A Model of Scientists' Information Seeking and a User-Interface Design

Tamar Sadeh

A Thesis Submitted in Fulfilment of the Requirements for the Degree of Doctor of Philosophy

City University London
School of Informatics
Information Science

Supervisors

Dr. Andrew MacFarlane
Prof. Stephen Robertson

December 2010

Tamar Sadeh, Cape Cod, November 2008

The real voyage of discovery consists not in seeking new landscapes, but in having new eyes.

Marcel Proust

Contents

List of Figures

Acknowledgements

I wish to thank Dr. Andrew MacFarlane and Professor Stephen Robertson, my supervisors, for their support, encouragement, and invaluable suggestions throughout my PhD studies. I especially appreciate the wealth of knowledge in the area of information seeking and retrieval that they shared with me and their willingness to support me from afar. A special thank-you goes to Dr. MacFarlane for his interest and attentiveness and for navigating all the administrative hurdles on my behalf.

I extend my sincere thanks to the high-energy physics (HEP) information specialists—Anne Gentil-Beccot, Annette Holtkamp, Jens Vigen, and Salvatore Mele from CERN; Travis Brooks from SLAC; and Heath O'Connell from Fermilab—who shared with me the raw results of the HEP survey and their vision of the INSPIRE information system and who provided me with constructive comments on the manuscript of my first publication on the information-seeking practices of the high-energy physics community. I also wish to thank this team for inviting me to their meetings and giving me the opportunity to play a modest role in the development of the new HEP information system.

To Professor Yosef Nir, dean of the physics faculty at the Weizmann Institute of Science, who introduced me to the high-energy physics arena and, through a long series of conversations and e-mail exchanges, helped me understand the special character of the scientists and their information-seeking behaviour, I am extremely grateful. My thanks go also to the other HEP scientists who shared with me their perceptions and practices related to information seeking.

I thank Matti Shem Tov, president and CEO of Ex Libris, and Nancy Dushkin, corporate vice president, who supported my wish to study, facilitated the time that I needed to achieve my goal, and showed their appreciation of my efforts.

Special thanks go to Professor Ady Stern, a dear friend, whose optimism, warm support, and excellent scientific advice helped me overcome many obstacles along the way.

I thank Ricka S. Rak, a great friend and mentor, who was there for me all along with her wise advice and who over the years has enthusiastically shared with me her many tips of the craft of writing.

I am deeply and forever indebted to my beloved parents, Miriam and Prof. Yochanan Ramon, whose wide horizons, endless curiosity, and striving for excellence have been my guiding light.

Very special thanks go to my beloved husband, Kai, and sons, Jonathan and Michael, who were supportive and proud of me throughout my studies and were willing to sacrifice time with me so that I could devote my energies to the thesis.

Last but not least, I'd like to thank my lifelong support group—my sister, Noga Carmel, and my friends Jenny Walker, Dr. Yael Brahms, Dr. Lilach Lerman, and Renat Wegrzyn—each one accomplished in her own area, who believed in me and encouraged me all along.

Abstract

Information systems that are available today do not optimally address the information-seeking behaviour of scholars, particularly those who belong to scientific communities; as a result, scholarly discovery is often cumbersome and incomplete. The hypothesis of this study is that an information-seeking system that is designed to address the nature of scholarly materials and the information-seeking behaviour of scholars, particularly the members of one scientific community, will increase the effectiveness of the scholars' searches and enable them to find and obtain relevant materials with greater ease and precision than current practices do.

The information-seeking behaviour and search practices deployed by high-energy physics (HEP) researchers are explored through a series of interviews and observations. More than 2,100 responses obtained from a HEP survey are also examined; in particular, the participants' open-ended responses are analysed. On the basis of qualitative and quantitative research regarding the characteristics of HEP scientists and their information-seeking practices, a set of six personas, representing typical members of the HEP community, is constructed.

An original model is developed that leverages existing models of information behaviour, information seeking, and information searching and reflects the full spectrum of active information-seeking and information-searching practices of HEP scholars and the nature of the data that these researchers seek. The model is then evaluated by means of seven scenarios involving the personas constructed earlier.

On the basis of the information-seeking model, a software user interface is designed as the future interface for the HEP INSPIRE information system. The user-interface design is corroborated through the model, and the personas are used to evaluate the design. Methods are suggested for long-term quantitative and qualitative monitoring of the ways in which this design supports HEP researchers. It is argued that the proposed user interface, which provides an information environment that accommodates the information-seeking practices of the HEP community in a friendly and efficient manner, will support HEP academic

research—and research of other scholarly communities that share some of the HEP community's characteristics—by shortening the search process and improving the findability of quality materials.

This thesis contributes to the body of information-science knowledge in the novel modelling of information-seeking behaviour of a well-defined scientific community, the use of personas for the modelling, and the concretization of the model into a new user-interface design.

Chapter 1 Introduction

Scholarly materials have always been the basis for academic research, and in the past, a major differentiator of research institutions was the size and quality of their library's physical collections. The electronic era, and later the invention of the Internet, introduced a new reality, where physical limitations of location and space were eliminated and access to materials could be granted from anywhere and at any time.

The first information systems to be provided electronically were library catalogues, enabling users to search for carefully described physical materials; yet the provision of the materials themselves—books, manuscripts, maps, music scores, audio and video recordings, and more—was still physical. Hence, users were limited to a library that they could access in person, or else they had to wait for materials that were brought to their library from other locations.

Providers of abstracting and indexing collections started digitizing their collections in the mid-1980s, first offering the digitized versions to users on physical devices that were available only at a library and, more recently, providing access to the collections through the Internet. Such materials are typically licensed and are thus available only to users affiliated with institutions that can afford to pay a subscription fee.

The next step was the electronic publishing of the materials themselves—first articles, and later, conference proceedings, reports, books, and other scholarly materials. Today, most articles and many books are published electronically; in some disciplines (primarily science, technology, and medicine), almost all recent research materials are available electronically, and researchers tend to seek out these electronic versions. However, the majority of the materials are licensed, and hence access is regulated.

Theoretically, users can access all electronic materials no matter where the users are or what day of the week or time of day it is. However, in practice, users face several obstacles when they seek scholarly materials. The first, already mentioned, is the licensing issue: not only is access limited to materials licensed by the researcher's institution, but gaining access is not always a straightforward process, especially when the researcher is not physically present at the institution. Furthermore, because materials are typically available from several providers and several Web sites, hypertext links that facilitate the access to such materials do not always link users to the copy of the material that is licensed by their institution (referred to in the library world as the 'appropriate' copy). Various technologies address these challenges, including proxy servers and OpenURL link resolvers.[1]

The actual locating of the desired materials is the other—and even greater—challenge for users. A variety of information systems is available today: some are very general, all-purpose systems, such as Web search engines; others are general academic systems, such as library catalogues, multidisciplinary databases, and emerging discovery services offering materials in local library collections as well as electronically published materials; and still others are very specific, targeted systems, such as subject-specific databases. As a result, a researcher is confronted with various user interfaces, search mechanisms, and levels of quality of the retrieved materials. Furthermore, the exponential growth of the number of publications poses the challenge of how to review a long list of search results that all have some relevance to a specific topic.

Surveys monitoring the changing behaviour of library users show a clear tendency towards the use of Web search engines and other Web resources instead of library catalogues and other library resources (OCLC 2005, 2006, 2007; Markey 2007a; CIBER/UCL 2008; Connaway and Dickey 2010). The publication of the OCLC Online Computer Library Center survey *Perceptions of Libraries and Information Resources: A Report to the OCLC Membership* in late 2005 (OCLC 2005), followed by a report focusing only on college students who participated in that survey (OCLC 2006), confirms what libraries already knew: that the behaviour of library users, particularly students, has changed since the

[1] An OpenURL-compliant link resolver is a software service component configured by the user's library. The link resolver generates context-sensitive services—such as a link to the 'appropriate' copy of the full text or, if the full text is not available to the user, to a document-delivery form. Such links are based on the metadata of the items in which the user indicated an interest. The OpenURL standard is a protocol for interoperability between an information system and a service component (a link resolver) (Van de Sompel and Beit-Arie 2001).

emergence of Web search engines, online bookstores, and other Web-based services that enable many users to satisfy their information needs without the help of the library. According to the 2006 OCLC report, 89% of the undergraduate and graduate students surveyed start their search for information with Web search engines, and only 2% start at their library Web site (OCLC 2006). A later OCLC survey, published in 2007, reveals that while respondents had taken advantage of Web services such as Web search engines, blogs, and online bookstores even more intensively in the two years since the 2005 survey, one Web service that was included in the survey exhibited *decreased* usage during that period—library Web sites (OCLC 2007).

Other statistical data support this trend, even though the picture portrayed by such data is not as grave. For example, statistics compiled by the Association of Research Libraries (ARL), whose membership spans North America, indicate a significant decrease in reference transactions (33%) and circulation transactions (6%) at ARL institutions between 2000 and 2006, despite an increase of about 12% in the number of students enrolled in the member institutions at that time (see http://www.arl.org/stats/annualsurveys/arlstats/statxls.shtml).

The findings described thus far address the behaviour of the general body of library users. When the behaviour of researchers is examined, the findings are slightly different (Brown 1999; Murphy 2003; RIN 2006; Hemminger et al. 2007; Jamali and Nicholas 2008; Gentil-Beccot et al. 2008; Haines et al. 2010). However, it is clear that even in research communities, users are drawn to the simplicity and ease of use of Web search engines. Because most scholarly materials are discoverable through multiple interfaces, users may well be able to obtain the same materials through Web search engines and academic systems—without facing the challenges posed by the latter. However, Web search engines come with their own drawbacks, particularly the limited searching and filtering options available to users and a search scope that comprises a universe of materials of unequal quality. Furthermore, library-driven services such as bibliographic tools and citation analyses are not available through Web search engines. Hence, most researchers rely on more than one type of information system and typically use both library-oriented systems and Web search engines. This dual information-seeking behaviour that characterises the academic community is of great interest to designers of information systems.

1.1 Motivation

Overwhelmed by complex human-machine interfaces, users today are drawn to software solutions that require little expertise and minimal effort. However, as academic users are surely aware, such systems may not be as reliable and trustworthy as systems that libraries and information providers offer. Furthermore, systems that are simple to use do not provide the means for users to conduct sophisticated searches and discover the exact materials that are needed; hence, users are often frustrated and end up settling for results that are just ‘good enough’. This thesis addresses the challenge of providing scholars with systems that are easy to use and yet offer tools that help the scholars find and obtain the materials that they need.

An information-seeking model that is defined on the basis of the practices and expectations of today’s users, as is the model that is proposed in this study, should inspire designers of search interfaces and provide value to the academic community: a user interface like the one that the author has created according to the principles of such a model will shorten the search process, improve the findability of quality materials, and thus support academic research.

1.2 Research Hypothesis

An information-seeking system that is designed to address the nature of scholarly materials and the information-seeking behaviour of scholars, particularly the members of one scientific community, will increase the effectiveness of the scholars’ searches and enable them to easily find and obtain the precise materials that they need.

1.3 Aims and Objectives

The aim of this study is to propose an innovative model of an information-seeking process conducted by researchers. Addressing the characteristics and the needs of a specific scientific community for the first time, this model can enable designers of information systems to help the community members find and obtain relevant materials with greater ease and precision than current practices do. The study goes on to suggest an original software user interface that is based on the model and can be feasibly implemented by providers of scholarly information.

The high-energy physics (HEP) community is the target research community of the study. However, the model and the interface design derived from this model may serve other scientific communities as well.

To achieve its aim, the study adopts the following objectives:

- Examine information-seeking behaviour and search practices of HEP researchers. In particular, observe and interview researchers to ascertain the ways in which they use HEP-specific and other information systems for scholarly research.

- Examine the more than 2,000 responses obtained from a HEP survey that took place in the summer of 2007; analyse the open-ended responses; and generate a set of criteria that describe the search-related behaviour of HEP scientists.

- Develop a set of personas that represent typical members of the HEP community.

- Design a general model that describes information-seeking practices by HEP scholars, and use the personas to evaluate the model.

- Use the proposed model as the basis for designing a software user interface that can be deployed as the interface of the future HEP SPIRES information system (called INSPIRE) now under development.

- Evaluate the proposed user interface by means of the information-seeking model.

- Use the personas and follow-up interviews with HEP community members to corroborate the interface design.

- Develop a set of measurements to evaluate the effects of the new interface on the information-seeking behaviour of HEP scientists who use that interface.

1.4 Contribution

The contribution of this thesis to the body of information-science knowledge consists of the definition of an information-seeking model that, for the first time, addresses a specific scientific community that is highly focused on research and has distinct information-seeking patterns. The model is created using an innovative approach—the development of personas that represent scholars of the target community and the application of the personas in the theoretical context of the model; such personas are usually applied in practical contexts. In addition, the concretization of an information-seeking model into a new user-interface design that can be feasibly implemented provides a novel means of evaluating the theoretical model empirically and of supporting a specific scientific community with its well-defined information needs.

1.5 Case Study

This research focuses on the high-energy physics (HEP) community.

The HEP community is relatively well defined and is highly focused on scholarly communication; hence, the community is a natural candidate for this research. Innovative methods for the exchange of and access to publications have been developed by community members to address their special needs, and a team of information specialists from HEP centres has been working for more than 35 years to provide access to scholarly materials in an easy and reliable way. The team invests much effort in gathering scholarly materials and in developing communication channels.

A survey that the team launched in summer 2007 reveals the perceptions of community members about the HEP information systems. These perceptions demonstrate the importance that the community members attribute to information seeking and their reliance on community resources and tools.

The author of this thesis was intrigued by the HEP scientists' tremendous loyalty to their information systems over such a long period of time. As an outsider, the author found the search interfaces of HEP information systems—the SPIRES database and the arXiv e-print repository—rather complex and very different from interfaces that are becoming the standard for academic searches. To better understand the ‘magic’ that attracts the HEP researchers, the author decided to investigate the information-seeking behaviour of HEP community members and assist the HEP information specialists in their effort to gain a better understanding of their users and to optimally address the evolving community’s needs within the existing information framework.

The survey responses, some of which the author analysed for the HEP information specialists, along with interviews that the author conducted with community members, provided both quantitative and qualitative measures for understanding and modelling the information-seeking behaviour of the community members. In addition, the author was involved in an ongoing process to redesign the SPIRES information system’s user interface, a process that began in 2008.

Although this case study addresses the high-energy physics community, the fruits of this research may be relevant to the scholarly population as a whole, as can be inferred from other scientific communities’ adoption of the arXiv e-print service, originally a tool for HEP members. By ascertaining the commonalities between the HEP community and other scientific communities, this research can serve to create better information-seeking interfaces for scholars in other fields.

1.6 Research

Although the author originally set out to examine the information-seeking behaviour of scholars in various disciplines, she decided to focus only on the HEP community because of her greater holistic understanding of its needs, its materials and resources, and the perceptions of its members.

The research included the phases described in Figure 1.

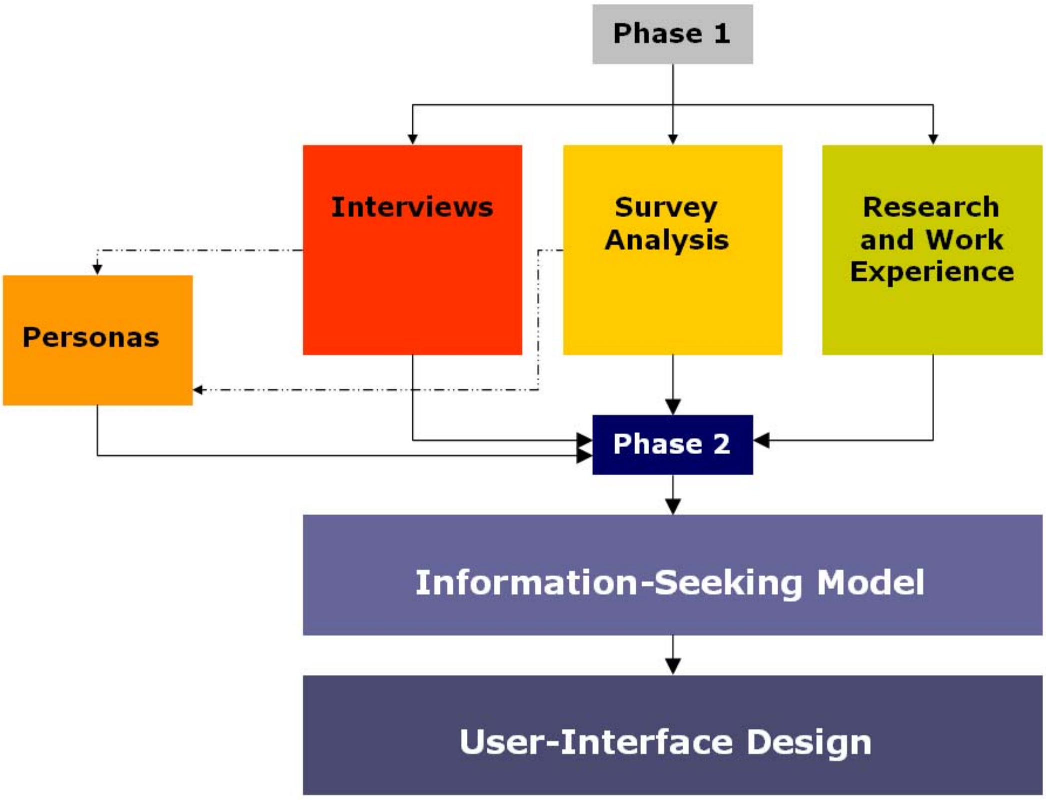


Figure 1: Process and products of this research

1.6.1 Interviews

The author interviewed seven HEP scientists ranging from PhD candidates to the dean of the physics faculty; two are experimental physicists, and five, theoretical physicists. The interviews took place in the scientists' offices at the Weizmann Institute of Science (in Rehovot, Israel). Information from the interviews helped the author understand the information-seeking behaviour of these scientists and the aspects of their daily routines that involve the seeking, gathering, and exchange of information. Numerous additional interviews were conducted as the author advanced in her research and as she modelled the information-seeking behaviour of the researchers and designed the proposed user interface.

1.6.2 Survey Analysis

The HEP information specialists provided the author of this thesis with the results—about 2,100 responses—of the survey that they had conducted in 2007 (see Chapter 3) and requested an analysis of the open-ended responses. The survey results, including this author's analysis, yielded quantitative measures for assessing the information behaviour and, in particular, the information-seeking behaviour of the community members.

1.6.3 Personas

The information obtained from the interviews and the survey analysis served for defining six personas that represent six groups of 'typical' HEP scholars. The HEP information specialists approved the representations and arranged these personas in order of the degree to which they consume scholarly materials. This order also represents the perception of the information specialists regarding the effort that one needs to invest to support the groups of scholars represented by each persona.

The use of personas in a theoretical study constitutes a novel approach to modelling information-seeking behaviour.

1.6.4 Research and Work Experience

For more than two decades, the author has been involved with software development for information systems. In particular, for the last 12 years she has been working for an international software company that develops solutions for libraries and information centres, where she was responsible for the design of the user interfaces of two flagship products. This work included the building of a conceptual model of the interface, partnering with user-interface experts to design the interface, working with the development team to develop the interface, and collaborating with focus groups and development partners to obtain feedback and conduct usability studies, both of which contributed to the further development of the products. During that time, this author published 17 papers addressing various aspects of information systems in the library (see Appendix B).

1.6.5 Information-Seeking Model

Using the knowledge acquired through the interviews, the survey analysis, work experience, and a survey of the relevant literature, the author created an original model of the information-seeking behaviour of the HEP community scientists in the context of their work-related information behaviour and used the personas to evaluate the model.

The new information-seeking model contributes to information-seeking research in the focus on a specific scientific community that has distinct and well-defined information-seeking patterns. To the author's knowledge, this is the first time that an information-seeking model has been based on a specific community of scholars. Furthermore, the model addresses the full spectrum of active information-seeking of the community.

1.6.6 User-Interface Design

The model that the author constructed served as a basis for the author's design of a new user interface for SPIRES, the HEP database of scholarly publications of community members. The author presented this design to the HEP information specialists in writing and verbally, in a meeting that took place at the European Organization for Nuclear Research (CERN) in Geneva. Because of limited human resources, the team adopted only a few of the author's recommendations for their initial version of the successor of SPIRES, called INSPIRE, but the implementation of more of these recommendations has already been planned. As of this writing, the INSPIRE database is in beta testing, and the author continues to advise the information specialists about specific issues that come up.

The concretization of an information-seeking model into a user-interface design is an innovative approach to the way in which a theoretical model can be assessed and applied in a practical context—in this case, to improve the support for the information-seeking practices of a specific scientific community.

1.7 Design and Evaluation Processes

This research consists of a series of steps, each providing the basis for the next step (Figure 2).

