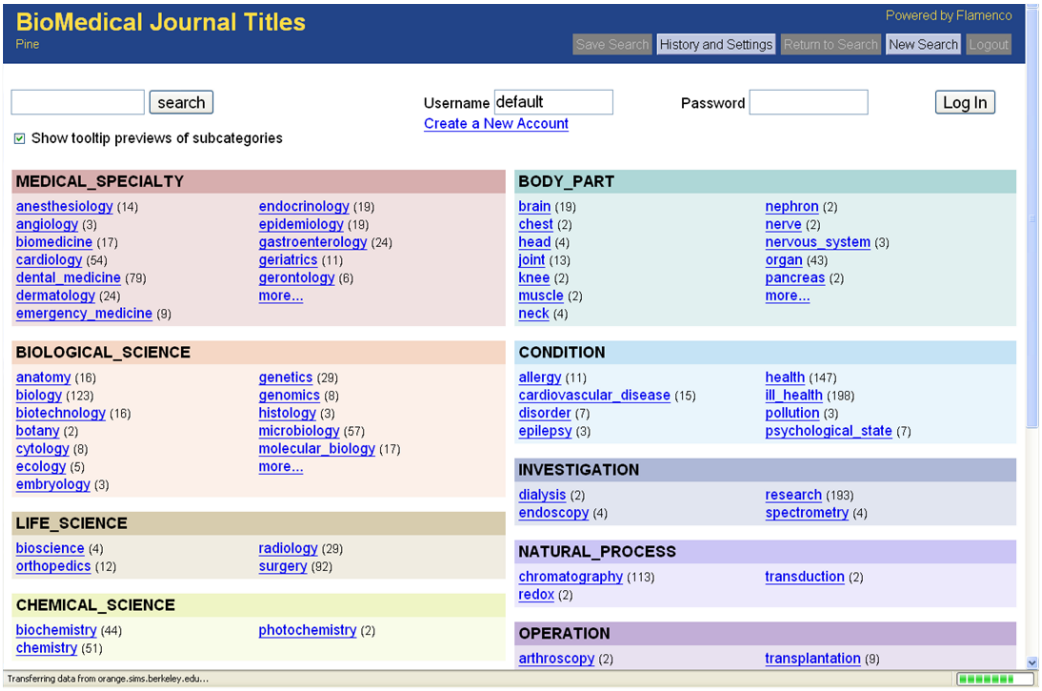

Figure E.5. *Faceted search [Hea06]*

Yahoo!, which originally used a hierarchy-based interface. In this form, Yahoo! was unable to compete with the simple search box used in Google, MSN, or AOL. This suggests that users do not like to browse categories to find the terms they need and prefer to try their own vocabulary and learn along the way [Shi05].

E.2.5 Clustering-Based Interfaces

An alternative to hierarchical-based interfaces are clustering-based interfaces, which, instead of depicting a predefined hierarchy, depict on-the-fly discovered themes and similarities. Clustering is an approach that helps users evaluate retrieved documents by grouping them thematically rather than as a linear list of results. It creates multiple clusters of documents, in which all the documents share one or more common values. Clusters are represented by thematic lists or by nodes in a graph-based interface.

The clusters are sometimes organized as a hierarchy of themes from broad themes to

more specific and depicted as hierarchical trees.

Clustering provides an overview of query results, filters out duplicates, and might be useful in revealing unexpected trends in the set of retrieved documents. Examples of clustering-based interfaces are *Scatter/Gather* [CKPT92], *ThemeScape* [WTP+95], *Clusty*[1], *Exalead*[2], *Grokker*[3], *KartOO*[4], and *Ujiko*[5].

Scatter/Gather: A classic example of clustering is the Scatter/Gather tool, which groups retrieved results in a series of lists rather than a single list. Each list clusters documents according to related keywords that they contain. Clusters are displayed in scrollable panes in textual form; each cluster displays the titles of retrieved documents. Users can discard uninteresting clusters, which causes the remaining clusters to be reorganized. Empirical evidence suggests that, by contextualizing the results of a query according to keywords, this approach is effective in enabling users to isolate relevant documents [HP96, PSHD96].