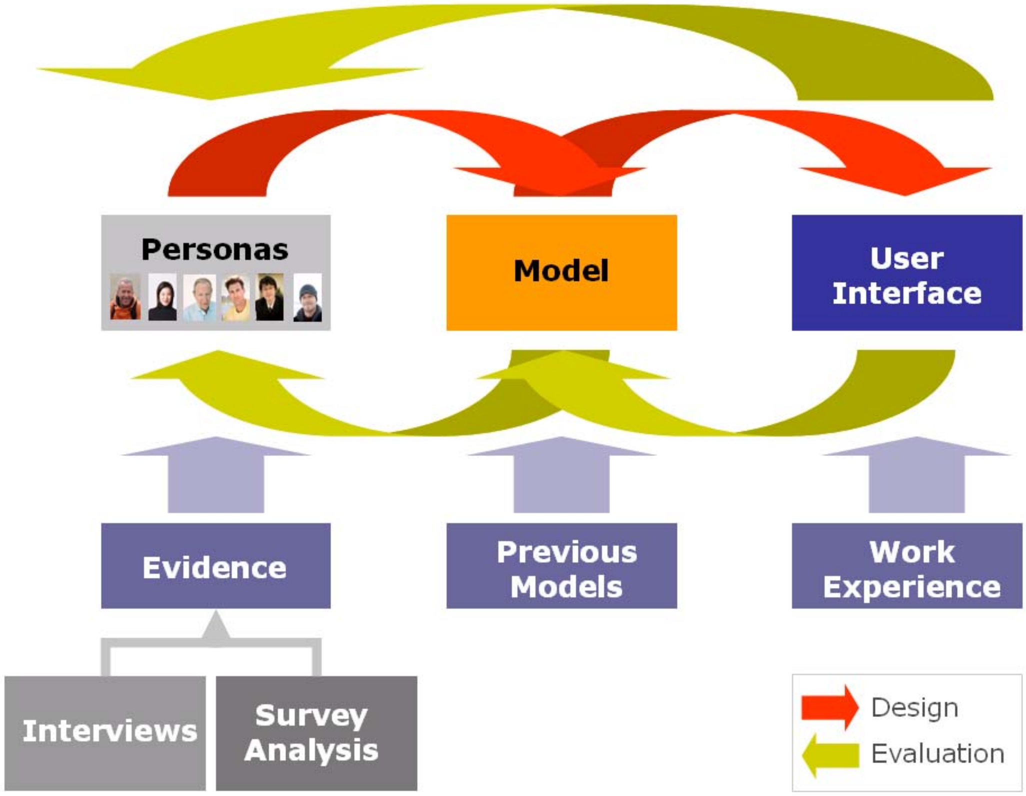


Figure 2: Design and evaluation

The purpose of the first steps—interviews with researchers and the analysis of the HEP survey—is the gathering of evidence, which then serves as a basis for the definition of personas. The personas, in turn, along with a survey of existing models, serve as the basis for the design of the information-seeking model, including its subcomponent, the information-searching model. The information-seeking model is evaluated with the aid of the personas.

The models and the author's work experience serve as the basis for the user-interface design. The design is evaluated by means of the model and the personas.

1.8 Structure of the Thesis

This thesis is divided into seven chapters. Following the introduction (Chapter 1), a literature overview in Chapter 2 traces the evolution of information search-and-retrieval models that emphasise human information-seeking behaviour. The models are described in light of the development of research in this area, and their relevance to this study is explained.

Chapter 3 describes the high-energy physics community and its information-seeking behaviour and search practices as ascertained by observations of, and interviews with, HEP researchers. An analysis of HEP scientists' open-ended responses to a survey of their information-seeking practices is provided, and a set of criteria that describe those practices is formulated.

Chapter 4 explains how the findings from the interviews and survey responses were leveraged to create a set of six personas that represent typical members of the HEP community. These personas later serve in the evaluation of an information-seeking model proposed by the author and a user-interface design that the author derives from this model.

In Chapter 5, an original model that aims to represent the information-seeking practices by HEP scholars is presented. The personas that are defined in Chapter 4 are used to evaluate the information-seeking model.

Chapter 6 explains how the information-seeking model is used to derive a software user-interface design. Aspects of the user interface are discussed in terms of the components and workflows making up the model, and the model is then used to evaluate the user-interface design. The design is also corroborated through the use of personas and additional interviews with HEP community members. Finally, a set of measures is defined to provide an informed evaluation of the interface's effects on the information-seeking behaviour of HEP scientists who use it.

Chapter 7 summarizes the research findings and suggests directions for future work in this area.

Appendix A contains summaries of the interviews that the author conducted with the HEP researchers in 2008 and the beginning of 2009.

Appendix B lists the author's publications.

1.9 Previous Publications by This Author

Parts of this thesis are drawn from previous publications, as follows:

- Parts of Chapter 3 appear in Sadeh 2008a.
- The discussion about clustering and faceted browsing (Chapter 6) was presented in Sadeh 2008a.
- Discussions about users' expectations and behaviour in the library domain are discussed in several publications (Sadeh 2007a, 2007b, 2007c, 2008c). Some background information is taken from Sadeh 2007b.

Other publications by this author are listed in Appendix B.

Chapter 2 Literature Review

2.1 Introduction

The research for this thesis spans several scholarly areas. The topic of users' information-seeking behaviour is examined through user studies, on the one hand, and through literature related to information seeking and retrieval, on the other hand. While literature about information seeking and retrieval models has contributed to establishing the theoretical background for describing user behaviour in this thesis, user studies provide evidence regarding the actual use of information systems. Both the theoretical understanding and the practical findings were instrumental in the author's design of the information-seeking model and, later, in the user-interface design.

Once the high-energy physics (HEP) community was selected as the target of this research, literature about the community and its information systems provided insights about the distinct information behaviour of the community members. This literature contributed to the author's understanding and interpretation of the members' information-seeking behaviour in the specific context of HEP research. User studies that focus on this community have been of great value to the discussions in this thesis, but user studies that relate to similar communities or more general scientific communities are also relevant as measures for comparison between the behaviour of the HEP researchers and other scientists.

Information gathered in the area of user-interface design was helpful in translating the abstract model into a proposal for an actual information system. The use of personas as part of the whole research process was also supported by evidence from relevant literature.

This thesis is therefore positioned at the intersection of several research areas (Figure 3):

- Information seeking and retrieval and models of information behaviour and information seeking

- User studies focusing on academic communities' information-seeking perceptions and practices
- User-interface design
- Information systems and scholarly communication practices of the high-energy physics community

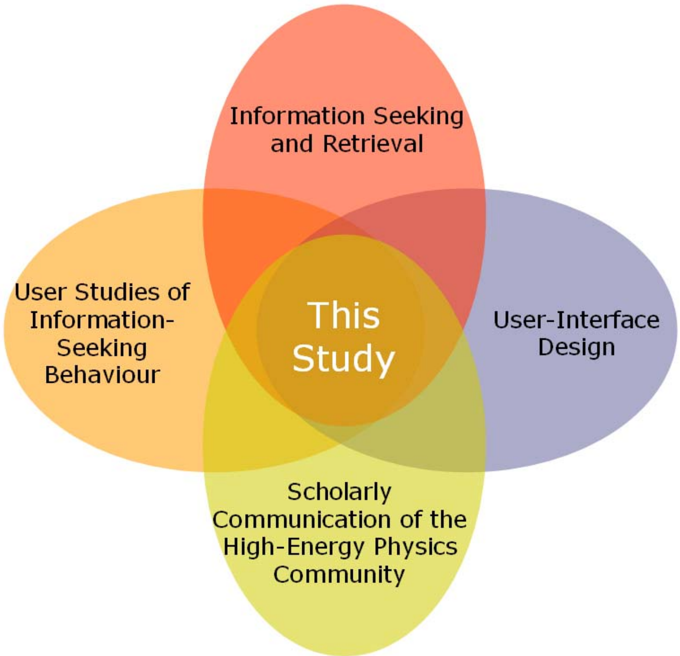


Figure 3: Context of this study

This chapter provides a general survey of applicable literature, with certain topics covered in greater detail later in this thesis. Out of the many works consulted by the author, some were selected to be described in this chapter on the basis of their concrete contribution to the discussion at hand. For example, information-seeking models that are discussed in this thesis are those that are general enough to apply, to some degree, to the HEP community and hence could serve as a basis for the definition of a new information-seeking model that is specific—focusing on the HEP community—and still inclusive, covering the full spectrum of information-seeking practices of the community.

2.2 Information Seeking and Retrieval: Terminology and Models

In 1979, the philosopher Jean-François Lyotard coined the term ‘postmodernism’, giving a name to a school of thought that influenced the second half of the twentieth century, gaining momentum at the end of the 1960s. Postmodernism—a reaction against the rationalism, scientism, and objectivity of modernism—signifies a shift of human philosophical perceptions toward the acceptance of pluralism, decentralization, dissemination, and networked knowledge. Representing scepticism toward universal and authoritative views, postmodernism acknowledges the notion of relative truth and the human interpretation of reality within a specific context.

When one projects postmodernist notions on information systems, it becomes clear that human diversity is such that the user’s context is crucial to the success or failure of technological solutions that aim at satisfying information needs. Therefore, it is no wonder that research in the area of information seeking and retrieval, which initially focused on a system-oriented view, has shifted toward a user-centric approach. Furthermore, while information systems in their early years were scarce and struggled to provide basic functionality, by the mid-1980s these systems formed a robust infrastructure that could allow information researchers to turn their eyes—and thoughts—to the user. Not surprisingly, this trend has been accompanied by a shift from quantitative methods of investigation to qualitative methods that enable researchers to gain an in-depth understanding of individuals and their information needs. Bates describes the shift to qualitative research as follows: ‘Whereas the scientific perspective was driven by a fundamental assumption that we can be easily fooled by our own perceptions, and so we had to place intervening controls between ourselves and the thing observed, the qualitative approach revels in the ingenuity and uniqueness of insight of the individual who is doing the research’ (Bates 2002).

Because of the focus of the present study on researchers and researchers’ needs, this chapter discusses selected trends and models that cohere with the user-centric approach to information-behaviour research and information-seeking research. The way in which these models apply to information-seeking behaviour of HEP researchers is discussed in detail in Chapter 5.

2.2.1 Terminology

To ensure a common understanding of terminology, this section proposes definitions of the terms *information behaviour*, *information need*, *information-seeking behaviour*, *information-searching behaviour*, and *information use*.

2.2.1.1 Information Behaviour

Ingwersen and Järvelin (2005) define information behaviour as the 'human behavior dealing with generation, communication, use and other activities concerned with information, such as, information seeking behavior and interactive IR [information retrieval]' (Ingwersen and Järvelin 2005, 21).

Wilson (2000) proposes a slightly different definition of information behaviour: 'the totality of human behavior in relation to sources and channels of information, including both active and passive information seeking, and information use. Thus, it [information behaviour] includes face-to-face communication with others, as well as the passive reception of information as in, for example, watching TV advertisements, without any intention to act on the information given' (Wilson 2000, 49).

In the context of this thesis, the definition of information behaviour is derived from Wilson's definition but is more specific: the behaviour of the members of a research community in connection with their *work-related* sources and channels of information, including both active and passive information seeking, and information use. Limiting the definition of information behaviour to the work context eliminates any reference to the passive reception of information without any intent on the part of the recipient to act on the information given. In the context described in this study, the fact that the recipient is a member of the research community already implies some degree of intent. It could well be that information is passively acquired when, for example, researchers take part in a conference and attend lectures that are not directly related to immediate work tasks or have a discussion with colleagues over dinner; however, any such information may become relevant at a later stage.

2.2.1.2 Information Need

The definition of an information need is far from straightforward, although many consider the term self-explanatory. Dervin and Nilan (1986) discuss at length the shift in the definition of information need from a focus on the system—what it is in the information system that the user needs—to a focus on the user—what it is that the user needs. An information need, in their view, is a situation 'in which the individual's internal sense "runs out." The person must create a new sense' (Dervin and Nilan 1986, 21). The person needs to bridge a gap to return to a state of possessing internal sense.

Ingwersen and Järvelin (2005) provide a definition that acknowledges the same gap: for them, information need 'signifies a consciously identified gap in the *knowledge* available to an actor. Information needs may lead to *information seeking* and formulation of *requests* for information' (Ingwersen and Järvelin 2005, 385).

This work differentiates between information seeking and information searching and therefore adopts the definition of information need from Ingwersen and Järvelin, with a slight modification: a consciously identified gap in the knowledge available to a user that may lead to information searching.

2.2.1.3 Information-Seeking Behaviour

Wilson (2000) defines information-seeking behaviour as 'the purposive seeking for information as a consequence of a need to satisfy some goal. In the course of seeking, the individual may interact with manual information systems (such as a newspaper or a library), or with computer-based systems (such as the World Wide Web)' (Wilson 2000, 49).

According to Ingwersen and Järvelin (2005), information-seeking behaviour is 'human information behavior dealing with searching or seeking information by means of information sources and (interactive) information retrieval systems' (Ingwersen and Järvelin 2005, 21).

For the purposes of this study, information seeking is defined in a slightly different way, acknowledging the reliance of researchers on finding information that they do not explicitly specify but that is nevertheless vital to their work tasks. This study views information-seeking behaviour as information behaviour that consists of an active pursuit of information through the use of information systems, by the members of a research community. By not referring to a specific need, this definition is applicable to both directed and undirected searching (see Bates's modes of information seeking in 2.2.2.5).

2.2.1.4 Information-Searching Behaviour

Information-searching behaviour, according to Wilson (2000), 'is the "micro-level" of behavior employed by the searcher in interacting with information systems of all kinds. It consists of all the interactions with the system, whether at the level of human computer interaction (for example, use of the mouse and clicks on links) or at the intellectual level (for example, adopting a Boolean search strategy or determining the criteria for deciding which of two books selected from adjacent places on a library shelf is most useful), which will also involve mental acts, such as judging the relevance of data or information retrieved' (Wilson 2000, 49).

In some studies, searching and seeking are regarded as synonyms. Ingwersen and Järvelin (2005), for example, do not differentiate between information seeking and searching (see 2.2.1.3 for their definition of information seeking).

This thesis follows Wilson's approach, with his distinction between information seeking and information searching. Information-searching behaviour is, then, the aspect of information-seeking behaviour that deals especially with active, directed searching in information systems for data that can be specified to some degree.

2.2.1.5 Information-Use Behaviour

According to Wilson (2000), information-use behaviour 'consists of the physical and mental acts involved in incorporating the information found into the person's existing knowledge base. It may involve, therefore, physical acts such as marking sections in a text to note their importance or significance, as well as mental acts that involve, for example, comparison of new information with existing knowledge' (Wilson 2000, 50). This definition, albeit useful, fails to relate to the information need that may have triggered the search process (although a person may find information useful even without having a perceived need for it). Therefore, the present study modifies Wilson's definition of information-use behaviour as follows: the physical and mental acts involved in extending a person's existing knowledge base with new information.

2.2.2 Models

Until the 1990s, research in the area of user needs, information needs, and information-seeking behaviour did not support the accumulation of a body of theory and empirical findings that could serve for further research. Wilson (1999) states a number of reasons for this situation. First, the quantitative research methods used could not provide insights that could support the development of a theory or practice in this area. Second, researchers focused on information science and did not adopt theoretical models of human behaviour from other fields. Third, general models of user behaviour started to emerge only in the mid-1980s.

In another retrospective review of earlier research, Dervin and Nilan (1986) observe that 'almost without exception "information needs" have not been defined as what users think they need but rather in terms that designate what it is in the information system that is needed. The definitions have not focused on what is missing for users (i.e., what gaps they face) but rather on what the system possesses' (Dervin and Nilan 1986, 17).

With the adoption of qualitative research methods and the acceptance of theories and models originating in the social sciences, the situation has changed. New models have emerged and have served as a basis for further research, such as the research conducted as part of the present study. These models include those proposed by Wilson (1981, 1999); Dervin—in her paper 'An overview of Sense-Making research: Concepts, methods, and results to date', presented at the annual meeting of the International Communication Association, Dallas, TX, in May 1983 (cited by Dervin and Nilan 1986); Ellis—in the paper 'A behavioural approach to information retrieval design', published in the *Journal of Documentation* in 1989 (as described by Wilson [1999] and Ingwersen and Järvelin [2005], among others); Bates (1989, 2002); Kuhlthau (1991); Ellis (1993); Belkin et al. (1995); and Marchionini (1995).

Theories and models that are of the greatest significance to this thesis are briefly presented in this chapter and discussed in more detail in Chapter 5.

2.2.2.1 Wilson's Models of Information Behaviour and Information-Seeking Behaviour

Wilson's 1981 paper, which includes definitions of concepts and models of information behaviour and information-seeking behaviour, has significantly influenced information research since then. By defining basic concepts and offering a new approach to the motivation underlying information-seeking behaviour, Wilson aims to reduce the confusion related to user studies and the perception of information needs. As noted by Bawden (2006), 'by 1980, the field [of user studies] was burgeoning in interest and publication, but lacking clear foundations of method and conceptual framework. Wilson's article must be seen against that background' (Bawden 2006, 672). After discussing the challenges involved in the use of the term 'information', Wilson presents a model of information behaviour (discussed in Chapter 5 of this thesis) as 'a way of thinking of the field "user studies"; its aim is not to "model" information-seeking behaviour but to draw attention to the interrelationships among concepts used in the field' (Wilson 1981).

Wilson then explores the concept of information needs and shows the complex relationships between a user in a specific context and an information source. Addressing the context of an individual, Wilson asserts that 'the "user's life world" can be defined as the totality of experiences centred upon the individual as an information user. Within this life-world one important sub-world will be the world of work, within which will exist various "reference groups" with which the user identifies: fellow professionals, the peer group within an organization and so on' (Wilson 1981).

Wilson notes that information-seeking behaviour research can have two forms: it 'can stand on its own as an area of applied research where the motive for the investigation is pragmatically related to system design and development', as in the present study, or it can serve to help one 'understand why the information seeker behaves as he does. This is an area of basic research and, although the resulting knowledge may have practical applications, there is no necessity that it should' (Wilson 1981).

Through a further exploration of the concept of 'information needs', Wilson presents a model of information-seeking behaviour (discussed in Chapter 5 of this thesis) and proposes 'to remove the term "information needs" from our professional vocabulary and to speak instead of information-seeking towards the satisfaction of needs' (Wilson 1981).

Wilson goes on to discuss the context of the user and the influence that this context has on the user's information-seeking behaviour, suggesting that 'the search for determining factors related to needs and information-seeking behaviour must be broadened to include aspects of the environment within which the work-role is performed' (Wilson 1981). One of the consequences of broadening one's understanding of the user's context that Wilson points out is that 'if we wish to uncover the determining factors of behaviour we must do so by first undertaking in-depth studies of well-defined categories of persons, developing explanatory concepts and then testing these concepts in related but different settings' (Wilson 1981).

Wilson predicts a shift in the focus of research—from looking at information sources and systems to 'an exploration of the role of information in the user's everyday life in his work organization or social setting' (Wilson 1981)—and foresees three consequences. The first is the reliance on qualitative research, instead of or in addition to quantitative research methods. The second is the attempt to determine the behaviour's underlying factors by in-depth studies of well-defined groups, and the third is the examination of user behaviour in a wider perspective of psychological and sociological studies.

Wilson's 1981 article had a great impact on information-behaviour research. As Bawden (2006) asserts in his review of Wilson's article, 'one of the strengths of Wilson's 1981 proposals—and, I believe, the reason they are still widely cited—is that their relative simplicity enables a consensus understanding rare in the field. The same is surely true of Wilson (1999) representation of the relation between human information behaviour, information seeking and information retrieval; simple indeed, but with the power to bring clarity where there was none before' (Bawden 2006, 673).

The applicability of Wilson's models to the HEP information-seeking practices is described in Chapter 5.

2.2.2.2 Dervin's Sense-Making Model

The *sense-making* approach 'consists of a set of conceptual and theoretical premises and a set of related methodologies for assessing how people make sense of their worlds and how they use information and other resources in the process' (Dervin and Nilan 1986, 20).

To describe the behaviour of the information-seeker, or sense maker, Dervin proposes a 'situation-gap-use' model in her 1983 paper *An overview of Sense-Making research: Concepts, methods, and results to date*, presented at the annual meeting of the International Communication Association, Dallas, TX (cited in Dervin and Nilan 1986). The elements that make up the model are a situation (in time and space), which defines the context; a desired situation; a gap between the contextual situation and the desired situation, and a bridge that closes this gap (Figure 4). According to this model, the sense maker is stopped at a situation because of some kind of gap and needs to make use of a 'bridge' across the gap to resume the situation of sense making.

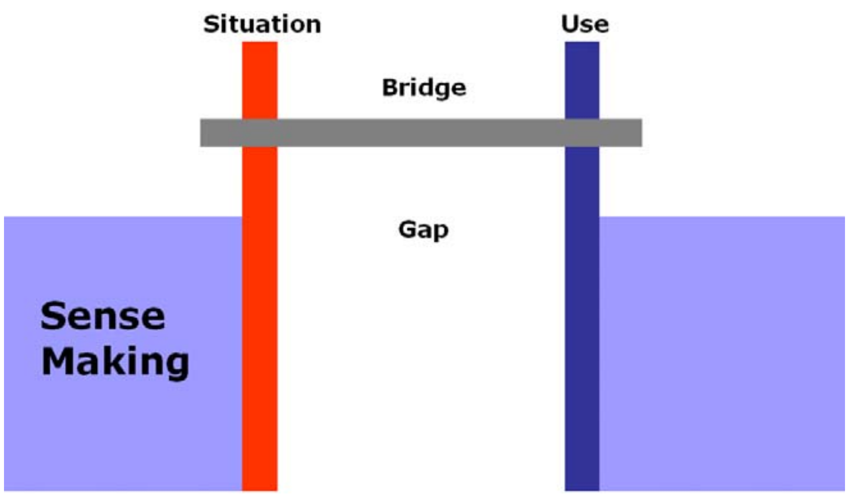


Figure 4: Dervin's situation-gap-use model

In his review of this model, Wilson (1999) sees the strength of Dervin's approach in its methodological effects. When applied to information behaviour, this approach can provide a framework for investigating the nature of the situation, the degree to which the gap can be bridged by information, and the nature of the outcomes following the use of information. Such questioning can serve as a trigger to obtaining insights that may help the designers of information services.

Ingwersen and Järvelin (2005) see the sense-making approach as an important advancement in the research on information seeking. According to them, the sense-making approach 'draws attention to individual sense-making (problem solving) in varying situations, and focuses on the actor [the information seeker] and process viewpoints rather than a systems (or traditional assumptions') viewpoint' (Ingwersen and Järvelin 2005, 62).

2.2.2.3 Ellis's Feature Set

Ellis (1989, as described by Wilson [1999] and Ingwersen and Järvelin [2005], among others) identifies eight features of information-seeking behaviour that characterize both academic and non-academic searchers in various fields and professions.

The eight features that Ellis identifies are the following:

- *Starting*: performing activities that characterize the initial stages of a search process, such as looking for review literature or asking the help of a colleague
- *Chaining*: following leads in known materials to other materials—for example, following references at the end of an article
- *Browsing*: conducting semi-directed searches, such as identifying articles in a journal's table of contents
- *Differentiating*: using differences in the examined information as filters
- *Monitoring*: maintaining an awareness of the current developments in the field
- *Extracting*: systematically identifying relevant materials
- *Verifying*: checking information for accuracy
- *Ending*: ending the process, possibly by a final search

The way in which these eight features interrelate to form an individual information-seeking pattern depends on the specific context. Ellis (1993) demonstrates that variations of these features and of the way in which they relate to each other emerge when the information-seeking behaviour of individuals belonging to different communities (social sciences, economics and social history, physics, chemistry, and English literature) is examined.

Referring specifically to the physicists, Ellis (1993) notes:

> The study of the physicists employed slightly different terminology and derived a model that differed in detail from that derived from the interviews with the social scientists. Five main categories were identified: *initial familiarization*—activities undertaken at the earliest stages of information seeking; *chasing*—following up citation links between materials; *source prioritization*—ranking sources based on perceptions of their relative importance; *maintaining awareness*—activities involved in keeping up-to-date; and *locating*—activities engaged in to actually find the information. (Ellis 1993, 482)

However, as Ellis (1993) points out, 'although the models differ in terminology and detail, when the features and characteristics of the models are compared they can be seen to be representing fundamentally the same activities' (Ellis 1993, 483).

Wilson (1999) divides Ellis's set of features into several types:

> Browsing, chaining and monitoring are search procedures, whereas differentiating is a filtering process and extracting may be seen as an action performed on the information sources.
>
> The remaining behaviours do not necessarily take place in a specific sequence and may be initiated in different sequences at different times in the overall search process. Ellis's account, therefore, in terms of the different kinds of features it embodies, appears to sit between the micro-analysis of search behaviour (starting, chaining, extracting, verifying, ending) and a more macroanalysis of information behaviour generally (browsing, monitoring, differentiating). (Wilson 1999, 7)

Because Wilson (1999) sees Ellis's features as functioning at different levels of an information-seeking process than the elements that Wilson had defined in his information-seeking model (1981), he explores the possible inclusion of these features in his model. Further discussion is provided in Chapter 5 of this study.

2.2.2.4 Kuhlthau's Process Model

Kuhlthau (1991) presents the process of information searching from the user's perspective as a 'series of encounters with information within a space of time' (Kuhlthau 1991, 361). In the context of her research, this information search process (ISP) 'is the user's constructive activity of finding meaning from information in order to extend his or her state of knowledge on a particular problem or topic' (Kuhlthau 1991, 361).

Following Dervin's sense-making approach (see 2.2.2.2), Kuhlthau (1991) explains:

> Information seeking is viewed as a process of sense-making in which a person is forming a personal point of view… The individual is actively involved in finding meaning which fits in with what he or she already knows, which is not necessarily the same answer for all, but sense-making within a personal frame of reference. Information from various sources is assimilated into what is already known through a series of choices. Formal organized sources from information systems interact with informal sources from everyday life experiences. The ISP culminates in a new understanding or a solution which may be presented and shared. Evidence of the transformation of information into meaning is present in the products or presentations in which users share their new knowledge with others. (Kuhlthau 1991, 361)

Kuhlthau (1991) defines six stages, or encounters, of the information search process—*initiation* (the user is becoming aware of a need for information), *selection* (identifying the topic for searching), *exploration* (seeking information on the topic), *focus formulation* (structuring and fixing the problem at hand), *collection* (gathering relevant information), *and presentation* (using the results of the search). The model presents three dimensions for each stage: affective (feelings), cognitive (thoughts), and physical (actions) (Figure 5).

	Stages in ISP	Feelings Common in Each Stage	Thoughts Common to Each Stage	Actions Common to Each Stage	Appropriate task According to Kuhlthau's Model
1	Initiation	Uncertainty	General/Vague	Seeking Background Information	Recognize
2	Selection	Optimism			Identify
3	Exploration	Confusion/ Frustration/ Doubt		Seeking Relevant Information	Investigate
4	Formulation	Clarity	Narrowed/ Clearer		Formulate
5	Collection	Sense of Direction/ Confidence	Increased Interest	Seeking Relevant or Focused Information	Gather
6	Presentation	Relief/Satisfaction or Disappointment	Clearer or Focused		Complete

Figure 5: Kuhlthau's information search process (ISP).
Modified from Kuhlthau (1991), Table 2.

Wilson (1999) notes that 'in effect, what Kuhlthau postulates here (and confirms by empirical research) is a process of the gradual refinement of the problem area, with information searching of one kind or another going on while that refinement takes place' (Wilson 1999, 8). Wilson draws similarities between Kuhlthau's model and Ellis's feature list (2.2.2.3)—although he finds Kuhlthau's model more general than Ellis's feature list because the former draws attention to the feelings that are associated with each stage—and tries to bring them together by mapping Ellis's categories to Kuhlthau's stages (Figure 6). Wilson (1999) concludes as follows:

> Through this merger of the two models, we can see strong similarities and the major difference appears to be that Ellis specifies the modes of exploration or investigation. The point must be reiterated, however, that Ellis does not present his characteristics as stages but as elements of behaviour that may occur in different sequences with different persons, or with the same person at different times. Thus, the two models are fundamentally opposed in the minds of the authors: Kuhlthau posits

stages on the basis of her analysis of behaviour, while Ellis suggests that the sequences of behavioural characteristics may vary.' (Wilson 1999, 9)

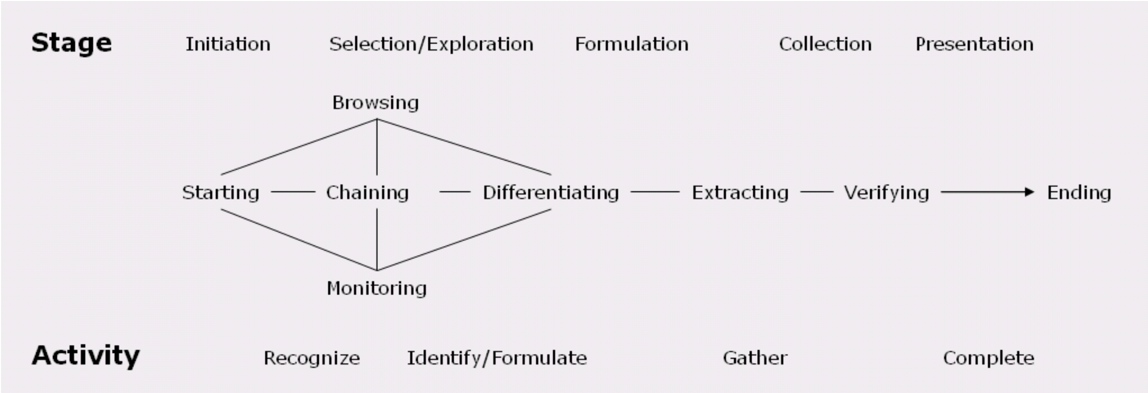


Figure 6: Wilson's comparison of Ellis's and Kuhlthau's frameworks.
Modified from Wilson (1999), Figure 6.

2.2.2.5 Bates's Models

In her discussion of searching, Bates (1989) puts an emphasis on the evolving nature of a search process, as is evidenced from the observation of manual information-seeking behaviour of individuals, and introduces a model that describes this process.

Bates (1989) first challenges the classic model of information retrieval (Figure 7) on several grounds. She argues that the model is not inclusive regarding the type of searches it represents and that it has many limitations 'as a realistic representation of actual searches' (Bates 1989). Furthermore, the model limits the 'creativity in developing IR systems that really meet user needs and preferences' (Bates 1989).

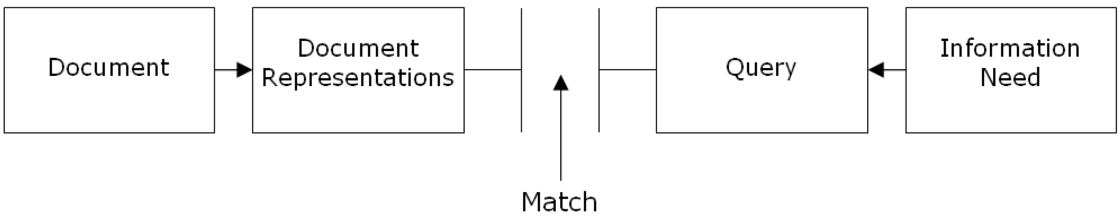


Figure 7: The classic information-retrieval model.
Modified from Bates (1989), Figure 1.

The berrypicking model (Figure 8) that Bates (1989) proposes suggests a different perspective on the nature of the query and the search process, embeds the search process in an information territory, and applies to a range of search techniques.

According to Bates (1989),

> users may begin with just one feature of a broader topic, or just one relevant reference, and move through a variety of sources. Each new piece of information they encounter gives them new ideas and directions to follow and, consequently, a new conception of the query. At each stage they are not just modifying the search terms used in order to get a better match for a single query. Rather the query itself (as well as the search terms used) is continually shifting, in part or whole….
>
> Furthermore, at each stage, with each different conception of the query, the user may identify useful information and references. In other words, the query is satisfied not by a single final retrieved set, but by a series of selections of individual references and bits of information at each stage of the ever-modifying search. A bit-at-a-time retrieval of this sort is here called berrypicking. This term is used by analogy to picking huckleberries or blueberries in the forest. The berries are scattered on the bushes; they do not come in bunches. One must pick them one at a time. One could do berrypicking of information without the search need itself changing (evolving), but in this article the attention is given to searches that combine both of these features.' (Bates 1989)

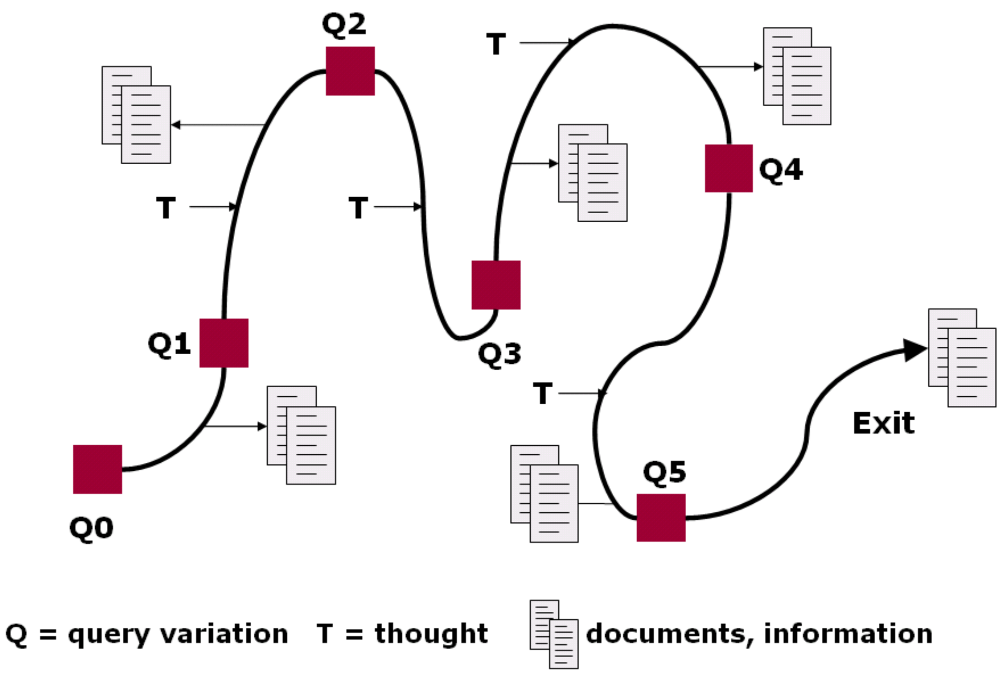


Figure 8: The Bates berrypicking model, with an evolving search.
Modified from Bates (1989), Figure 2.

Bates (1989) also emphasizes that the search is taking place in the context of a 'universe of interest', which is within the larger context of the 'universe of knowledge' (Bates 1989). The behaviour of the searcher is the focus of attention in this model, and 'the continuity represented by the line of the arrow is the continuity of a single human being moving through many actions toward a general goal of a satisfactory completion of research related to an information need. The changes in direction of the arrow illustrate the changes of an evolving search as the individual follows up various leads and shifts in thinking' (Bates 1989).

Bates (1989) lists several searching techniques: searching for documents referenced by a given document, searching for documents that cite a given document, browsing systematically through volumes and issues of a specific journal, looking at materials that are physically set in one location, searching by subject, and searching by author. Bates refers in her description to *manual* searches and mentions information systems only in regard to subject searches through abstracting and indexing databases. However, the relevance of her model to current technologies is most significant. In fact, today information systems can provide a means to support all these techniques; furthermore, an environment such as the one that Bates (1989) describes—comprised of multiple information sources—is supported by current technologies that enable a searcher to move from one system to another. However, most search techniques that Bates discusses are not considered in this study as 'searching'; rather, they are viewed as linking or navigating. This distinction is discussed further in Chapter 5.

Acknowledging the partial role of active searching in the knowledge acquisition process of humans (according to Bates (2002), 'it is not unreasonable to guess that we absorb perhaps 80 percent of all our knowledge through simply being aware, being conscious and sentient in our social context and physical environment [Bates 2002, 3]). Bates (2002) proposes a matrix, which describes modes of information seeking, as a way to integrate information seeking and information searching (Figure 9).

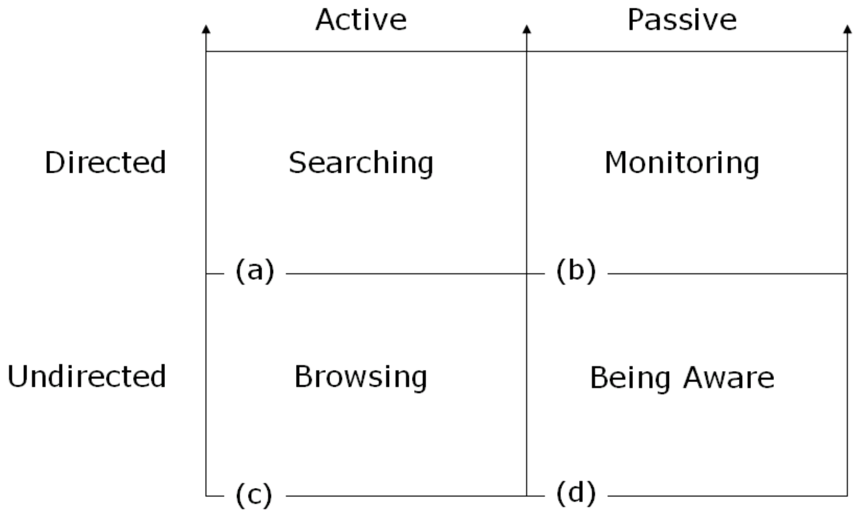


Figure 9: Modes of information seeking proposed by Bates.
Modified from Bates (2002), Figure 3.

The first row of the figure ('directed' modes) refers to modes in which a person seeks information that he or she can specify, to some extent, while the second row ('undirected' modes) refers to modes in which the person exposes himself or herself to information in a random manner. The columns differentiate between modes in which a person is actively looking for information or is just absorbing information in a passive way.

As explained in Chapter 5 of this study, the modes of information seeking suggested by Bates (2002) cohere with information-seeking practices of HEP researchers. See Chapter 5 for a detailed discussion of these modes in the context of the HEP community

2.2.2.6 Belkin's Modes of Interaction

Belkin et al. (1995) emphasize the interactive nature of the information-seeking process and join with Bates (1989) in arguing that such a process consists of multiple interactions with an information system and that 'people's conceptions of their information problems change through their interactions with the IR system' (Belkin et al. 1994). Furthermore, each type of information need requires a different kind of interaction with the information system.

Belkin et al. (1995) go on to describe 'a multidimensional space of information-seeking strategies (ISSs)' (Belkin et al. 1995) that are the basis of an information-seeking behaviour model. An ISS is a specific behaviour employed by a person who interacts with a 'knowledge resource' (Belkin et al. 1995) when searching for information. Each such behaviour—or ISS—is derived from a certain context (the person's information-seeking goals and the knowledge that is available to the person before starting the process, such as a specification of the information needed). The ISSs constitute interactions between the person and an information system, and several such interactions form an 'episode' (Belkin et al. 1995). Furthermore, the searcher's knowledge and goals evolve through the information-seeking episode, changing the specific values of the ISSs.

The Belkin et al. (1995) model suggests that 'any single information-seeking interaction is a complex activity, which can be characterized according to its values on a relatively small set of factors, or dimensions' (Belkin et al. 1995). Four modes[2] of interaction are proposed: method of interaction, goal of interaction, mode of retrieval, and resource considered (Figure 10).

---

[2] Although Belkin et al. (1995) refer to *modes* of interaction, perhaps more suitable terms would be *facets* or *dimensions* of interaction. The original term is used in this chapter for the sake of consistency with Belkin et al.

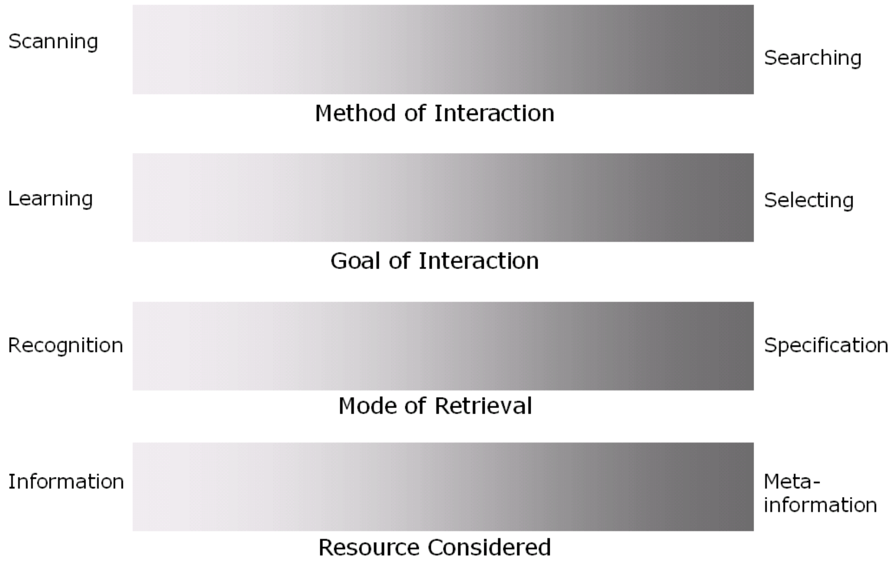


Figure 10: Belkin et al.'s modes of interaction.
Modified from Belkin et al. (1995).

'According to our conceptualization', note Belkin et al. (1995), 'information-seeking behavior is characterized by movement from one strategy to another within the course of a single information-seeking episode, as a person's problematic situation changes' (Belkin et al. 1995). They go on to explain that 'having ISSs described by, and located in, the kind of space we suggest gives us a means to describe movement from ISS to ISS, as well as to describe the individual ISSs, and potentially the means to understand such movement well enough to devise methods for supporting it in a principled fashion. From this point of view, we can consider ISSs as types of user interactions within the IR system, rather than as queries or demands put to that system' (Belkin et al. 1995).

This model and its applicability to the HEP information-seeking process are discussed further in Chapter 5.

2.2.2.7 The Information-Seeking Process According to Marchionini

Marchionini (1995) points out that 'the information-seeking process is both systematic and opportunistic. The degree by which a search exhibits algorithms, heuristics, and serendipity depends on the strategic decisions that the information seeker makes and how the information-seeking factors interact as the search progresses' (Marchionini 1995, 49). Marchionini describes the information-seeking process as a set of subprocesses that a searcher can carry out linearly; however, the searcher can also move from one subprocess to others as the information-seeking process advances.

A further discussion of the way Marchionini models the information-seeking process is provided in Chapter 5.

2.3 User Studies

User studies are discussed in this chapter as means to shed light on the actual information-seeking behaviour of researchers in the context of their work.

Studies focusing on the younger generation—primarily members of academic communities, including high-school students and undergraduates—offer insight on the characteristics of the researchers of tomorrow (OCLC 2005, 2006, 2007; CIBER/UCL 2008; Connaway and Dickey 2010). Studies focusing on scientific communities, addressing information-seeking behaviour of researchers, serve to elucidate the commonalities and the differences between HEP researchers and others (RIN 2006; Hemminger et al. 2007; Jamali and Nicholas 2008; Haines et al. 2010). Earlier studies (Brown 1999; Murphy 2003) emphasize the considerable change that has taken place in the information-seeking behaviour of researchers in the last decade. Markey (2007a, 2007b) summarizes twenty-five years of end-user searching and looks at her findings in the context of information-retrieval models. Gentil-Beccot et al. (2008) describes the practices of the HEP community as depicted by a survey launched in summer 2007. Findings of these studies are described in this section.

General studies show that students are using Web search engines as their research tools and point out that the use of the physical library, the library Web site, and the information systems offered by the library is negligible when compared to the use of Web search engines (OCLC 2005, 2006). The first OCLC report (2005) analyses the results of a survey of more than 3,300 respondents aged 14 to 65 from Australia, Canada, India, Singapore, the United Kingdom, and the United States. A companion report, focusing primarily on the perceptions of college students at the undergraduate and postgraduate levels, was published in April 2006 (OCLC 2006).

Only 2% of the undergraduate and postgraduate students surveyed for the OCLC 2006 report stated that they would begin their search for information on a particular topic at the library's Web site, despite the fact that 61% used the library Web site at least once and 85% indicated that they have a 'favourable view' of the online library. Furthermore, 77% believed that library resources (online and physical) are trustworthy or credible, and 76% described them as accurate (only 23% described Web search engines as trustworthy and credible, and 24% considered them accurate). Most of the students (75%) agreed that librarians add value to the information-search process.

When the students were asked about Internet tools and services such as Web search engines, e-mail, instant messaging, online news, online bookstores, blogs, and RSS feeds, their answers indicated that many are familiar with and use most of these tools—primarily e-mail, Web search engines, and instant messaging. On the other hand, more than 50% of the students replied that they were not aware of their library's e-book collection, and only 62% were certain that their library offers online databases and e-journals. When asked which resource they turn to first when they are looking for information, 89% of the students indicated Web search engines, 2% indicated online databases, another 2% indicated their library's Web site, and the rest indicated other Internet tools and services. The students considered Web search engines faster (90%), more convenient (84%), and easier to use (87%) than the online or physical library.

A later OCLC survey, published in 2007, revealed that while users had taken even more advantage of the Web services such as Web search engines, blogs, and online bookstores since the 2005 survey, one Web service actually featured decreased usage—the library's Web site (OCLC 2007).