ThemeScape: The ThemeScape was originally proposed by Wise et al. [WTP+95] and is currently owned by Cartia Inc. The interface uses a 3D landscape metaphor to represent the retrieved results clustered by themes, see Figure E.7. Shading is used, as on a map, to represent elevation, which in turn represents a set of documents clustered thematically according to the frequency of occurrence of certain keywords within them. Thus, predominant themes are visually represented as high mountains, whose peaks are snow-covered for easy identification, while minor themes are depicted as smaller mountains or hills. All such elevated points in the landscape are identified by their corresponding keywords and are distributed in the landscape in such a manner as to represent the relationships between them. For example, adjacent hills or mountains share more keywords than more distant ones.

[1] http://www.clusty.com
[2] http://www.exalead.ca/search
[3] http://www.grokker.com
[4] http://www.kartoo.com
[5] http://www.ujiko.com/

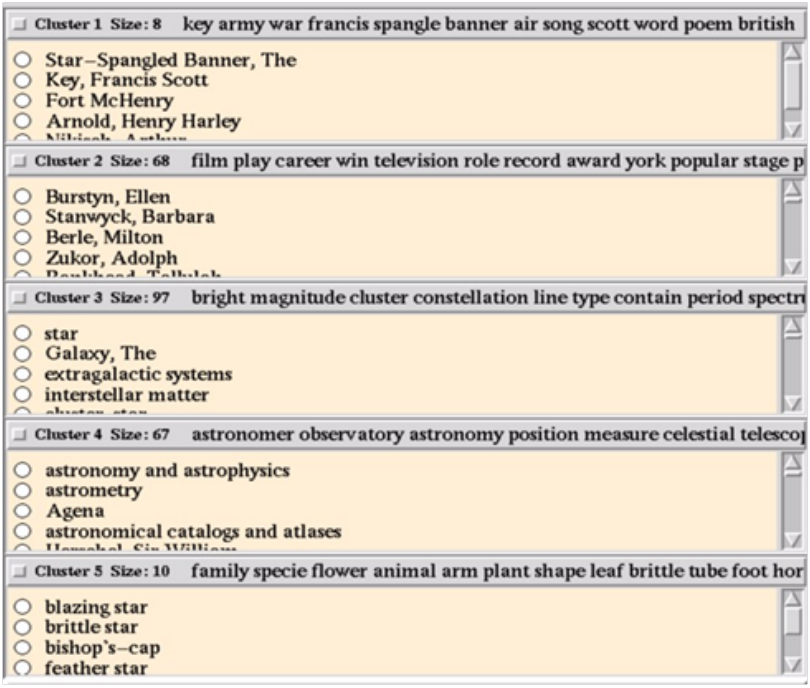

Figure E.6. *Scatter/Gather's output [CKPT92]*

Figure E.7. *ThemeScape's output*

Clusty, Exalead, and Grokker: These are three general-purpose search engines, which display the results of a query in the form of a traditional list of documents, while using a sidebar to present a list of clusters in which the same documents are grouped according to keywords that occur within them. Users can navigate through the clusters using a expansible/collapsible tree hierarchy(see Figure E.8).