A study undertaken by the Centre for Information Behaviour and the Evaluation of Research and University College London (commissioned by the British Library and JISC) (CIBER/UCL 2008), aiming 'to identify how the specialist researchers of the future, currently in their school or pre-school years, are likely to access and interact with digital resources in five to ten years' time' (CIBER/UCL 2008, 5) reveals some of the information-seeking behaviour patterns that will need to be addressed by scholarly information systems. The study states that 'in general terms, this new form of information seeking behaviour [digital information-seeking behaviour] can be characterised as being horizontal, bouncing, checking and viewing in nature. Users are promiscuous, diverse and volatile and it is clear that these behaviours represent a serious challenge for traditional information providers, nurtured in a hardcopy paradigm and, in many respects, still tied to it' (CIBER/UCL 2008, 9).

The study concludes that the information literacy of young people has not improved despite the exposure to technological tools from an early age. Furthermore, young people do not invest time in understanding their information need or developing search strategies, and spend little time in evaluating the information that they find. The study goes on to suggest that 'it would be a mistake to believe that it is only students' information seeking that has been fundamentally shaped by massive digital choice, unbelievable (24/7) access to scholarly material, disintermediation, and hugely powerful and influential search engines. The same has happened to professors, lecturers and practitioners. Everyone exhibits a bouncing/flicking behaviour, which sees them searching horizontally rather than vertically. Power browsing and viewing is the norm for all' (CIBER/UCL 2008, 8).

In a study that analyses and synthesizes twelve selected OCLC, RIN, and JISC user-behaviour projects[3] from the United States and the United Kingdom, Connaway and Dickey (2010) describe common findings across all or most user studies. These findings include the following points:

- Disciplinary-based differences in researcher behaviours exist and apply to both professional researchers and students.

---

[3] The twelve studies include two OCLC reports (2005, 2006), the CIBER/UCL report (2008), and the RIN report (2006) discussed individually in this chapter.

- Online access to materials—primarily e-journals—is extremely important and users favour it over print; users expect access to the full text, including that of older materials.
- Google and other search engines are central to students' information seeking.
- Speed and convenience are important for all users.
- Users feel confident in their information literacy skills although these have not necessarily improved in the recent decades.
- Users expect more functionality in library systems and high-quality metadata to help them search and evaluate content.

Numerous studies have focused recently on academic researchers and throw light on information-seeking behaviour that is more comparable to that of the HEP community. One such study was commissioned by the Research Information Network (RIN 2006) with the goal of assessing the way academic researchers in the United Kingdom perceive and use discovery services. The survey, in which 395 academic researchers and 55 academic librarians participated, yielded the following information:

- Researchers look primarily for journal articles. Almost all researchers (99.5%) rely on journal articles for their research, and 71% of them indicated journal articles as the materials most important to them. Other material types include monographs, book chapters, organisational Web sites, newspapers, historical records, conference proceedings, datasets, and preprints; however, none of these was indicated as the most important research material by the majority of the researchers.

- The most common search strategy, reported by almost 50% of the survey participants, is to start with a broad search and then refine the initial result set. Researchers expressed concern about missing important information; hence, they are willing to risk facing irrelevant results when starting with a broad search. A relatively small group—17% of the respondents—reported that they start with a precise search, and about a third of the respondents alter their strategy depending on the specific search: when they are conducting an exploratory search, they prefer scanning long result lists and identifying items of interest.

- The discovery tool mentioned specifically by the largest number of people is Google, with Google Scholar the second, followed by Web of Science, PubMed, and Web of Knowledge (RIN 2006, 29). Furthermore, the report states that 'there is a sharp fall from the most frequently mentioned to the remainder, and the least-mentioned from the top fifty are mentioned by only three people' (RIN 2006, 20).

- Researchers match the discovery tool to the task at hand. For example:
  - Google is used 'for a variety of general search tasks, but not significantly for those that are critical to research: in tasks such as finding a reference or researching a new area most users also use other tools as well. Locating datasets and non-text sources are important uses of general search engines as they are not well-identified elsewhere' (RIN 2006, 29).
  - Google Scholar is typically used for known-item searches. Furthermore, 'researchers appear to be using it for convenience rather than relying on it for research in depth' (RIN 2006, 29).
  - Colleagues are often asked for information about areas that are not familiar to the requester or that relate to materials other than articles and books (e.g., datasets).
  - Bibliographic databases and citation databases (Web of Science and Web of Knowledge) are used for researching new areas.

The report (RIN 2006) concludes the following:

> Bibliographic databases and Web of Science/Web of Knowledge are used for the three core activities of researching a new area, literature review, and finding references. Researchers, contrary to some hypotheses, are not relying on general web search for these mission-critical tasks, but are using the general search engines to support the other tools they use.
>
> The main uses of all the more general tools (such as Google, other forms of general web search and asking colleagues) are for more general background information activities. It is, however, worth noting that the general tools are also used to locate datasets whereas bibliographic tools are not. This suggests that datasets have not yet entered the mainstream as far as perceptions of how

> to locate them and that more focused tools are not providing the detail that researchers need. (RIN 2006, 29)

The RIN study also provides more specific findings about the perceptions of researchers and breaks down results by discipline. However, disciplines are defined very generally in the study, and hence findings that apply to physical sciences do not necessarily apply to high-energy physics. For example, the RIN report does not mention information systems that are crucial to HEP researchers—SPIRES and arXiv—and preprints are indicated as one of the three most important types of scholarly materials by only 5.1% of the respondents, probably because of the complete lack of or only small number of HEP researchers among the 395 survey participants. Also, the 29 participants in the section of the survey that relates to postdoctoral students represent all research disciplines. Because of the distinct nature of the HEP community (discussed in Chapter 3), the relevance of this part of the RIN findings to the present study is negligible.

Hemminger et al. (2007) describe a survey that took place at the University of North Carolina at Chapel Hill in spring 2005. The survey's goal was to quantify the transition to electronic scholarly communication and the impact of this transition on various aspects of information seeking. Of the 902 academic researchers who participated, 34 were researchers in physics or astronomy.

The survey analysis shows that the type of materials sought by researchers changed from the types indicated in earlier surveys. Participants listed journals, Web pages, databases, and personal communication, in that order, as the types of resources that they use most frequently. Preprints were found to be used much more frequently by researchers of basic sciences than by medical researchers. The survey indicates that the search tools that the researchers use the most are bibliographic and citation databases (47%) and Web search engines (30%) and that there is a strong preference for obtaining materials electronically and at no cost. Furthermore, 'many comments in the survey indicated a strong preference for a single "meta" search tool where the user could enter a single search string that would result in all content in all resource collections being searched, as opposed to manually identifying resource collections and individually searching them' (Hemminger et al. 2007, 2214).

Hemminger et al. (2007) point out that 'as free, Web-based literature databases such as Google Scholar continue to grow, the distinction between bibliographic/citation database and Web search engines is blurring' (Hemminger et al. 2007, 2210). This conclusion corresponds with findings described in Chapter 3.

Another observation from the University of North Carolina survey that is relevant to the present study is that all researchers—regardless of their professional experience—'are increasingly using simple single text box search interfaces such as those provided by search engines like Google' (Hemminger et al. 2007, 2214).

Haines et al. (2010) describe a study of the information-seeking practices of basic science researchers at the University of Vermont College of Medicine, conducted in 2008. Nine participants from several basic science disciplines were interviewed regarding five main topics: the information sources they use, the search techniques they employ, their working environment, the library services they use, and library services that they would consider using.

The survey results include the following:

- All participants expressed a strong preference for online materials.
- Most participants reported starting their searches with either PubMed or Google, depending on the nature of their information need.
- To find information about a given topic, researchers look for background information such as review articles, books, book chapters, and presentations. Google is mentioned as a starting point by four participants.
- Primary literature is gathered for specific information on a topic; almost all participants use PubMed for locating the required materials.
- Following references and citations and asking an expert were listed as methods of finding specific materials.
- Participants use the simplest query form available; only one of them mentioned using advanced search features.
- Participants use alert services to keep abreast of new materials.

Interviews with medical researchers conducted by this author (presented at a City University London seminar in April 2008) and the author's in-depth discussions at the United States National Library of Medicine (NLM) in January 2008 suggest that the role of PubMed for medical researchers is similar to the role of SPIRES for

HEP researchers. In both cases, the information system includes most, if not all, of the information that the researchers require—in the form of metadata only—and is freely available to users. Furthermore, both systems are community based (PubMed is provided by NLM), and their user interfaces are the object of much attention by the teams that manage them. However, the HEP community and the community of medical researchers and practitioners differ considerably. Further discussion about the special characteristics of the HEP community and the way in which it differs from other communities is provided in Chapter 7.

From the descriptions that the participants in the University of Vermont survey gave of their search processes, ‘it became clear that searching bibliographic databases was a small part of conducting a literature review, a finding similar to behaviors described by early models of online search behavior’ (Haines et al. 2010, 77) and that researchers are likely to interact with colleagues in the institution and elsewhere to find the information they need. Furthermore, the authors of the study conclude that ‘basic science researchers valued and relied heavily on their community and desired even more collegiality’ (78).

Because the ultimate goal of the University of Vermont study was to support the design of a suite of library services that would better meet the needs of the library’s users, most of the other conclusions of the study relate to the role of the library and are therefore not discussed here.

In 2006, a study at University College London (UCL) examined the information-seeking behaviour of 114 scholars—students and faculty—in the Department of Physics and Astronomy (Jamali and Nicholas 2008). The study focused particularly on the scholars’ practices for keeping up to date and for finding articles, as well as on the influence of academic status and intradisciplinary dissimilarities on the surveyed researchers’ behaviour.

In the introduction to the report, the authors comment that physicists are ‘renowned for their information prowess’ (Jamali and Nicholas 2008, 445); the authors also pay tribute to the community for having ‘played a significant role in scholarly communication and publishing, especially in areas such as e-print culture and electronic publishing’ (Jamali and Nicholas 2008, 445) and for being trendsetters in the area of scholarly communication. Physicists and astronomers were selected as the subjects of the survey because the authors acknowledge the importance of efficient information systems to the community's research capabilities.

Findings of the research show that 'the majority of respondents believed that it was important for them to keep up with the developments of their subfields' (Jamali and Nicholas 2008, 448) and that they use various tools for keeping abreast, primarily personal communication, e-mail alerts, and lists of new submissions. Older researchers are more likely to rely on personal communication, while younger ones tend to rely on e-print archive services, with HEP researchers using these services more extensively than members of other groups.

For searching, the survey participants reported using various methods and information systems, with searching in Google the most popular method used on a daily basis (18%), followed by searching in subject databases (11%), browsing or searching on journal Web sites (9%), and looking up references provided in articles (8%). However, when asked about how they found the article that they had read most recently, the researchers gave responses that paint a different picture: 35% reported finding the most recently read article through recommendations from colleagues; 20% followed a reference trail; 13% searched in a database; 10% searched in Google; and 5% browsed through e-journals tables of contents.

Jamali and Nicholas (2008) reach the following conclusions:

> Physicists in HEP relied mostly on searches in subject databases (arXiv.org) for identifying articles they read. The second most used method was searching in Google. The fact that Google was the second used means by which articles were found in the field of HEP might be because of high availability of open access material in HEP that makes everything searchable by general search engines such as Google. In order to keep up-to-date with the developments in HEP they mainly depended on browsing e-print archives, word of mouth and meetings. (Jamali and Nicholas 2008, 460)

A survey of HEP researchers conducted in 2007 (Gentil-Beccot et al. 2008) and the author's interviews with HEP researchers (see Chapter 3) show that HEP researchers search in SPIRES and not in arXiv, as Jamali and Nicholas assert.

Furthermore, because these authors do not refer to findings related to SPIRES (even though there is a reference to 'one of the, apparently, most efficient information systems' [Jamali and Nicholas 2008, 445] without giving it a name), it appears that they do not fully grasp the difference between arXiv and SPIRES. The blurred distinction between Web search engines and bibliographic and citation databases noted by Hemminger et al. (2007) may have also influenced the responses of the survey participants and the interpretation of these responses.

The survey that is the focus of the present study took place early in the summer of 2007 (Gentil-Beccot et al. 2008). Over 2,100 responses from HEP researchers were analysed first by the team of HEP information specialists who conducted the survey and then by the author of the present study. The survey and results are described in detail in Chapter 3.

Several other user studies that focus on researchers were examined for the present research. However, because of the rapid change in the information-seeking behaviour of scholars, some of their findings (from the 1980s, the 1990s, and the beginning of the twenty-first century) are not necessarily indicative of the situation today.

The first of these studies was conducted in 1998 (Brown 1999) with the aim of examining the information-seeking behaviour of scientists at the University of Oklahoma and the way in which they were responding to changing scholarly communication practices such as the provision of materials in electronic form. The responses of 49 faculty members in three disciplines—biochemistry, mathematics, and physics/astronomy—revealed the following practices:

- All three groups of scientists listed textbooks and, in second place, monographs, as the tools they use for teaching. Physicists also listed journals and preprints, in that order.
- Physicists listed journals, preprints, monographs, conference proceedings, and textbooks, in that order, as the tools they use to support their research activities. These findings are similar to those of the other two groups of scientists.
- The three members of the HEP community that took part in the survey indicated the use of arXiv and SPIRES as sources for information.
- Almost all the participants reported using the references at the end of articles to find previously published information.

- Fewer than 50% of the respondents used electronic subscription journals, and about two-thirds indicated print journals as their preference; about a quarter preferred the electronic version, and the rest indicated both print and electronic forms as their preference (but added ‘printability’ as an important feature of electronic journals).

These findings from the 1998 survey, and specifically the preference for obtaining materials in a print format, highlight the drastic changes that have occurred in the last decade, as evident from later surveys.

A study of the information-seeking behaviour of scientists engaged in interdisciplinary research is described by Murphy (2003). The study took place at the United States Environmental Protection Agency campus in Research Triangle Park, North Carolina. Of the 149 participants, whose areas of specialization cover a broad range of subjects, approximately 89% indicated at least two subject areas with which they need to be familiar for their research. While interdisciplinary scientists share certain information-seeking practices with other communities—for example, reliance on personal networks for information gathering—the locating and understanding of relevant materials present challenges that are unique to interdisciplinary researchers, such as the need to be familiar with area-specific information resources, jargon, and the context in which research has been done in two or more fields. As a result, the majority (85%) indicated that keeping up with literature is challenging and that they need to invest time after work hours and obtain assistance from others to keep abreast of new developments in their fields.

In Markey’s (2007a, 2007b) two-part series of articles, ‘Twenty-five years of end-user searching’, she first examines literature describing intervention-free studies of users’ searches (that is, studies based on analyses of search log files); the second article in the series looks at the research findings in the context of information-retrieval models.

Because Markey’s review spans 25 years of studies, some of which relate to systems that disappeared years ago, the findings should be viewed with some caution. Nevertheless, some of the findings are noteworthy:

- Users enter short queries—one or two words—when searching in library catalogues, and two to three words when searching with Web search engines. The longest queries—four to eight words—were 'request- and question-format queries sent to Web search engines' (Markey 2007a, 1073).

- o The use of Boolean operators is rare (less than 20%). Fewer than 15% of the searches include the operator AND, fewer than 3% use the operator OR, and fewer than 2% use the operator NOT.
- o Advanced search options in addition to Boolean operators are used even more rarely than the latter. Fewer than 15% of the searches include phrases enclosed in quotation marks, fewer than 5% include truncation symbols, and nested Boolean logic is 'uncommon' (Markey 2007a, 1077).
- o Most users are happy with the default values of the information systems and do not change them.
- o Very few users scan more than two pages of results.

In the second article of the series, Markey (2007b) looks at her findings in the context of three specific information-retrieval models that 'recognize that information retrieval is not a one-stop event—people search repeatedly for the same topic of interest and their searches involve changes in cognition, feelings, and/or events of the information seeking process' (Markey 2007b, 1123).

The first model discussed by Markey is the berrypicking model proposed by Bates (1989) (for more information, see 2.2.2.5). In Markey's words, the Bates model acknowledges that 'retrieval is not a direct route from information need to final retrieved set. Instead, the search changes direction, pauses, and meanders as the user reads retrieved documents, follows up on leads, and responds to shifts in thinking. New information gives information seekers new ideas, new directions to pursue, and a new conception of their information needs' (Markey 2007b, 1123).

The second model is Kuhlthau's (1991) (for more information, see 2.2.2.4) information search process model, which breaks the information seeking process into stages and describes the way information seekers move from one stage to another. As Markey (2007b) points out, 'when users are engaged in an extensive inquiry project, their thoughts evolve from vague and unclear to focused and personalized; their actions change from general and exploratory to specific and comprehensive; and their feelings emerge from uncertain and hesitant to interested and directed' (Markey 2007b, 1123).

Markey goes on to examine a third model, the multiple information seeking episodes model presented by Line and Belkin in 2000 (for more information, see 2.2.2.6), which 'explains why people search repeatedly for the same information need across multiple episodes' (Markey 2007b, 1124). According to Markey (2007b), this model describes the information seeking process through an information need that changes, evolves, or even is replaced by another need during the process, causing corresponding changes in user behaviour.

Markey (2007b) focuses on the information-seeking process that involves repeated searches and raises questions about the changes in user searches during the process and the way in which information systems support repeated searches. She notes that 'although answers to these research questions will be of interest to IR researchers, their value to system designers should be paramount. If there are stark differences between end-users' initial and subsequent searches, system designers should take notice because it may be an opportunity for them to make their systems more adaptive to user needs' (Markey 2007b, 1125).

Taking the simplistic approach adopted by users as a given, Markey (2007b) discusses system features that could help users locate the most relevant materials. She points out that the major reason for failed searches is the initial choice of query terms: users typically start with broad and often inaccurate concepts and then tend to narrow down the results by using more specific terms or by employing 'a variety of flip-flopping[4] patterns' (Markey 2007b, 1125). One way, then, of helping users is by providing them with search suggestions based on an inspection of their queries. However, Markey (2007b) explains, 'what information systems report back to users should make for easy reviewing and instantaneous recognition of relevant possibilities....Much experimentation is needed to determine when users want intervention, the types of intervention they will tolerate, and their preferred reporting formats' (Markey 2007b, 1126). The Bates model, describing three levels of search activities (moves, tactics, and

[4] Markey defines flip-flopping as 'following up an initial broad-based concept with a specific term, flipping back to a broad-based concept, flopping to a specific term, adding a new term to express a narrower concept or an entirely new concept, and so on' (Markey 2007a, 1125)

stratagems) is suggested by Markey as a starting point for considering system intervention.

Nevertheless, Markey (2007b) states:

> I have come to believe that making online IR systems more complicated with additional functionality, frequent and unanticipated interruptions in the form of direct system intervention, and detailed instructions and tutorials in system use, is not the right way to proceed....Let us use what we learn from these observations to build future systems that covertly teach and advise at the same time their users conduct business.' (Markey 2007b, 1128)

Addressing future research, Markey concludes with a statement that is relevant to the present study:

> Finally, I invite IR system designers to put future research findings to work by building systems that are sensitive to the progress users are making in their ongoing searches, intervene with complex search features that are likely to solve user problems, and monitor users to determine whether these complex features help them achieve their goals. (Markey 2007b, 1129)

Markey's findings regarding the queries that users formulate are supported by research that was conducted as part of the initial research for the present study. Before focusing on the HEP community, the author examined searches logged in two information systems and presented her findings at City University London in April 2008. The two systems that were monitored were the journal search provided on the Web site of the Institute of Physics (IOP) and SearchHUB, the metasearch[5] system of Cranfield University, both facilitating searches for scholarly publications.

The analysis of an IOP log file created during November 2006 shows that scholars are drawn to simplicity: users of the IOP Electronic Journals search service tend

[5] *Metasearching* (also referred to as integrated searching, simultaneous searching, cross-database searching, and federated searching) refers to a process in which a user submits a single query to heterogeneous information systems. The metasearch system facilitates the interaction with the user, broadcasts the user's query to each information system, and returns the results to the user.

to submit simple queries, almost always consisting of one word or one phrase (91% of the logged queries). Very rarely do the users add the truncation symbol (2% of queries). Furthermore, although the IOP site added a clustered search feature that can help users to focus on the relevant results when scanning long result lists, only 0.2% of the searchers in the sample took advantage of this new feature, and even fewer were curious to find out what it is.

The query log from February 2007 suggests that scholars at Cranfield University tend to use the search interface of SearchHub as is, relying on the default search options or changing them only to a small degree. The queries themselves are simple and short, consisting of a mean of 2.67 words. The users do not tend to take advantage of advanced search features such as wildcards or Boolean operators other than AND. When users conduct a search in specific metadata fields, they usually look at the title, but they often look at the author and subject fields as well. Users select the date of publication only in conjunction with another search term, such as author name, and do not search in ISSN or ISBN metadata fields at all.

2.4 User-Interface Design

Although many aspects of user interface are crucial to the design of a new user interface for a scholarly information system, it is believed that one of the fundamental challenges of today's information systems relates to their ability to help users sort out long result lists. Literature that addresses the 'findability' of items, in general, and particularly the post-search processing of result lists, is reviewed in this section.

Morville (2005) defines the term *findability* as follows:

> a. The quality of being locatable or navigable.
> b. The degree to which a particular object is easy to discover or locate.
> c. The degree to which a system or environment supports navigation and retrieval. (Morville 2005, 4)

Current scholarly systems, as well as popular search engines, present search results in a one-dimensional list, regardless of the number of items in the list.

However, with the huge amount of data that is available today (which is growing exponentially) and the indifference shown by typical users toward the phrasing of their search queries (Hemminger et al. 2007, Markey 2007b, and CIBER/UCL 2008, among others), many searches in the scholarly environment generate a large number of results, thus decreasing the findability of items that are not at the top of the list. A purely linear presentation of so many results is inadequate, despite the use of sophisticated relevance-ranking algorithms to prioritize the result list. When applied to scholarly materials, these algorithms are questionable in that they lack the context in which the query was defined and hence cannot assist the system in tailoring the presentation of the results to the specific person's needs. A further problem is that relevance-ranking algorithms alone cannot bridge the gap between a user's intended query and the way in which the user phrases it.

To increase the findability of specific items buried in the list, the proposed information system should employ various methods, as suggested later in this study (Chapter 6). For example, the system can analyse the results and then suggest that the user search the whole collection again for one of the topics associated with the result list. If a search yields a large number of results, the system can enable the user to drill down to subsets of the result list (for more information, see Chapter 6).

What these options have in common is the idea that the result list itself is a tool that can be exploited. One of the ways in which the system can exploit the list is by analysing it and presenting it as a multidimensional structure that corresponds to common methods of classifying information; users can then examine the list from various angles, gaining a better understanding of its content and its relevance to their query.

In his discussion, Morville (2005) explains that classification coheres with the way in which people perceive information:

> We classify to understand….In a formal taxonomy…properties flow from class to subclass through the principle of inheritance. Each object and category is assigned a single location within the taxonomy. We live at an address within a nested hierarchy of streets, cities, states, and countries. We exist as *Homo sapiens*

> within the taxa of domain, kingdom, phylum, subphylum, class, order, family, genus, and species. (Morville 2005, 7)

Furthermore, argues Morville, classification provides the flexibility that helps searchers grasp the rich and complex nature of the information that is now available to them; and the disengagement of library objects from their physical location enables systems to arrange elements in more than one way. Morville (2005) continues:

> Of course, the world doesn't always cooperate with this Platonic approach to classification. Fish with lungs….Words with many meanings. Meanings with many words. Reality confounds mutually exclusive classifications, and so we find ourselves debating which existing category works best or defining new categories to allow a perfect fit…
>
> We embrace faceted classification…using multiple fields or "facets" to describe the objects within our collections. First defined in the 1930s by Indian librarian S. R. Ranganathan, faceted classifications have flourished in digital domains, where objects can exist simultaneously in many locations. (Morville 2005, 127-8)

In their introduction to the evaluation of the way information systems provide innovative searching options, Wilson et al. (2009) point out that keyword search—the 'Google-like' search—has become the standard for new information systems. This kind of searching does not provide good tools for users when their goals are not well defined, when their requests for information are complex, and when their knowledge of the subject matter is inadequate. Faceted browsing and clustering, described in Chapter 6, are two technologies that are being added to information systems to help users 'filter and navigate through information using recognition rather than recall' (Wilson et al. 2009).

Wilson et al. (2009) propose using a combination of information-seeking models for building an evaluation framework, aiming to quantify the strengths and weaknesses of the design of three systems. Their findings, presented as recommendations for user-interface design, include the following:

- Maintain keyword search.
- Enable users to change selections and make multiple selections of facets.

- Show the hierarchy of the selections made, to describe the current context.
- Show the results along with the selections.
- Provide sorting and filtering options ('The sorting and filtering of lists is an important part [of] finding and organising data. The ability for the user to arrange data and results so that they can effectively find the information they want supports a number of key tactics' [Wilson et al. 2009]).
- Facilitate information gathering.
- Provide initial display (preview) of results to help users make decisions at an early stage.

These recommendations were taken into account for the user interface described in Chapter 6 of this thesis.

Hearst (2006) discusses faceted browsing and clustering as ways for an information system to group results meaningfully 'in order to help make sense of the results, and to help decide what to do next' (Hearst 2006, 59). Hearst explains the differences between these two technologies and provides evidence of the change in users' behaviour once grouping mechanisms became available to them.

Hagedorn et al. (2007) suggest a way to combine the easy searching of Web search engines with the rich and multifaceted structure of scholarly data to create a simple interface to complex online information systems. Their paper describes a joint project of the University of Michigan and the University of California, Irvine. The project team built a mechanism for generating clusters from metadata records (originating from various OAIster[6] databases) and labelling the clusters. The labels were matched against a classification system. The team then created a prototype user interface to the corpus of data, deploying a clustering mechanism. Cluster labels were used for both providing browsing functionality and narrowing down search results. Findings show that 'the real power of including new subject terms [cluster labels] was on the search results page' (Hagedorn et al. 2007),

---

[6] OAIster is a union catalogue of digital resources. Metadata records in OAIster are harvested from multiple digital information systems via the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH).

and the general outcome is described as an improvement of the findability of relevant materials.

Kaki (2005) describes a study that explored how faceted categorization is used in real settings (as opposed to an earlier study that was a controlled experiment) and how beneficial it is to users. A Web-based search interface called Findex was provided to 16 users for two months, for any use that the users had in mind. The findings of this study show that in 25% of their searches, users took advantage of the categories displayed alongside the result lists to find relevant results. On average, 2.3 categories were selected per search process. Participants reported that the categories were most useful when queries were general, broad, or vague, such as those that characterize exploratory searches; the ranking algorithm failed to display the required items at the top of the list, but the categories helped users find results nevertheless. Evidence also shows that the use of categories facilitated the selection of multiple results, once again assisting users in exploratory searches more successfully than the sorting order of the result list.

2.5 Personas

Chapter 4 discusses the creation of personas to be used in the evaluation of an information-seeking model and a user-interface design proposed in Chapter 5 and Chapter 6, respectively. Personas come from the field of user-interface design. Introduced by Alan Cooper (1999) as part of his goal-directed design methodology, a persona is described by Blomkvist (2002) as ‘a model of a user that focuses on the individual’s goals when using an artefact’ (Blomkvist 2002, 1). The use of personas was adopted by user-interface designers in the late 1990s and later examined in both academic and commercial environments (Blomkvist 2002; Perfetti 2002; Randolph 2004; Dantin 2005). Research suggests that the process of using personas ‘helped introduce clarity and a form of accountable reasoning into the UI evaluation process’ (Dantin 2005, 7). Blomkvist (2002) discusses the creation of personas and proposes the use of a set of guidelines; these guidelines, along with those suggested and described in detail by Mulder (2006), were implemented in this work.

A more detailed discussion of research surrounding the use of personas and their applicability to this study is provided in Chapter 4.

2.6 Information Systems and Scholarly Communication in the High-Energy Physics Community

The high-energy physics (HEP) community has a long tradition of innovation in providing means for researchers to access scientific information, publish their work, and take part in community-based initiatives. HEP initiatives have been discussed in the literature since the establishment of the SPIRES suite of HEP databases in the 1960s (described by Addis 2002). The creation of arXiv in 1991 was described in a retrospective article by the arXiv's founder, Paul Ginsparg (2008).

HEP information specialists employed by Cornell University and by the major HEP centres—the European Organization for Nuclear Research (CERN), the Stanford Linear Accelerator Center (SLAC), Deutsches Elektronen-Synchrotron (DESY), and the National Accelerator Laboratory (Fermilab)—dedicate their time to maintaining and improving the HEP information systems, primarily SPIRES and arXiv, as is evident in the information specialists' publications (Kreitz and Brooks 2003). Information about the content of the information systems and about new developments is published occasionally by the hosting institutions (Cornell University Library 2008; DESY 2008; CERN 2010) and on the Web sites of the information systems (arXiv, http://arxiv.org/Stats/hcamonthly.html; INSPIRE, http://www.projecthepinspire.net/).

In her review of scholarly communication of the HEP community, initially published in 1965, Goldschmidt-Clermont (2002) examines the communication patterns of the HEP researchers and the techniques that were developed by the community to address the information needs of these researchers. Robbins (2007) reviews SPIRES ('the Mainstay of High-Energy Physics', as indicated in the review title) and the practices involved in maintaining it. Jamali and Nicholas (2008), as mentioned earlier, address the HEP community's information-seeking behaviour.

Theoretical discussions of HEP scholarly communications are described by Aymar (2008), Heuer et al. (2008), Gentil-Beccot et al. (2008), and Jamali and Nicholas (2008). Gentil-Beccot et al. (2008) also present the initial results of a survey that the information specialists from HEP institutions conducted in 2007.

A more detailed discussion of HEP-related literature and the way it relates to this research is provided in Chapter 3.

Chapter 3 Information-Seeking Behaviour in the High-Energy Physics Community

3.1 Introduction

The high-energy physics (HEP) community has a long tradition of innovation in providing means for researchers to access scientific information, publish their work, and take part in community-based initiatives. The main information systems used by the community are the SPIRES suite of HEP databases and the arXiv e-print service in physics and other fields of science. Both resources offer innovative services and are free, in the spirit of the community's support for open access to knowledge. According to the approximately 2,100 responses to a survey that information specialists from HEP institutions conducted in 2007, the great majority of the community members (88%) use SPIRES, arXiv, or both as their primary information system (Gentil-Beccot et al. 2008). This chapter describes the information-seeking behaviour of the HEP community as inferred from published results of the information specialists' survey, conclusions from the author's analysis of the survey responses, and the author's interviews with HEP researchers.

September 2008 saw a flurry of attention focused on high-energy physics (also referred to as particle physics), when the Large Hadron Collider (LHC) was started up at CERN, the European Organization for Nuclear Research. The launch of the new experiment was the culmination of an intense collaborative effort of tens of thousands of high-energy physics scientists from all over the world. According to the LHC project leader, Lyn Evans, the project aims to usher in 'a new era of understanding about the origins and evolution of the universe' (CERN 2008). This unparalleled international scientific project sheds some light on the special nature of the community of HEP researchers, particularly the information systems on which the community members rely and the information-seeking behaviour that they have developed for conducting their research.

Besides the high level of their scientific expertise, HEP scientists are characterized by a strong sense of community affiliation and willingness to contribute their time, knowledge, and skill to community-based initiatives (Kreitz and Brooks 2003; Robbins 2007; Gentil-Beccot et al. 2008). Decades before the terms *Web 2.0* and *social networks* were coined and even before the World Wide Web as such was invented at CERN by Tim Berners-Lee, HEP researchers' need to efficiently carry out projects that had a rapid turnaround and that involved a large number of participants across the globe led them to develop new publishing and sharing practices and establish several information systems. The relatively modest size of the HEP community—about 20,000 people around the world (Gentil-Beccot et al. 2008)—and the fact that many researchers collaborate on projects that are based in just a few regions undoubtedly contribute to the members' strong sense of ownership regarding the quality and innovation of the information systems.

3.2 HEP Community-Based Information Systems

The two major information systems that serve as the backbone of research for the HEP community are the SPIRES suite of HEP databases and the arXiv e-print service in physics and other fields of science.

The SPIRES suite was started in the 1960s at the Stanford Linear Accelerator Center (SLAC), was computerized in 1974 by teams from SLAC and Deutsches Elektronen-Synchrotron (DESY), and is today a joint project of SLAC, DESY, and the National Accelerator Laboratory (Fermilab) (Addis 2002). For more than three decades, SPIRES has been providing access to the literature, people, institutions, research, and experiments in the fields of particle and astroparticle physics. The first service in the United States to implement a Web server, SPIRES 'has helped lead the transition from a totally print-based system to an almost totally electronic-based system. In the process, it has expanded to provide worldwide subject-specialized access not only to the field's journal literature (as do database vendors), but to a wider set of information objects comprising a significant amount of the intellectual "ecology" of the field' (Kreitz and Brooks 2003, 6). Operated as a community-based resource, SPIRES 'is not an effort that is particularly well-funded. In fact, it operates only through a careful use of every (automated, cost-lowering) software program it can implement, a judicious use of hands-on intellectual oversight and cataloging, an aggressive commitment to collaborative and consortial information sharing, and—most radically and uniquely—the volunteer efforts of many of our users' (Kreitz and Brooks 2003, 6). In recent years, the overuse of ad hoc labour-saving devices has led SPIRES to software paralysis, a situation that drove HEP information scientists at CERN, SLAC, DESY, and Fermilab to choose the CDS Invenio digital library technology developed at CERN as their new technological infrastructure. The development of a new service, called INSPIRE, was already under way in 2008 (DESY 2008), and in March 2010, HEP community members began testing it (CERN 2010).

The arXiv e-print repository, another example of a community-based initiative that predates the Web, was established at the Los Alamos National Laboratory in 1991 by Paul Ginsparg, whose aim was to create 'a centralized automated repository and alerting system, which would send full texts only on demand. That solution would also democratize the exchange of information, levelling the aforementioned research playing field, both internally within institutions and globally for all with network access' (Ginsparg 2008). In 2001, Ginsparg joined the faculty of Cornell University, taking arXiv with him, and since then, the repository has been owned and run by Cornell University. In October 2010, the repository had over 630,000 postings of research articles in physics, mathematics, statistics, computer science, and quantitative biology, all published online. Figures published on the arXiv site indicate that although arXiv was initially created for HEP literature and is extensively used by HEP researchers, the arXiv model proved to be beneficial for other disciplines (Figure 11). At the end of December 2009, only 23.5% of the submissions in arXiv were HEP materials and the submission rate of the HEP community was 14.1% of the total submission rate (arXiv, http://arxiv.org/Stats/hcamonthly.html).

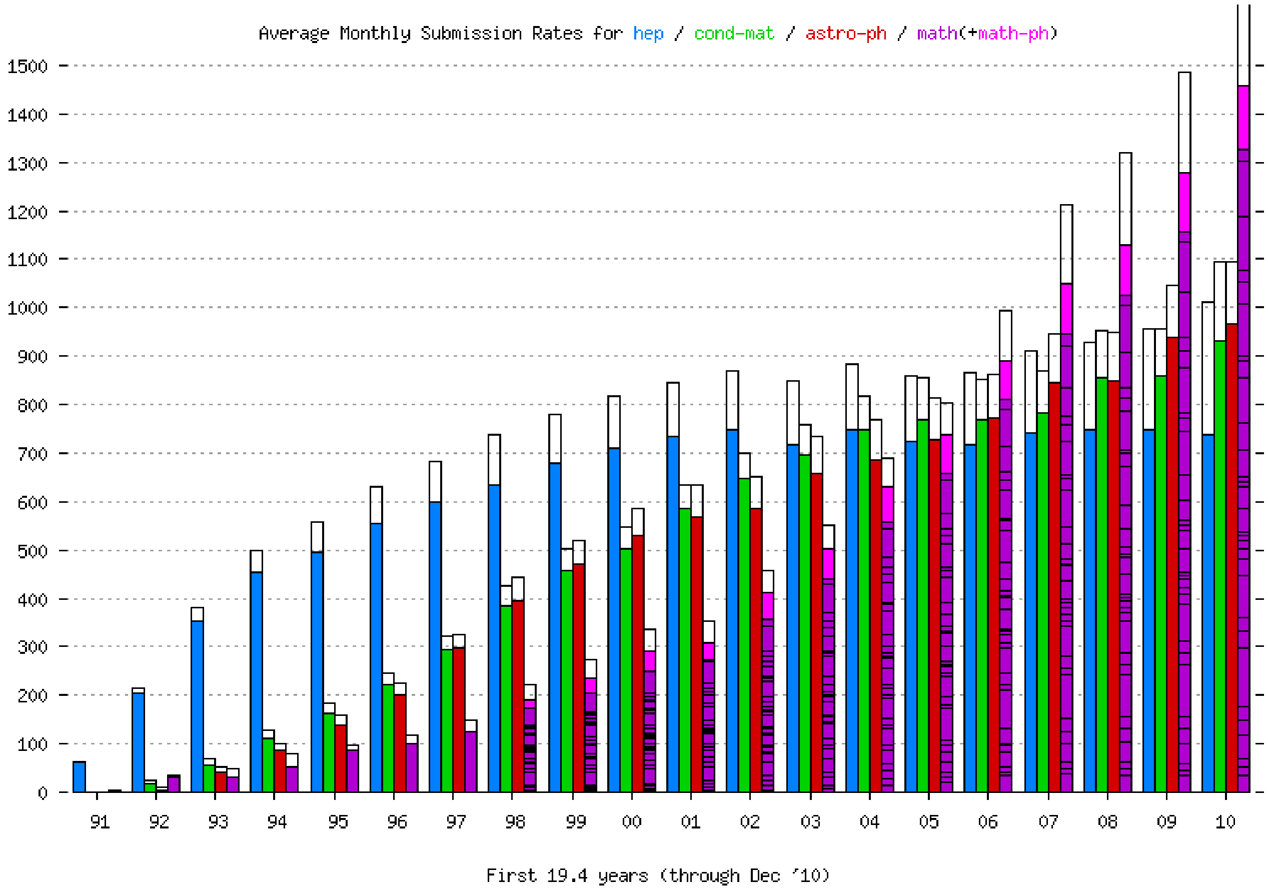


Figure 11: Average monthly submission rate to arXiv, by discipline, 1990-2009. HEP submissions are indicated in blue. Source: http://arxiv.org/Stats/hcamonthly.html.

Researchers upload their own articles to arXiv, along with structured metadata, and after some screening by volunteer moderators, the articles are made freely available to researchers, typically within a day. In this manner, the community gains access to the preprint[7] version of an article before the article is published in peer-reviewed journals, resulting in 'immediacy of dissemination without barriers' (Aymar 2008). Later, authors can submit revised versions of the preprint and even the post-peer-reviewed, author-formatted version of the article. In October 2008, Cornell University Library reported that 'more than 200,000 articles are downloaded from arXiv each week by about 400,000 users, and its 118,000 registered submitters live in nearly 200 countries' (Cornell University Library 2008).

[7] Preprints are defined as 'near-printed copies of manuscripts submitted for publication in journals. They are temporary documents whose function is to bridge the time-gap created by publication delays' (Goldschmidt-Clermont 2002).

The results of a survey that took place in early in the summer of 2007 (Gentil-Beccot et al. 2008) and interviews that the author conducted with HEP scientists clearly show the huge importance that HEP researchers attribute to SPIRES and arXiv and their pride in these resources. As noted by the team of information specialists (from HEP institutions) that conducted the survey, even the overwhelming response to the survey—about 10% of the community members—'is *per se* a result, signifying the engagement of the community with its information resources' (Gentil-Beccot et al. 2008, 9).

3.3 HEP-Related Research for This Thesis

The prominence of SPIRES and arXiv as research tools for the HEP community forms an antithesis to the common perception that academics are moving away from complex search interfaces and toward Web search engines. To better understand the 'magic' that attracts the HEP researchers, the author interviewed several HEP scientists and analysed the results of one section of the HEP survey. The information-seeking behaviour patterns that emerge from the survey and from the interviews shed some light on the factors that determine the success or failure of a search process and on the effects of the human-computer interface on a user's search experience. This chapter focuses on some of the issues that the HEP information specialists' analysis revealed and describes findings from the author's analysis. These findings fall into two main areas: the HEP scientists' perceptions about their information-seeking practices and the information systems that they use; and the influence of the special nature of the HEP community and the SPIRES and arXiv systems on the information-seeking process of the community's members.

3.4 Methodology

To gain some understanding of HEP community members' views of the information systems that they use, the author interviewed three senior researchers (two experimental physicists and one theoretical physicist) about their information-seeking practices and their preferences regarding the HEP resources. Later, four researchers—three PhD candidates and one postdoctoral scholar—were interviewed to complement the qualitative data and include researchers who are less familiar with the present and past literature than more seasoned researchers yet spend more time looking for scholarly materials. These semi-structured interviews took place in 2008 and the beginning of 2009, at the offices of the HEP researchers at a leading scientific institution. A detailed description of the interviews, including the questions asked, is provided in Appendix A . The interaction with some of the researchers continued on a less formal basis in the form of e-mail correspondence and occasional meetings.

Next, the author reviewed the responses to a survey conducted by HEP information specialists (the survey is described in detail by Gentil-Beccot et al. 2008). Running from April 30, 2007, to June 11, 2007, the survey collected 2,115 responses from HEP researchers in more than 35 countries.

In one section of the survey, the respondents answered multiple-choice questions. In another section, the respondents were asked to select the 'HEP information system' that they use the most; they were presented with a list of predefined options—arXiv, CDS (CERN Document Server), Google, Google Scholar, and SPIRES—as well as a text box in which they could type a different choice (such as the Astrophysics Data System, ADS). Then the respondents wrote answers to the following four open-ended questions:

a. 'Why do you mostly use this system?'
b. 'What do you like the most and the least about arXiv?'
c. 'What do you like the most and the least about SPIRES?'
d. 'What do you like the most and the least about Google (Scholar)?'

Of the 2,115 respondents, most (2,018) did indicate a preferred information system (arXiv, 847; SPIRES, 1,017; Google or Google Scholar, 182; other, 77[8]) and answered at least some of these questions (a—1,849 respondents; b—1,314; c—1,200; and d—602). The information specialists who conducted the survey kindly provided the author with the answers to the open-ended questions.

---

[8] Respondents used the 'Other' text box to name an additional resource or list multiple preferred resources.

The answers to the open-ended survey questions were interpreted and analysed on the basis of an understanding of the nature of the HEP community and its information-seeking practices (as revealed in the interviews that the author had conducted earlier) and the published analysis of the survey results that relate to the multiple-choice questions (Gentil-Beccot et al. 2008).

Because of the manner in which the open-ended questions were phrased and the likelihood that the respondents, when composing their answers, considered other answers that they had given in the questionnaire, the replies do not necessarily reflect a comprehensive and systematic evaluation of the information system on the part of the respondents. Rather, the replies highlight patterns of the respondents' information-seeking behaviour, aspects of the information systems that the researchers find most important, and faults that the researchers find in the information systems that they use or avoid using.

By nature, replies to an open-ended question are not structured in any way and do not necessarily use predetermined terminology; therefore, a set of criteria were defined for this study to allow an analysis of the replies. According to those criteria, the replies fell into a positive group and a negative group. Because not all the respondents refer to the same issues in their answers, two sets of figures are usually presented in this chapter: the first figure is the percentage of respondents (out of those who answered the question) who commented on a specific issue, and the second figure is the percentage of respondents in this subgroup who commented positively or the percentage who commented negatively about that issue.[9]

This discussion of preferences and practices refers to both the results of the survey's multiple-choice section—marked with '[PA]' (published analysis)—and the author's analysis of the replies to the open-ended questions, marked with '[AA]' (author's analysis).

---

[9] For example, '26.9% of the researchers who responded to the question "What do you like the most and the least about arXiv?" referred to arXiv's coverage. Of these respondents, 68% indicated that they like the coverage, while 32% felt that the coverage is not sufficient.'

3.5 Researchers' Preferred Information Systems

In the electronic era, when information systems are interlinked, researchers often look upon a research environment as one entity, even when it consists of multiple information systems. Furthermore, because information is often available from more than one system or via more than one workflow, the information alone is not the only criterion that a researcher weighs in deciding where to seek information; the user experience and the services that are provided by an information systems can lead the researcher to favour one search path over another.