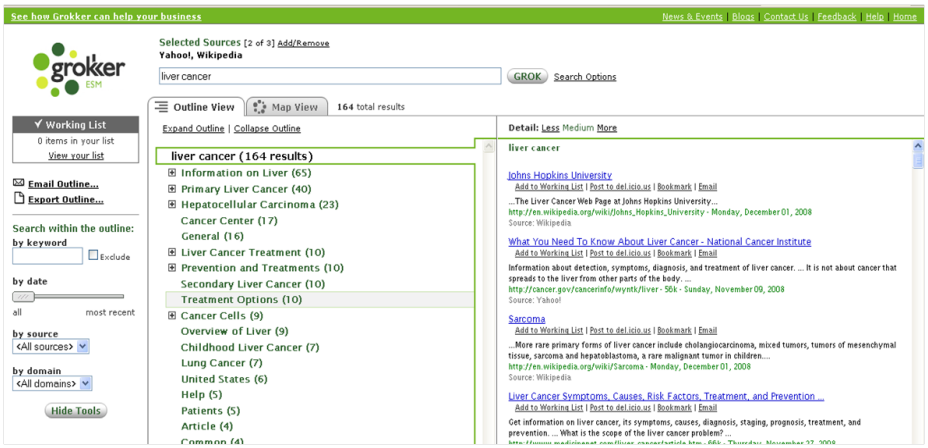

Figure E.8. *Grokker's output for the query "liver cancer"*

KartOO and Ujiko KartOO and Ujiko are a general-purpose search engines. Similarly to ThemeScape, KartOO represents the retrieved results by using shading to depict them in the form of a coloured 3D shadow. The results are displayed on the screen in terms of their semantic proximity, which is also represented by keywords whose font size corresponds to the degree of shading. Near the shadowed areas are icons of individual webpages. Figure E.9 displays the results for the query "liver cancer". Hovering one's cursor on a keyword, for example, *treatment*, reveals yellow lines connecting the term to relevant websites: <www.liverfoundation.org>, <www.livercancertreatment.org>, <www.answers.com>, <www.medicinenet.com>, <cancer.gov>, and <www.mayoclinic.com>. Clicking on a term refines the query and updates the list of results. On the other hand, hovering on the icon representing a retrieved webpage reveals a thumbnail representation of the page in the left sidebar, along with a short summary of its contents. When the cursor is not hovered over an

item in the results, the sidebar displays a list of clusters or related topics similar to those used in Clusty, Exalead, and Grokker.

The Ujiko graphical interface takes the form of a metallic oval with a large black circle in the middle, vaguely reminiscent of a personal electronic device see Figure E.10. When the user types in a query, the black circle displays a list of keywords in differently coloured fonts, representing the results as clustered thematically. The list is associated with bars, shaped like segments of concentric circles, whose colours correspond to the keywords. A traditional list of results appears in the oval surrounding the black circle and is organized thematically in association with the coloured bars inside the circle. As the user hovers the curser over a keyword, the corresponding bar and the retrieved results adjacent to it are highlighted. When the user hovers on an item in the list of results, the keywords and bars disappear from the black circle and are replaced by a thumbnail representation of the webpage and a short description of its contents. Users can customize their search by removing unwanted results or marking others as favorites; Ujiko retains these preferences during subsequent searches.

As in the case of previously considered types of interfaces, those that employ clustering have certain **disadvantages** [Hea06, BYRN99, PF00]:

- Although clustering provides cognitive support by grouping information together, this support is not optimal because clustering is a non-deterministic technique whose results are unpredictable; every time the same query is submitted, although the same results will be retrieved, they may be clustered differently.

- Just as there are multiple ways of clustering a set of documents, so there are different lists of terms that may be used to represent the clusters. In any given search, terms that may be useful to the user could be discarded by the algorithm that generates the cluster.