According to the published analysis of the HEP survey, an overwhelming percentage of researchers (87.9% [PA[10]]) marked one of the community-based resources—SPIRES or arXiv—as the information system that they use the most (48.2% and 39.7%, respectively [PA]) (Figure 12). Furthermore, the researchers' detailed responses to the open-ended questions clearly indicate that the majority use both systems, each for a different purpose. As one researcher explains, 'SPIRES is the most user friendly and comprehensive system for finding articles in HEP, whether they be published or in preprint form, and hence is the most useful research tool, while the arXiv is obviously the indispensable tool for providing us with new preprints on a daily basis'.

[10] Results of the survey's multiple-choice section are marked with '[PA]' (published analysis).

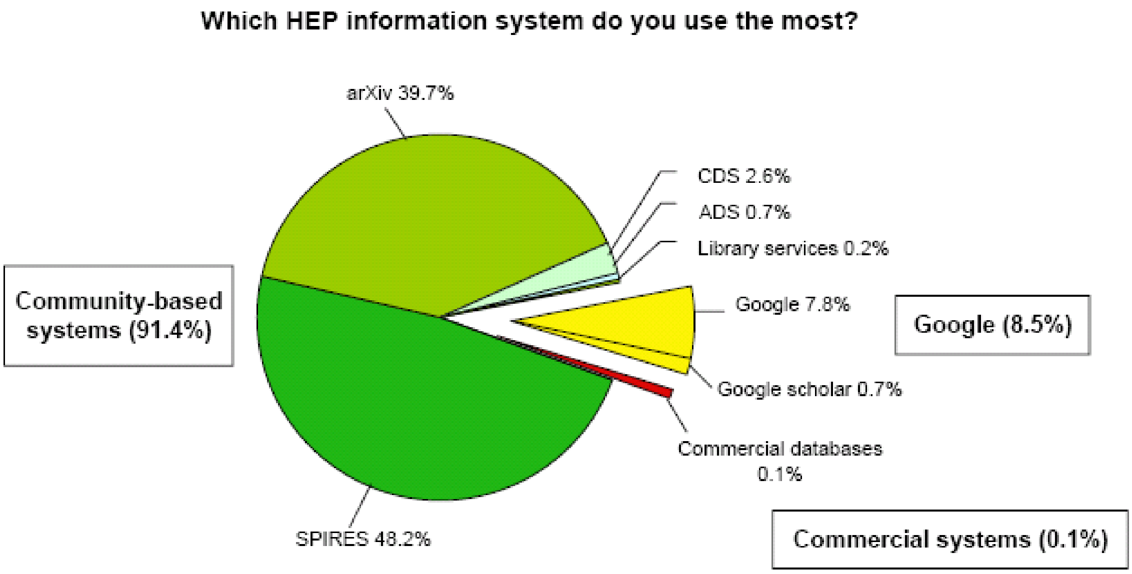


Figure 12: HEP information systems used most by HEP community members. Reprinted from Gentil-Beccot et al. (2008), Figure 1.

Google is next in popularity as the most used system (7.8% [PA]; Google and Google Scholar combined—henceforth referred to as Google services—8.5% [PA]), although when looking at its popularity according to the researchers' number of years in the field (which usually corresponds to their age), one sees that newcomers (22% of those who joined the field in the last two years [PA]) are more inclined to use the Google services (Figure 13).

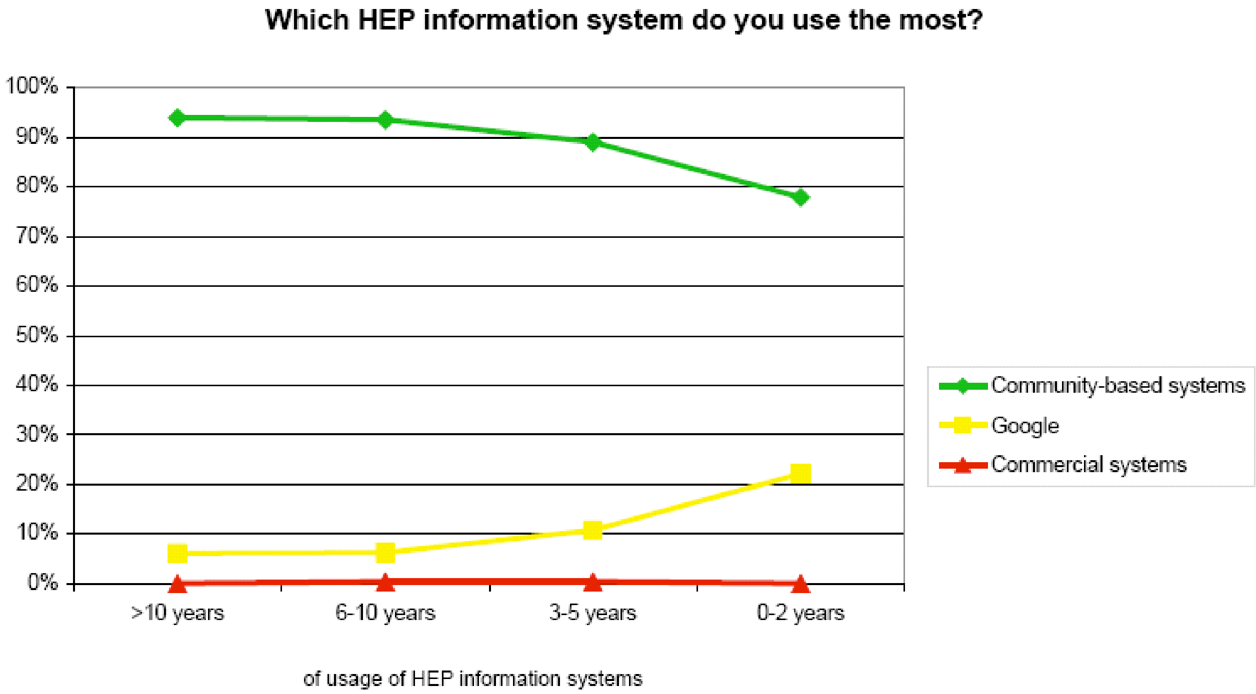


Figure 13: Percentage of HEP community using HEP information systems. Percentage broken down by the members' number of years in the field; based on the multiple-choice survey responses. Reprinted from Gentil-Beccot et al. (2008), Figure 2.

In their replies to the open-ended questions, 28% of all the respondents referred to their use of Google services (Google and Google Scholar) to some degree as a means of finding scholarly information (once again, the percentage varies by the number of years that the researchers are in the field and reaches 36.7% of the junior set) ([AA[11]]). Researchers commented about the comprehensiveness of the information-seeking process when they use all three resources: 'Papers are published daily there [in arXiv] and one can look for a given article easily by the arXiv number. SPIRES is better when the authors and/or exact title are known and GOOGLE is probably the best when one only knows roughly what they are looking for'. Or, as another researcher wrote, '…arXiv is essential to keep up with and to contribute to research in HEP, etc. SPIRES is essential for locating older items, for names, etc. For outside HEP and also as a complement search tool, and for searching for scientists' webpages Google and Google Scholar are essential'.

[11] Figures arising from the author's analysis of the replies to the open-ended questions are marked with '[AA]' (author's analysis).

Although this discussion focuses on the resources that HEP researchers use the most—SPIRES, arXiv, and the Google services—the published analysis shows that some respondents indicated CDS (2.6% [PA]) and the Astrophysics Data System (ADS) (0.7% [PA]), both of which are community-based information systems, as the information system they use the most. Commercial services are used most by only 0.1% of the respondents ([PA]). These impressive findings signify the scientists' unique reliance on the resources that the community has created and has maintained for decades.

3.6 Researchers' Information-Seeking Process

The information-seeking process is supported by four major components: the core research data (for example, articles and books); the metadata that describes the core research data (metadata in this sense is cataloguing information, such as the author, title, and publication date of a book or article); the indexing system and search engine; and the user experience—the interaction between the researcher and the system in terms of workflow, screen design, on-screen terminology, services that are offered at each stage of the search process, and more.

In the HEP context, we can see that core research data, if available electronically, is stored in several locations—mainly arXiv, publishers' sites, and institutional archives; and metadata is typically assigned by the provider of the repository where the item is stored or by another provider. For example, a preprint article that is stored in arXiv bears metadata that is mostly author generated and is available in arXiv. If the article is relevant to HEP, curated metadata is added in SPIRES and is made available to researchers through the SPIRES interface. Furthermore, the metadata and the full text of the article are indexed by Google and Google Scholar and are available through these services. If the article is eventually published in a journal as well, the publisher is likely to add metadata, to index both the metadata and the final full-text version, and to make the metadata and the full text available through the publisher's interface. Authors often add the post-peer-review version to arXiv, thus making the published version available not only through the publisher's interface. Researchers can choose the arXiv interface to locate the article; they can search for it through SPIRES and obtain the full text from arXiv or from the publisher's site; they can use Google or Google Scholar to locate the article and link to the full text; or they can use the publisher's site to search for and retrieve the article.

While some researchers keep to their preferred resource and workflow, many indicated in the HEP survey that they alter their behaviour according to the kind of material that they are looking for and their prior knowledge about the required item. For example, a researcher who is looking for an article that is stored in arXiv might use the arXiv interface and enter the arXiv ID (although many opt to enter the arXiv ID in Google or Google Scholar or use the appropriate command in SPIRES, leading them to the article in arXiv with just one more click); a researcher who knows only the name of the author is likely to use the SPIRES interface; and a researcher who is not familiar with a subject area is likely to consider using Google or Google Scholar. Another major factor that determines the path that a researcher takes is the degree to which the research material is associated with HEP subject matter. The fact that SPIRES focuses on particle and astroparticle physics is at times an advantage, because researchers using SPIRES know that they will not encounter any irrelevant results if they look for material in these areas, but at other times—such as when the researcher is looking for material that is interdisciplinary or tangential to particle and astroparticle physics—the SPIRES focus on HEP becomes a disadvantage. The type of material (such as published articles, preprints, theses, or conference proceedings) is another important criterion for the selection of the best resource, and researchers adapt their behaviour accordingly (Figure 14); for example, when looking for preprints, only 7.4% of the respondents use the Google services, but when looking for theses, the percentage grows to 32.6% ([PA], Gentil-Beccot et al. 2008).

	Community Services					Google		Commercial services		
	SPIRES	arXiv	CDS	ADS	Library services	Google	Google Scholar	Publisher website	Commercial databases	Fraction of answers
Which HEP information system do you use the most?										
	48.2%	39.7%	2.6%	0.7%	0.2%	7.8%	0.7%	0.0%	0.1%	99%
Which HEP information system do you use the most to find...										
preprints (known reference)?	46.0%	45.2%	2.9%	0.2%	0.2%	4.2%	0.9%	0.3%	0.1%	97%
articles (known reference)?	60.2%	20.6%	3.1%	2.1%	1.4%	6.1%	1.5%	4.5%	0.5%	96%
preprints on a given subject?	49.6%	34.2%	3.2%	0.8%	0.0%	9.7%	2.4%	0.1%	0.0%	96%
articles on a given subject?	55.2%	19.4%	2.3%	1.7%	0.3%	16.0%	3.9%	0.5%	0.7%	95%
preprints by a given author?	67.1%	24.0%	2.8%	0.5%	0.1%	4.2%	1.0%	0.0%	0.2%	96%
articles by a given author?	74.0%	12.1%	2.4%	1.8%	0.2%	6.7%	1.7%	0.3%	0.8%	96%
theses?	48.8%	10.9%	4.3%	0.9%	2.4%	27.2%	5.4%	0.0%	0.1%	83%

Table 5: Favorite information systems in general, first row, and for specific needs. The last column summarizes the fraction of respondents who answered these questions.

Figure 14: Breakdown of HEP scientists' favourite information systems. Reprinted from Gentil-Beccot et al. (2008), Table 5.

3.7 Researchers' Evaluation of Information Systems

The HEP survey sheds light on the way in which researchers evaluate information systems. Gentil-Beccot et al. (2008) describe in detail the importance that researchers attribute to features such as the access to full text, depth of coverage, quality of content, and search accuracy. The free-text replies provide further information about the way in which these preferences apply to specific information systems.

According to the author's analysis, the surveyed researchers addressed some or all of the following factors when evaluating the information systems that they use:

- The coverage offered by the information system
- The means of finding the information
- The effort required to find and obtain information
- User friendliness of the information system
- Related services offered by the information system

The rest of this section examines how the researchers evaluated each of these factors in the HEP survey. Note that the remaining figures in this chapter refer to the author's analysis ([AA]).

3.7.1 Coverage Offered by an Information System

Obviously, the content that researchers can find via an information system's interface is the most important factor when they evaluate that system. If the information that they are looking for cannot be located, no other factor will convince them to use the system. Metadata and full text can both be considered content, and depending on the context, one or the other is the focus of researchers' interest. For example, metadata is important when a researcher only wants to find out whether there are publications on a specific topic or by a specific author. Similarly, when an article that a researcher needs is not published electronically, the metadata can help the researcher obtain the article in print form or through a mediatory service.

In the HEP survey, the respondents addressed several issues related to the content of information systems. The issues fall into five groups.

Focus on HEP topics: HEP researchers are aware of the HEP coverage of arXiv, SPIRES, and the Google services. SPIRES covers all materials that are specific to high-energy physics; arXiv covers some other scientific domains as well but focuses on preprints; Google covers everything, regardless of the materials' relationship to HEP or to science in general; and Google Scholar, which presents scholarly materials but does not disclose the origin of the information that it indexes, is not limited to HEP (Figure 15). As explained earlier, a scientist's area of research and specific needs affect that person's view of the HEP focus as an advantage or disadvantage.

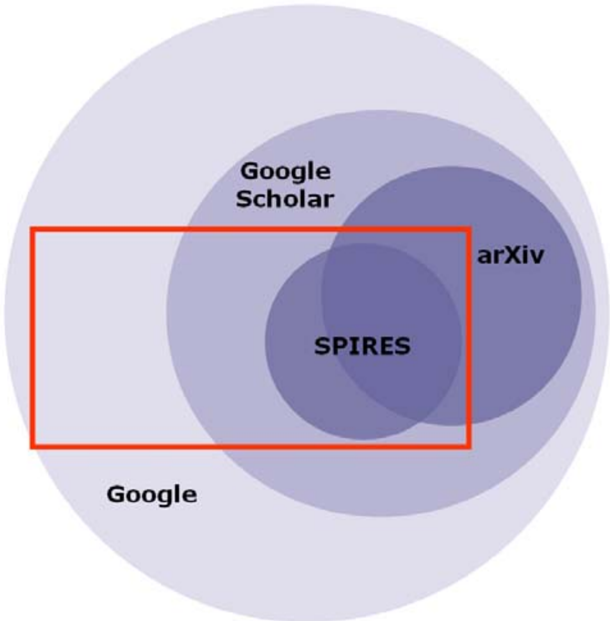


Figure 15: HEP-related information landscape.
The red box indicates HEP-related information that is available through SPIRES, arXiv, and the Google services.

In their replies to the open-ended questions in the HEP survey, 43.5% of the respondents referred to the coverage of SPIRES. Of these respondents, 80% noted that they like the SPIRES coverage in general. Of the 11.8% of the respondents who commented about the SPIRES focus on HEP literature, 62% said that they do not like this focus. Some regarded this focus as both an advantage and a disadvantage (for example, '[What I like the] most [about SPIRES]: good for hep; [what I like the] least [about SPIRES]: only for hep'). Those who felt that the focus on HEP is a disadvantage indicated that the coverage is poor or lacking in adjacent areas (e.g., cosmology, astrophysics, gravitational physics, mathematical physics, and condensed matter physics) and commented that today these areas are strongly connected to HEP. As one respondent put it, 'as research becomes more interdisciplinary (quark matter <-> superfluidity) it is necessary to have a search engine which harvests all arxiv subject classes and analyzes citations correspondingly!'

Regarding arXiv, 26.9% of the respondents referred to its coverage; 68% of that group indicated that they like the coverage of arXiv, while 32% said that the coverage could be better. Most of the latter wrote that arXiv should also host old preprints, from before it was established. However, we can infer from the overall survey responses that most researchers do not search in arXiv. They typically see the new submissions, and they link to specific papers after searching in other information systems (primarily SPIRES and the Google services). The completeness of arXiv's coverage, therefore, is not as important for the researchers as the completeness of the other information systems, as can be seen by the relatively modest discussion of the arXiv's coverage.

The majority of users who described what they like most and least in Google and Google Scholar—60%—referred to the coverage of these services. According to 79% of that group, the coverage of Google and Google Scholar is extensive and sometimes includes relevant documents that they cannot find in arXiv and SPIRES. Nevertheless, 13.3% feel that the coverage is incomplete when it comes to HEP materials. Other concerns raised by those who referred to the Google services is the lack of focus on HEP (11.7%) and the non-academic nature of the materials offered through these services (7%). Some of the respondents (3%) listed the coverage of the Google services as what they like both the most and the least, as exemplified by the following comment: 'I like the fact that it covers many outlets and sources of scholarly work. I sometimes find sources I might not have found any other way. At the same time, this is also one of its weaknesses: I often have to wade through many useless references and links'.

As is evident from the survey, the breadth of content was of much more concern to those addressing Google and Google Scholar in their replies and the least concern to those addressing arXiv (Figure 16). The findings indicate that the coverage of the Google services is a major draw for the HEP scientists who decide to use these resources, particularly for specific information needs, despite the criticism that the scientists articulate about the quality of the materials that these systems make available.

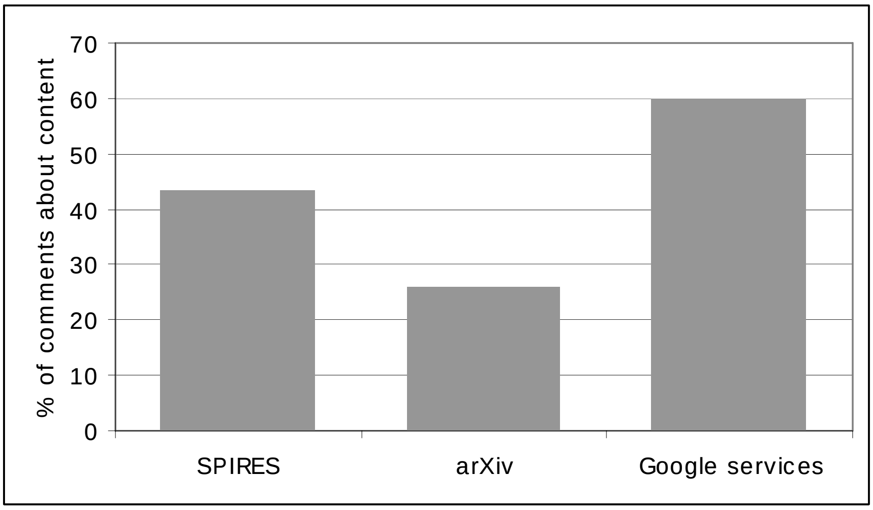


Figure 16: Comments on content coverage of information systems. Percentage of users who commented on the coverage of SPIRES, arXiv, and the Google services, out of the total number of users who commented about these information systems.

Availability of materials: Although access to materials that are published electronically—the majority of the materials that HEP scientists use today—can be immediate, some publications are subject to license and hence are not available to all of the community members. The HEP researchers express considerable awareness of open-access issues surrounding scientific information and often prefer the freely available preprint version of an article to the final, published version, even if the latter is also available to them. An outcome of this awareness is that seminal papers by large collaborations are published exclusively in open-access journals. In the replies to the open-ended survey question about SPIRES, 16% of the researchers referred to the availability of the full text. Whereas 53.5% of this group of researchers like the links to the published materials and 23.7% like the links to the preprints, 42% do not like the fact that SPIRES sometimes links them to publisher sites that are open only to subscribers or require a fee. ‘The uncertainty about access to the full text is probably what I like least about SPIRES’, explains one respondent; another notes that ‘fortunately, it covers the most of HEP. Unfortunately, it does not allow access to the all texts’.

Regarding arXiv, only 12.6% of respondents referred to the availability of the full text—86.7% of these respondents cited the immediate availability of the materials as what they like the most, and the rest listed what they like the least, such as the absence of an article's publication status or the lack of a link to the published version of an article.

Access to full text was addressed by 8% of those who referred to the Google services. The occasional inability to access articles' full text from Google or Google Scholar—as a result of missing or outdated links or links to copies that are unavailable to the user because of the lack of institutional licensing—was listed as a drawback by 72% of these respondents ('Least: often leads to the links without complete texts of the references, even though the full text is available on other sites'; 'the main link is inconsistent for physics. It doesn't always link to arxiv, so it is more of a pain to get full text'). On the other hand, the rest of this group of respondents mentioned easy access to the full text as what they like the most about the Google services.

Quality of materials: Measuring the quality of scholarly information is not a simple task and yet is of great importance to the scientific community. Knowing that an information system applies quality measurements is, of course, crucial to a researcher who relies on an information system to provide quality materials. Furthermore, some systems use such measurements to apply a relevance-ranking algorithm that determines where a specific document appears in a result list.

SPIRES applies quality measurements to the materials it hosts; arXiv performs only minimal screening of the deposited materials; and the Google services index all materials without any quality assessment. The issue of quality was addressed by many of the surveyed researchers (6% of those who described what they like the most and the least in arXiv, 8.5% of those who referred to SPIRES, and 23% of those who referred to Google and Google Scholar). Depending on the context, some prefer strict quality control and others prefer the democratic nature of an open repository. Whereas 91.6% of those who referred to the quality of the materials covered by SPIRES indicated in their responses that this characteristic is what they like the most, only a few researchers (8.4%) said that quality is lacking in SPIRES. Regarding arXiv, 64% of the few (8%) who referred to the quality of its materials do not like the quality. The Google services are perceived as providing data of lesser quality, and 97% of those referring to the issue of quality wrote that they least like the irrelevant or untrustworthy results provided by these services.

As expected, the figures show that the quality of materials was more an issue for users who referred to Google and Google Scholar than for users who referred to the community information systems. While the Google services enable researchers to find quality materials that they cannot locate in SPIRES and arXiv—as well as quality materials that are available in those systems—the researchers need to spend more time to find materials that are of high quality, as can be inferred from comments such as 'I find it contains many irrelevant entries, far more than spires,' 'too much garbage to sift through', and 'too much noise! google scholar returns, it seems like, everything on the web'. One of the researchers summarized the problem as follows: '[What I like] the most: quantity. The least: quality (not filtered by expert referees)'.

Period of coverage: Because arXiv was established only in 1991, it does not hold preprints that precede that year. SPIRES contains metadata that goes back to the 1970s and represents print materials as well as materials that were digitized and are stored in institutional repositories or are available through publishers' Web sites. Using Google and Google Scholar, researchers occasionally stumble upon even earlier materials. The coverage of old materials in arXiv was addressed by 5% of the respondents, all of whom wrote that what they like the least is the lack of such coverage. Of the 11% of respondents who referred to the coverage of old materials in SPIRES, 69% like the coverage and 31% think that this coverage is not sufficient and would like to see more scanned papers online.

Type of materials: Although the major corpus of research materials consists of articles, the availability of additional types is of great value to researchers. These types include conference proceedings, PhD theses, research reports, presentations, lecture notes, data sets, and non-text-based materials. In the HEP survey, however, very few respondents to the open-ended questions referred to the issue of material types at all (only SPIRES was mentioned, and only by 3% of the respondents), and of those respondents, 25% liked the variety of materials offered by SPIRES. The other 75% mentioned the lack of other types of materials (e.g., books, multimedia, and presentations) in SPIRES as what they like least. Survey respondents who referred to Google and Google Scholar mentioned that what they like most is the reference to other types of materials, such as conference talks, annual reports, and Web pages.

3.7.2 Means of Finding the Information

We typically think of searching as the obvious means of locating scholarly materials. In the HEP context, we can examine active, directed searching[12] in all the information systems that we have discussed so far. Depending on the task at hand—searching for a specific document (a 'known' item) or exploring to see what materials are available about a subject or by an author—HEP researchers prefer one information system to another and often change their searching behaviour.

Many respondents referred to searching, although it was not always clear whether they were referring to the search interface, the search options, the search engine, or the metadata that is available to the search engine. For a discussion of these aspects of searching, see 3.7.3.

To obtain relevant information, HEP scientists use two techniques in addition to searching. One technique is to monitor new materials of interest. As expected, the HEP researchers praise services that enable them to see the latest submissions to arXiv: 17% of the respondents mentioned arXiv's daily listings and specified this feature as what they like the most in arXiv. Furthermore, when asked why they use arXiv the most, 21.8% of the respondents wrote that arXiv enables them to keep abreast of developments in their area by offering a means to monitor relevant new submissions. As one of the researchers explained, 'I check every morning the new articles. It is mandatory to be kept up-to-date'. For some researchers, scanning new postings is part of their morning ritual: 'I look at arXiv every day first thing in the morning and no other thing. With a cup of coffee. That has been my routine since 1992'.

The other technique for locating relevant materials relies on a mesh of links between materials. Such links are generated by the information system, and in the context of the HEP community, the best example is the mesh of links provided by SPIRES. In addition to the more obvious links from an article to other

[12] For a discussion of the information-seeking modes suggested by Bates, see 2.2.2.5.

publications that appear in the reference list, links to articles that cite a given article are available, as well as links that follow routes to other items through authors, institutions, and conferences. By following the links, a researcher can track the evolution of an idea and can also easily obtain an overview of the literature on a particular topic. In their responses to the HEP survey, 13.5% of the respondents referred to the citations available in SPIRES, with 78% of this group indicating that the citations are what they like the most (22% said that the citations should be improved). Another 5% referred to the references in SPIRES; 95% of this group named the references as what they like the most, and the remaining 5% found fault with the references. Because citations also serve as a measure of the academic excellence of researchers, respondents expressed concern about the lack of completeness of the citations and the resulting inaccuracy of the citation analysis provided by SPIRES for each paper. Some survey respondents (14%) addressed the issue of citation analysis, 74% of whom said that they like it the most. However, 16% were very critical, as is often the case in the assessment of impact in scholarly communication.

3.7.3 Effort Required to Find and Obtain Information

When commenting about the process of locating and obtaining information, the surveyed researchers mentioned three factors:

- The search interface: how easy the interface is to learn and what it enables a researcher to do
- The search engine: how fast the search process is, how accurate the results are, and how tolerant of user mistakes the search engine is
- The overall effort required for a researcher to actually find and obtain the required items

In addition, many researchers wrote about the search process, though it was not always clear whether their satisfaction or dissatisfaction related to the interface, the options that it offers, or the search engine. It was evident, however, that the user friendliness of the system is important for many researchers.

The search interface of SPIRES was addressed by 28% of the respondents. While some researchers commented that the search interface of SPIRES is very good, others complained that it is old fashioned and complex. Objectively speaking, the default search interface is not intuitive and requires expertise, as noted even by those who like it; however, once researchers master the interface, they consider it easy to use and say that it enables them to specify information needs very accurately.

Interestingly, although the percentage of respondents who addressed the search interfaces of SPIRES and arXiv is rather similar (27.7% and 31.8%, respectively), the respondents were very clear in their comments about arXiv and were divided equally in their opinions about SPIRES (51% and 52%, respectively[13]). Also, if we look at SPIRES alone, respondents who are experienced with SPIRES and yet have less seniority in the field are those who are more inclined to indicate the interface as what they like the most in SPIRES[14] (Figure 17). However, the complexity of the interface may be the reason for the decreasing appeal of the interface to those who belong to the more senior age group.

No. of years of professional activity	Like the SPIRES search interface the most	Like the SPIRES search interface the least
0-2 years	60.0%	40.0%
3-5 years	49.2%	57.6%
6-10 years	55.6%	46.9%
>10 years	47.8%	53.7%

Figure 17: The appeal of the search interface of SPIRES.
The positive and negative comments are broken down by the number of years that researchers are active in their field.

According to the survey responses, the lack of a full-text search and of fuzzy searching in SPIRES are the features that researchers miss the most. Although

[13] Some respondents indicated the interface as what they liked the most and the least; hence, the number of references to this issue is greater than the number of respondents.

[14] Although 60% of junior researchers indicated that they like the SPIRES interface the most, the total number of respondents from this age group who referred to the SPIRES interface is relatively small—20 researchers—as opposed to the number of more experienced respondents (134 respondents of more than ten years in the field, 81 respondents of six to ten years, and 59 respondents of two to five years). Hence the findings related to junior researchers are not unequivocal.

4.2% of the respondents like the search by author names the most, 3.5% least like the need to manually match up variations of author names and the difficulty of identifying the desired author name. These shortcomings make searching inconvenient for finding authors with common names, especially when it comes to Chinese names.

Other search-related factors mentioned by respondents who referred to SPIRES include the search engine (mentioned by 12% of the respondents, 62.5% of whom commented positively), speed (mentioned by 8.5% of the respondents, of whom 64.5% were positive), and aspects of the interface such as the overall design (mentioned by 4.3% of the respondents, with 59.2% commenting positively).

Searching in arXiv is clearly what respondents like least about that system. Many referred to having a workaround—searching through SPIRES, Google, or Google Scholar. For some, such workarounds do not pose any problem, but others would like arXiv to have a better search engine. Although arXiv provides full-text searching, respondents indicated that it should be improved.

The arXiv search interface was also addressed by many respondents—31.5%—most of whom (83.9%) indicated that they like it the least. However, the interconnection between arXiv and SPIRES, designed to provide a complete environment in which these information systems bear complementary roles, does succeed in meeting the expectations of the community. Again, to overcome the limitations of arXiv's search interface, many users search for arXiv materials through other systems, as evidenced in comments such as '[the] search feature in arXiv is not so easy to use; I use SPIRES instead for searching and then click back through to arXiv. They work well together'.

Respondents referred to the kind of searching that they can do with each of these systems. While they use the arXiv search typically when they have an arXiv ID, they find SPIRES convenient when they know the title or the author of the required material. Combining an author or title with a date works very well for them. Regarding keyword searching, 1% of the respondents liked it the most, but 3.6% liked it the least; most of the latter group think that the keywords and classification are not adequate. Whereas very few respondents indicated that they like the capability to conduct exploratory searches in SPIRES the most ('Though arXiv is excellent in many respects, Spires is usually more convenient in searching for references one doesn't already know'), 2.3% of the respondents liked this capability the least. These respondents thought that 'topic search is not as good as Google Scholar'; 'least: hard to explore by subject'; or 'Worst: It is a bit hard to locate papers on a given topic if you have no specific entree'. Google, on the other hand, is perceived by 7.4% respondents as a good tool for exploratory searching.

The Google search function was addressed by 60% of the respondents. The Google services, although viewed as having a search engine that is good (48%), quick (10.8%), and accurate (12.3%), received considerable criticism. Irrelevant or untrustworthy results (22.5%), inaccurate searches (7.7%), the lack of certain search options (7%), and inadequate sorting of results (5.2%) were among the least liked characteristics of the Google services. Apparently, although the search itself is quick, the process of sorting out the results is too long and cumbersome, despite the user friendliness that 12.6% of the respondents like. As one researcher says, '…it gives too many links and one has to really spend time in checking out what is actually useful'.

3.7.4 User Friendliness of an Information System

One of the themes that is repeated in the open-ended responses is the ease of use of each of the information systems discussed. Although ease of use was not always clearly defined, it can be assumed that users refer to the overall experience, which is a combination of screen design, terminology, search and display options, help screens, compliance with accessibility standards, and more.

In their responses, some respondents wrote very general statements such as 'user interface', 'efficient interface', or 'great interface'. Some were more concrete, addressing specific elements such as 'Most: no graphics. I like plain text', or 'most: formal interface'. In addition to such comments about each of the systems, 27.4% of the respondents referred to the ease of use (or lack of it) and 2.3% referred to the interface in a broad sense (as opposed to solely the search interface) in their replies to the question 'Why do you mostly use this system?'

The user interface was brought up by 4% of those who referred to SPIRES, 59% of whom like it the most, and the rest like it the least. Respondents did not refer to the user interface explicitly when commenting about the other information systems. Of the respondents who refer to SPIRES, 21.3% refer to its ease of use—which emerges as an issue of importance to them when the responses are compared to those about the other resources (for which ease of use was addressed by 18.4% of those who refer to arXiv and 11.4% of those who refer to the Google services). Furthermore, the majority of respondents expressed very clearly that they like the ease of use of SPIRES and the Google services and were divided almost equally in their evaluation of the ease of use of arXiv. The user interface of SPIRES was criticized by some, who found it outdated.

When one looks at the number of years that respondents have been using the systems (a way of estimating the ages of the respondents), it is evident that those who have been working with SPIRES for a longer period find it friendlier (Figure 18).

	Like the ease of use of SPIRES the most	Like the ease of use of SPIRES the least
0-2 years	70.0%	30.0%
3-5 years	80.4%	19.6%
6-10 years	87.7%	12.3%
>10 years	88.4%	11.6%

Figure 18: The appeal of the ease of use of SPIRES.
The positive and negative comments are broken down by the number of years that researchers are active in the field.

The responses to the question 'What do you like the most and the least about arXiv?' referred to two additional issues. The first was the site's accessibility, in the sense that the arXiv site is very simple and straightforward in its implementation and does not rely on JavaScript, Flash player, or other more advanced interface components. 3.8% of the respondents liked this simplicity most. The other issue was the organization and display of the information: 3% of the respondents liked it least.

3.7.5 Related Services Offered by an Information System

Searching is only part of the information-seeking process. Obtaining the full text of articles and additional information—such as the number of times an item has been cited—complements the process and is of great relevance to researchers. Other services, such as enabling researchers to download citations in various formats and to navigate to related information—authors, conferences, and more—are important as well. SPIRES offers several such services, which respondents indicated as most liked: citation analysis (10.4%), the ability to follow links (15%), and the ability to download citations in various formats (4.5%).

3.8 Characteristics of the User Community

Although this work focuses on the information-seeking behaviour of the HEP community members in the context of their scientific work, some other aspects of the community members should be considered in the design of a future information environment. Based on a thorough understanding of the goals and objectives of the information seekers, the model describing their information-seeking behaviour is more likely to provide an accurate account of the current practices and a more profound basis for a future system that will be designed in accordance with this model. Most of the insights about the HEP community members originate from the interviews; however some general characteristics can be concluded also from the responses to the survey.

In their published analysis, Gentil-Beccot et al. (2008) refer to several characteristics of the HEP community, as captured by the survey results: nationality, seniority in the field, and the frequency of use of the HEP information systems.

The figures regarding the nationality of the respondents (Figure 19) corroborate the information obtained through interviews: the HEP community is distributed all over the world, with CERN serving as a temporary base for people of many nationalities. Researchers spend time also at the other HEP centres, primarily SLAC and DESY.

Country	Fraction	Country	Fraction
United States	27.4%	Iran	0.9%
Germany	9.5%	Mexico	0.9%
Italy	7.7%	Australia	0.8%
United Kingdom	6.5%	Denmark	0.8%
CERN	4.9%	Sweden	0.8%
France	4.1%	Greece	0.8%
India	3.4%	Portugal	0.8%
Spain	3.0%	Argentina	0.7%
Canada	2.6%	Korea	0.7%
Brazil	2.4%	Austria	0.6%
Russia	2.4%	Poland	0.6%
Switzerland	2.2%	Chile	0.5%
China	2.1%	Finland	0.5%
Japan	1.7%	Taiwan	0.5%
Israel	1.5%	Czech Republic	0.3%
Netherlands	1.2%	Norway	0.2%
Belgium	1.1%	Hungary	0.2%
Turkey	0.9%	Others	4.8%

Table 1: Distribution of answers per country. Users based at CERN were asked to indicate "CERN" and not "Switzerland". 97% of respondents answered this question.

Figure 19: Distribution of survey respondents per country.
Reprinted from Gentil-Beccot et al. (2008), Table 1.

The number of years of experience in the field has been discussed in this chapter in relation to the search behaviour of the survey participants. Assuming that the number of researchers joining the field every year is more or less stable,[15] one might predict that researchers at each level of seniority would contribute a similar number of responses, but this was not the case. By dividing the number of respondents (2,110) by the number of years that researchers typically spend in their field (40), one would expect an even distribution of 50 responses from each experience bracket; however, as evidenced in Figure 20, the participation in the

[15] It is difficult to assess whether the number of HEP researchers is increasing, decreasing, or staying more or less the same. Interviewees and the HEP information specialist team estimate that the size of the community has remained stable in the last two decades or it may have even decreased slightly, with researchers drifting toward the adjacent astrophysics and cosmology fields due to the delay in launching the LHC project in CERN. However, since the beginning of the project in September 2008, the LHC data output attracts researchers back.

survey by the more senior researchers was less than expected. Those who have worked in the field for 3-10 years responded the most to the survey.

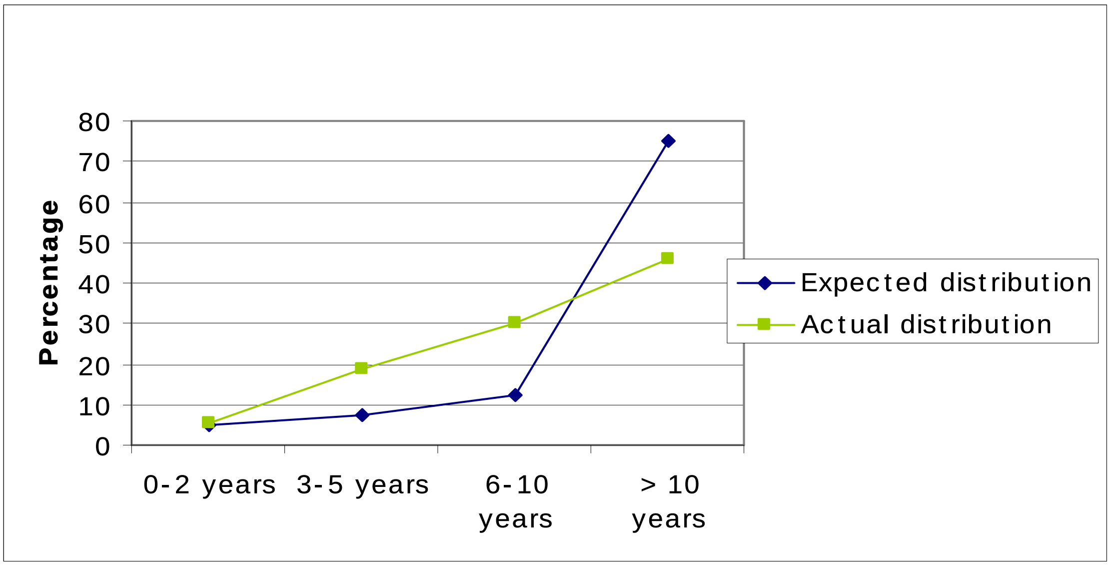


Figure 20: Distribution of survey responses by number of years of experience. This distribution is compared to the distribution expected if there were an even number of responses at each level of experience.

The specific field of the researchers is a major differentiator in regard to the response rate and, even more so, in regard to the reported intensiveness of the use of the HEP information systems. Experimentalists tend to spend less time on information seeking than theorists, as was evidenced in the interviews and by the nature and the number of responses to the HEP survey of each group. Although in personal communications, some of the interviewees estimated that the number of experimentalists is more or less equal to the number of theorists (an estimate that is shared by the team of HEP information specialists), the published analysis of the HEP survey reports that 61.3% of the respondents are theorists, 22% are experimentalists, and the rest work in the field of software (5.5%), instrumentation (3.5%), accelerators (2.7%), engineering (1.3%), or another field (3.5%) (Gentil-Beccot et al. 2008). Some of the respondents, however, associated themselves with more than one field (e.g., experiments, instrumentation, and software).

In terms of the frequency of usage of the HEP information system, the theorists report a much higher rate. Figure 21 shows the percentages of theorists and experimentalists who report using information systems at specific degrees of frequency. The graph represents only those who identified themselves as belonging to one of the groups.

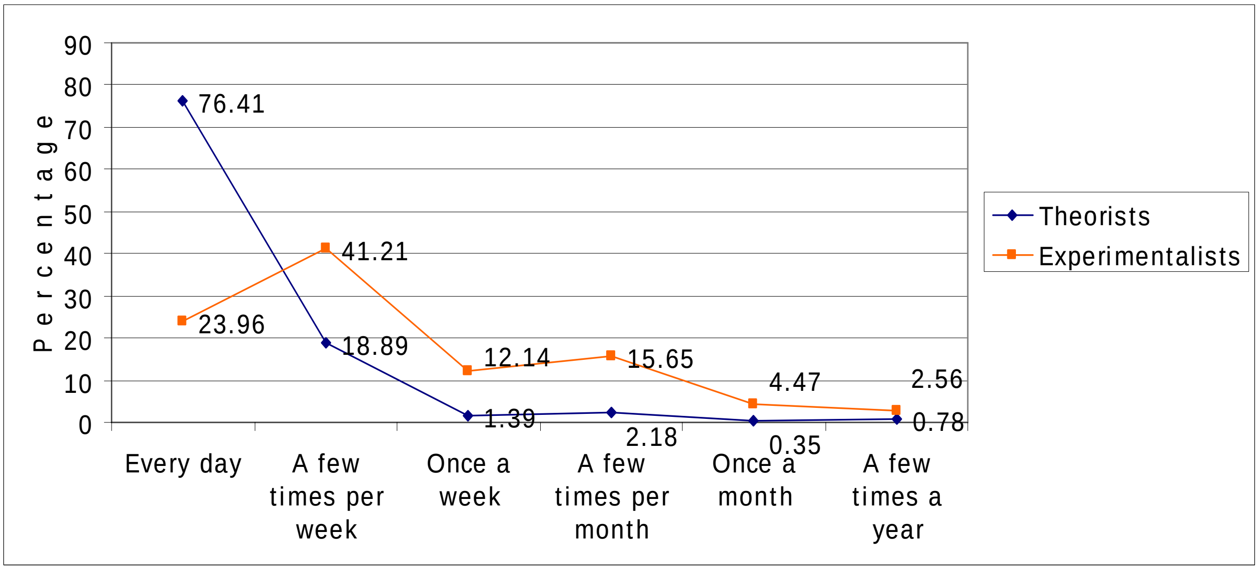


Figure 21: Frequency of HEP information system use by theorists and experimentalists

The interviews with the HEP researchers and personal communications from them corroborate the findings of the survey. Indeed, theorists spend much more of their time in searching for materials. The reason for this difference relates to the kind of work that the scientists are engaged in—theorists are less focused on a narrow area and less dependent on deadlines, equipment, and even collaboration with colleagues. They are typically open to broadening their understanding and changing or adjusting their theories on the basis of new information that becomes available to them. Experimentalists, on the other hand, generally work as part of big collaborative teams and have to focus on fulfilling their own tasks. Hence they spend little time searching for materials that are not directly related to their exact area.

Another characteristic that may be worth noting is that a great majority of HEP researchers is male. The survey does not indicate gender, but the low representation of women among the community members was noted by the interviewees as a traditional phenomenon that has started changing only in recent years.

3.9 Information Systems as Community Tools

The involvement of the HEP community in the development of the information systems and the ongoing maintenance of the systems is significant. Such involvement makes SPIRES and arXiv most valuable to the researchers, on the one hand, and a source of pride, on the other. Respondents expressed their feelings in various ways, starting with their willingness to fill out the survey and write lengthy responses. The tone of their replies and explicit comments reflect their feelings: about 14% of the survey participants added unprompted comments about SPIRES; of them, 96.2% were extremely positive. About 12% of the respondents added unprompted comments about arXiv, all of a positive nature.

3.9.1 Comments Related to SPIRES and arXiv as Community-Developed and Community-Maintained Resources

SPIRES was cited in the context of the community by 19 respondents (1.7%), and arXiv, by 157 respondents (12%). The unprompted comments about the community aspects of SPIRES and arXiv include the following:

- 'SPIRES is doing a wonderful service to the community.'
- 'SPIRES was a historical moment in science. Other fields should use it as a prototype.'
- 'I like the existence itself of the database [SPIRES].'
- 'Many thanks to SPIRES for the valuable service of processing references from all papers and putting it to the database - I understand how difficult it is, but it is extremely valuable to the scientific community.'
- 'Spires is simply the way we search for articles in HEP. Period. No competition. No competition needed.'
- 'We as a community are blessed to have SPIRES.'
- 'MOST: [SPIRES] It's an IRREPLACIBLE source of data.'
- 'good: it [arXiv] is THE place to post in my field, so everything is there and it almost replaces all journals.'
- 'I like most that everyone in my field uses it.'
- 'Likes: that it exists …. It's hard to imagine doing physics without arXiv.'
- 'the most is the tremendous service it does to the community for giving immediate access to new works and for the massive store it is.'