Figure E.9. *KartOO's output for the query "liver cancer"*

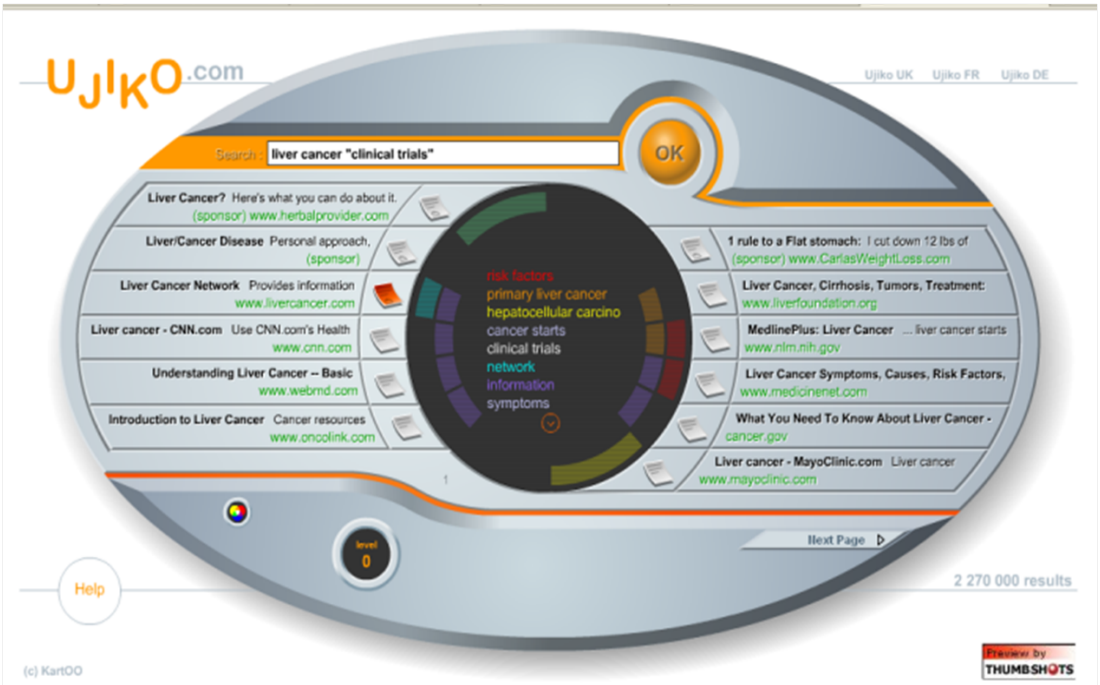

Figure E.10. *Ujiko's output for the query "liver cancer"*

- It is difficult to generate appropriate names for the clusters, which are consequently given uninformative headings such as "Cluster 1," "Cluster 2," etc.
- While some interfaces organize clusters hierarchically to make their representation more user-friendly, the sub-clusters represented in the hierarchy are subject to the same problems as are any clusters, and the terms used to identify them may be relatively uninformative or non-intuitive.

Studies show that users prefer hierarchical-based interfaces that do not employ clustering over clustering-base interfaces, because the latter use non-intuitive categories and arbitrary levels of granularity [PF00]. Also, research suggests that graphical representations of clustering are less effective than plain-text ones [KLP96].

E.3 Summary and Discussion

The problem of searching and comparing clinical-trial information has received scant attention in the computer-science literature. Users of such information need tools that help them to search or browse relevant databases, to refine their queries on subsequent iterations, and to compare the results. Existing tools support one or another of these needs more or less effectively, but none provides an overall solution for the process of seeking clinical-trial information. As we have seen, typical interfaces require a heavy cognitive effort from a user to satisfy his or her information need: that is, to retrieve relevant results, to differentiate interesting from non-interesting results, and to compare selected results. There are a number of problems with existing approaches: (1) alternative, semantically related terms are not provided (except in the tag clouds approach); (2) users are frequently obliged to navigate the controlled vocabularies used to describe the documents; (3) users find complicated interfaces difficult to manipulate; (4) users cannot determine whether query terms appear in the desired context within the retrieved documents (for example, whether "liver cancer" is the disease studied in the trial or is only mentioned in the discussion of methods,

eligibility criteria, or results); and (5) each of the visualization systems described above is optimized to support either searching, browsing, or comparison of results, rather than the entire process of clinical-trial information seeking. Thus, what remains to be developed is a tool that can support all three of those activities effectively.