- ‘Most: Not replac[e]able by anything else. It is like a morning newspaper with all hea[d]lines and columns and wealth of information.’
- ‘IT [arXiv] HAS CREATED A LEVEL PLAYING FIELD! This cannot be stressed too highly.’
- ‘most - that it [arXiv] exists. It's certainly the single most valuable online resource for high-energy physics.’
- ‘The arXiv has become the standard repository for scientific preprints in most of the research domains in physics.’
- ‘It would take a few articles to do justice to the historical role of arXiv in the evolution of scientific information mediation. I agree with the articles posted on the arXiv webpage. The speed, freedom and availability of published research results is by far the most important contribution of arXiv. Everything else is of course desirable and deserves our appreciation, but it is of secondary importance.’

3.9.2 Comments about SPIRES and arXiv as Open, Free Tools

As mentioned earlier in this chapter, the HEP community embraces openness and free access to knowledge. In their responses, 10% of those who described what they like about arXiv addressed this topic specifically. Only a few referred to the openness of SPIRES, perhaps because the availability of the metadata is taken for granted while the availability of the full text of the materials is not guaranteed and depends on the publisher if the article is not stored in arXiv. The comments that refer to SPIRES and arXiv as open, free tools include the following:

- ‘What I like the most about arXiv is the essential fact that most papers in HEP are there, freely available for anyone to read.’
- ‘most: free access of course, content generated by *us* [arXiv].’
- ‘The most: being freely available, it spreads knowledge faster and allows faster development of research fields. [arXiv].’
- ‘Science is available for free to everyone. This is democracy of knowledge. [arXiv].’
- ‘Most: That it is a free service that has developed to a standard and is therefore rather complete (‘a paper not in arXiv does not exist’).’
- ‘most: it is 100% free [arXiv].’
- ‘What I like the most is the freedom of exchanging scientific ideas for all interested people, irrespective of their association to an institute, whether

> their institute has enough money to pay for journal publication, whether their idea is in the main stream or something new, maybe not always correct or interesting, but it would be up to readers to judge. And readers learn to accept or reject a work not because a famous person is between authors or it is published in a famous journal. There is no policing by referees.'

Finally, some researchers wrote comments such as 'I like everything about SPIRES'; 'what's there not to like?'; 'excellent'; and 'fantastic'—without any explanation. Although these comments do not give much information to act on, they do serve to show the general attitude of the community toward its information systems.

3.10 Discussion

The survey analysis and the interviews that the author conducted with HEP researchers paint a picture of the HEP community as a distinct scientific community with clear information-seeking behaviour characteristics, some of which cohere with findings observed in other scientific communities whereas other characteristics are unique.

3.10.1 Creation of, Cultivation of, and Reliance on Collaborative Information Systems

It is evident that in addition to being heavy users of the community information systems, the HEP researchers also make an effort to ensure the systems' superiority for the HEP field and take pride in their success. In fact, the community information systems were initiated by community members and not by external providers, and the need for better communication channels among community members was actually the trigger for the creation of the World Wide Web by Tim Berners-Lee, a community member. The innovation in the HEP communication channels is well respected by information specialists and, as one of the librarians at the Weizmann Institute explained in a personal communication, watching the advances in HEP scholarly communication enables one to predict the trends that will arise in some other scientific disciplines.

3.10.2 Loyalty to and Trust in the Community's Information Providers

Most HEP researchers demonstrate a high degree of loyalty to and trust in the community information systems and the teams that maintains these systems, technically and content-wise. Unlike their attitude toward the remote, anonymous teams that develop and maintain Google and Google Scholar, the HEP researchers not only feel close to the teams that manage SPIRES and arXiv but also contribute time and effort in providing these teams with relevant and accurate information. The researchers show a great degree of empathy for the teams' members in their efforts to fulfil all their tasks. Such intimacy among information providers and information consumers—all of whom are community members—is fundamental to the success of the community information systems.

3.10.3 Open-Access Philosophy

HEP community members—both researchers and information providers—are very keen on providing open access to information. Their information systems are, of course, open to all. Furthermore, the community members prefer publishing in open-access journals and using the arXiv copies of articles rather than the publisher's version of articles whenever possible. In the last few years, HEP information specialists have been deeply involved in initiatives aiming to change the business model of publishers so that all papers are made available to all (http://scoap3.org/index.html).

3.10.4 Preference for a High-Quality, Focused Information Corpus

Although many, if not all, researchers use Google and Google Scholar, it is evident that in most information-searching interactions, HEP researchers prefer using their community systems. There are several objective reasons for this preference: primarily, the high scientific quality of the content and, in the case of SPIRES, the HEP-exclusivity of the content; the search options; and the related services (mainly lists of daily submissions of new articles, citation analysis, and bibliographic tools). The researchers' reliance on community resources has developed over time and has proved to be more beneficial for their purposes, in most cases, than reliance on alternative channels to information. In addition, using SPIRES and arXiv complies with the community's traditional information behaviour and its attitude toward open access.

3.10.5 Familiarity with Materials and Researchers

Because the HEP community is relatively small, senior researchers are familiar with the worldwide research community. Therefore, they can use their knowledge of the authors, institutions, or projects to assess content. Such knowledge may compensate for the lack of peer review of papers deposited in arXiv and the uncertain quality of many materials available on the Web—lecture notes, conference proceeding, reports, and the like. Junior researchers typically work with senior researchers, in small groups, and hence can benefit from the knowledge of the senior researchers when it comes to assessing materials.

3.10.6 Use of Several Information Systems to Satisfy All Information Needs

The HEP community members have developed information-seeking behaviour that takes advantage of several information systems—in particular, SPIRES, arXiv, and the Google services—which complement each other and together satisfy all the researchers' information needs. The systems differ in coverage, type of content, searching options and capabilities, access to full text, and related services; however, the systems are integrated through hypertext links to provide one environment. Information seekers learn to use the right tool for the specific task and feel comfortable with the interplay between these systems.

3.10.7 Perceived Need for Better Interdisciplinary Coverage by the Existing Information Systems

For researchers who focus solely on high-energy physics, the coverage of SPIRES is optimal: it includes all the documents they need, and it does not include information that is not relevant to them. However, for researchers whose areas are tangential to pure HEP or deal with interdisciplinary research, the coverage of SPIRES is too narrow. In a manner typical to the HEP community, researchers feel engaged and suggest improvements—in this case, the inclusion of other materials.

3.10.8 Need to Keep Abreast of New Materials

HEP researchers, like researchers in other fields, are always on the lookout for information. They sometimes conduct directed searches; they look at new materials daily; and they are always aware of relevant information, even that which might become handy in the future.

3.10.9 Use of Several Techniques Besides Searching to Satisfy Information Needs

Besides active, directed searching, researchers rely on a web of links that is provided through SPIRES. After finding one article of interest, the researcher navigates to other articles. Citation information provides one important path to related materials (and also serves for the evaluation of materials); however, other paths are often used.

3.10.10 Online Reading Used Almost Exclusively

Most HEP literature is available online, and the use of print materials is negligible. Even older materials are often scanned and made available online. Hence, the current information-seeking practices are adapted to an online environment.

3.10.11 Differing Information-Seeking Patterns According to Experience in the Field

The survey results suggest that researchers who are newer to the field tend to rely on Web search engines more heavily than other researchers. It is not clear, though, how the behaviour of these newcomers evolves over time: as they spend more time in the community, do they become more aware and more appreciative of the community information systems? Or do these researchers stick with their pre-HEP information-seeking habits, which will probably change the nature of information seeking by the community as time goes by?

3.11 Conclusions

The overall satisfaction that HEP scientists express about their community-based information systems suggests that the community effort that is invested in these systems renders them not only worthy from a professional point of view but also a source of pride to the community members. The aggregation of individual contributions over a long period of time improves the quality of both the information stored in the databases and the services that the information systems provide, making such services more focused on the real needs of the HEP researchers. However, the tendency of researchers who are newer to the field to rely on Google and Google Scholar implies that the research behaviour patterns of the HEP community may change in the future and may affect community members' contribution to the ongoing maintenance of their information systems. Improving the user experience of the dedicated HEP information systems while continuing to support researchers who want to conduct sophisticated searches might help halt the younger generation’s tendency to use other information systems. In addition, as scientific research becomes more interdisciplinary and as new types of materials gain in relevance, the HEP information systems will have to develop accordingly.

Providing better support for researchers in their efforts to satisfy their information needs would require, as described by Wilson (1981), a shift in the focus of information research ‘from an examination of the information sources and systems used by the information-seeker to an exploration of the role of information in the user's everyday life in his work organization or social setting’ (Wilson 1981). Wilson elaborates:

> Qualitative research seems particularly appropriate to the study of the needs underlying information-seeking behaviour because:
>
> - our concern is with uncovering the facts of the everyday life of the people being investigated;
> - by uncovering those facts we aim to understand the needs that exist which press the individual towards information-seeking behaviour;
> - by better understanding of those needs we are able better to understand what meaning information has in the everyday life of people; and
> - by all of the foregoing we should have a better understanding of the user and be able to design more effective information systems.
>
> ...Before a generally applicable theory of information-seeking behaviour can be evolved, the context of the research must be narrowed so that crucial determining factors can be identified and analysed. There can be little use, for example, in a national survey of the 'information needs' of any group (chemists, botanists, economists etc.) if members of these groups are undertaking widely differing kinds of tasks in totally different organizations with varying levels of information provision. If we wish to uncover the determining factors of behaviour we must do so by first undertaking in-depth studies of well-defined categories of persons, developing explanatory concepts and then testing these concepts in related but different settings. (Wilson 1981)

The general characteristics—and, in particular, the information-seeking behaviour of the HEP community members—inferred from the interviews and the analysis of the survey results described in this chapter serve as the basis for the creation of HEP personas, described in Chapter 4; the information-seeking model presented in Chapter 5; and the user-interface design derived from the model, described in Chapter 6. By understanding the nature of a scientific community and its requirements, one is more likely to create an information-seeking model and a user interface based on that model that will successfully address the community's research practices and support it in satisfying its information needs.

Chapter 4 Personas

4.1 What Are Personas, and Why Use Them?

It is obvious that creating a model that represents user behaviour requires a thorough understanding of the users in respect to their information-seeking behaviour: their goals, their expectations, their needs, and their form of action. Both qualitative research and quantitative research enable us to acquire such understanding, whereas a combination of the two research methods is likely to yield the best results. Because there is no ‘typical’ user—rather, every user is unique—this work suggests the use of *personas* to illustrate representatives of the high-energy physics (HEP) community and to serve as the basis for the design of the model—and later, of the derived user interface.

Another way of gaining an understanding of the goals, expectations, needs, and habits of a target population is by conducting psychometric tests. Hammond (2006) describes these tests as follows:

> Psychometrics means literally ‘measurement of the mind’, and psychometric tests are designed to measure the intrinsic mental characteristics of a person. …
>
> Owing to the lack of direct access to the mental characteristics under scrutiny, the discipline of psychometrics has developed a detailed set of procedures and models for statistical estimation. Essentially, these procedures rely on the presence of a large number of indicators allowing us to ‘focus in’ on or triangulate the characteristic being measured. In most psychometric tests these indicators may be viewed as the individual items or questions of which they are composed. (Hammond 2006, 184)

Using psychometric tests—typically consisting of well-formulated, structured questionnaires—to examine a representative sample of community members yields a series of indicators. However, as Hammond (2006) notes, ‘most psychometric tests in use today are normative or norm referenced, which means that data exist which tells us what range of scores is expected from the population under consideration. This requires that the means and standard deviation of a large representative sample are available to the tester so that she or he can interpret the meaning of an individual's score. These descriptive statistics are termed the norms' (Hammond 2006, 188). Hammond goes on to explain that psychometric tests can also

> use criteria other than test norms for interpreting test scores, as long as they are clearly specified in advance. This strategy is employed by a class of tests known as criterion referenced…. In this case an external performance criterion becomes the standard against which a respondent is judged. …The main point in criterion-referenced tests is that the respondent either reaches a prespecified criterion or does not. Obviously, this means that the criteria have to be established very accurately and precisely justified before the test is made available for use.' (Hammond 1995, 189)

Although analyses of the responses to psychometric tests can be useful in clustering community members on the basis of specific personality characteristics (Hylegårde 2009; Bawden and Robinson, forthcoming), psychometric tests were not the tool of choice for this study for several reasons:

- The HEP survey took place before this study began and was not designed in a manner that would support psychometric analyses.
- A questionnaire for use in psychometric analysis requires that the designer have a set of characteristics in mind and build the questionnaire in light of these characteristics. In this study, the special characteristics of the HEP community members were revealed as a result of the interviews and the HEP survey analysis; once the author had gained an understanding of the community traits, there was no need to start over with the design and execution of a psychometric test.
- The creation of personas, a method with which the author is familiar in practical contexts, provides not only a means to describe representative individuals but also enables one to use personas as tools. In this study, the personas are used for the development of the information-seeking model and the user interface derived from this model, as explained later in this chapter and described in Chapter 5 and Chapter 6.

Defining personas is an established practice in user-interface design. Blomkvist (2002) describes personas as follows:

> A *persona* is a model of a user that focuses on the individual's goals when using an artefact. The model has a specific purpose as a tool for software and product design. The persona model resembles classical user profiles, but with some important distinctions. It is an archetypical representation of real or potential users. It's not a description of a real, single user or an average user. The persona represents patterns of users' behaviour, goals and motives, compiled in a fictional description of a single individual. It also contains made-up personal details, in order to make the persona more 'tangible and alive' for the development team. (Blomkvist 2002, 1)

The concept of using personas was first introduced by Alan Cooper as part of his goal-directed design methodology (Cooper 1999). According to Cooper (1999), 'our most effective tool is profoundly simple: *develop a precise description of our user and what he wishes to accomplish*. The sophistication comes from how we determine and use that precise description' (Cooper 1999, 123). By defining personas, the designer of a system can better comprehend not just the way in which users operate within the limitations of their current or future environment but also how their form of action relates to their goals and perceptions. As Blomkvist (2002) explains, 'the main contribution of using personas in interaction design is that the process will be focused on the user's goals instead of tasks. The design process also regards personal objectives as important, which often is neglected in design methods and in theoretical models of users' (Blomkvist 2002, 7).

Referring to Cooper's concept of personas, Perfetti (2002) gives the following explanation:

> Rather than designing for all people or for averages, the Cooper approach suggests that designers focus on the unique goals of a specific person to develop a product that satisfies the needs of many users. A persona is a profile of a typical user; it is a description of an archetypal user synthesized from a series of interviews with real people and includes a name, a social history, and a set of goals that drive the design of the product or web site.

> By closely adhering to the goals of a specific persona, the designers satisfy the needs of the many users who have goals similar to those of the persona. The process is even more effective when designers design for several personas simultaneously, as they can satisfy an even larger number of users. Although designing for one to satisfy many may initially seem counter-intuitive, teams we've talked to who have employed it tell us it's a very effective technique.
> (Perfetti 2002)

Although personas are typically used in practical contexts of user-interface design, this work suggests that personas can be helpful also when defining a model of information seeking. Rather than referring to users in an abstract form, references to personas enable the design of the model to better capture the characteristics of the types of users that are actually taking part in the information-seeking process. Furthermore, personas help designers focus on the types of users that play a major role in the environment under study instead of paying attention to all potential users, some of whom may be too specific and may distract a designer from the main types of behaviour. Because ‘personas can be thought of as hypothetical users—fictional people who represent classes of users’ (Randolph 2004, 108), the challenge of a designer is to depict the most relevant representative users.

Dantin (2005) notes that ‘the approach of identifying personas and performing their tasks in evaluating UIs of…software systems was most definitely a process that helped introduce clarity and a form of accountable reasoning into the UI evaluation process’ (Dantin 2005, 7).

4.2 Creating Personas

In accordance with the user-experience design guidelines provided by Mulder (2006), the author of this study defined and described various personas on the basis of qualitative research—individual, recurring interviews with HEP researchers of various types—and quantitative research—the analysis of the HEP survey, including the open-ended questions.

Because they represent the potential users of a system, personas are defined on the basis of the needs and goals of actual users—both their personal goals and those that relate to their use of the system (Randolph 2004). As it is not likely that all users have the same needs and goals, multiple personas are necessarily created. However, the designer of the system must evaluate the degree to which the system is required to support the needs and goals of each persona and define at least one primary persona—'someone who must be satisfied with the system for it to be considered a success and who cannot be satisfied with an interaction designed for another persona' (Randolph 2004, 109).

Mulder (2006) suggests several approaches for creating personas, all involving qualitative research and 'segmentation' of the users. Segmentation, explains Mulder (2006), is 'the art of taking many data points and creating groupings that can be described based on commonalities among each group's members. For personas, the goal is to find patterns that enable you to group similar people together into types of users. This segmentation is typically based on their goals, attitudes, and/or behaviors…. It's less about science and more about sitting in a room reviewing your notes and listening to your gut' (Mulder 2006, 41).

The approach taken in this thesis project involved the following steps:

1. Conducting qualitative research, typically user interviews
2. Segmenting the users on the basis of the qualitative research
3. Testing the segmentation through quantitative research
4. Creating a persona for each segment

Mulder (2006) offers this recommendation for testing the segmentation: 'Through a survey or other form of quantitative research, test your segmentation model using a larger sample size to be more certain it accurately reflects reality. The goals are to confirm that these segments are in fact different, and to have evidence to back up your personas in front of stakeholders' (Mulder 2006, 45).

Defining personas is not a strict technical or scientific process; it requires that the designer employ creativity and intuition. Mulder (2006) adds that 'each type of user evolves into a persona as you add more detail to their goals, behaviors, and attitudes. Each one becomes realistic when you have supplied a name, a photo, demographic information, scenarios, and more' (Mulder 2006, 41). The quantitative research adds more certainty to the decisions taken at this stage, because they can be supported by statistical evidence. ‘With this approach’, Mulder (2006) explains, ‘you get a little more science and a little less art for your personas. The segmentation is still based on qualitative research, but you use quantitative research to obtain evidence to back up your decisions’ (Mulder 2006, 47).

The description of a persona includes a fictional name and life story, a picture, and a ‘tag line’—a phrase, supposedly said by the persona, that is likely to represent the character of the persona as related to the area that is being described or addressed in this work.[16]

The author of this work segmented the HEP community members on the basis of findings from the HEP survey analysis (detailed in Chapter 3), interviews with the HEP researchers (see Appendix A), and conversations with the HEP information specialists. The findings led to the following guiding principles for the segmentation:

- The HEP community is divided more or less equally between theorists and experimentalists.
- Theorists are more focused on searching than experimentalists; also, the former are more inclined to conduct broad searches, whereas the latter are typically focused on obtaining very specific materials.
- Information needs vary from one researcher to the next, depending on the person’s seniority in the field. However, the great majority of researchers closely follow the advances in their field.
- There are very few female researchers among the community members, but the number has been increasing slowly in recent years.
- Research is carried out worldwide, and researchers collaborate regardless of nationality or mother tongue. Furthermore, researchers tend to spend periods of time at HEP centres (months or even years) and carry out some of their studies at institutions that are not necessarily in their country of origin.

[16] Five pictures were obtained from Getty Images (US), Inc., and are licensed to be used in this work. The sixth picture (that of Ed) was taken by the author and was approved by the person who was photographed.

- Researchers respect the ongoing effort that is being invested by the HEP information specialists in maintaining the HEP information systems. The information specialists, for their part, respect the users and are highly focused on meeting the researchers' expectations.
- Whereas researchers who have spent more than ten years in the field have already built their information-seeking practices, newer entrants to the field vary more in their information-seeking behaviour, are more inclined to rely on Web search engines and other non-scholarly tools, and are more open to new technologies.
- Researchers are proud of their research field and have a strong sense of affiliation to the HEP community. They find their work interesting and challenging.

These guiding principles and the desire to represent the main characteristics of the community members led the author to divide the HEP researchers into six groups and develop a persona who stands for each group. The characteristics of the groups are not mutually exclusive. Hence, for example, half of the personas represent theorists and the other half represent experimentalists; three personas represent young researchers, two represent researchers in the prime of the career, and one represents a researcher who has been in the field for more than four decades; and five out of the six personas are male figures. It is obvious that not all female researchers are PhD candidates or theorists and not all post doctoral students move to a foreign country for a period of time; however, the concretization of the characteristics into personas paves the way for a better understanding of human behaviour, in general, and information-seeking behaviour, in particular, as in the context of this study.

The six personas were presented to the HEP information specialists in February 2009. The order in which the personas are listed in this chapter is based on the recommendation of the HEP specialists.

4.3 HEP Personas

4.3.1 Persona A: Ed

Ed is the head of a physics department in an academic institution located in a small city. He is a HEP theorist.

Ed is 54 years old, married, and has three children, age 25 (a postgraduate student in computer science), 22 (an undergraduate majoring in psychology), and 16 (a high-school student). Ed's wife is a primary-school teacher. The family lives in a townhouse, walking distance from the university, and Ed typically walks or rides his bike to work.

"Accuracy and clarity are what count. Think methodically, be precise, and ignore ornamentation that attempts to hide vagueness."

Ed is highly respected by his colleagues, and his career has been developing smoothly. He has been working at the institution for the last 25 years, except for two sabbatical years at SLAC and three months at CERN. He has also been a visiting professor at other institutions for shorter periods. He spends about half of his time doing research, and the rest is dedicated to teaching two courses, supervising three PhD candidates, and dealing with administrative tasks related to his role as department head. His primary area of research is Large Hadron Collider physics, but he is also interested in the phenomenology of supersymmetric theories and neutrino masses. He is often invited to speak at conferences and workshops and has won six prizes for his scientific achievements.

Ed's academic excellence is the result of methodical, intense work and he loves his area of expertise. He spends 10-12 hours a day working. It is important for him to be up to date on research in his field, and he invests time in acquiring expertise in related fields, too. He does not consider himself very computer-savvy—he uses the computer as means to do his work. He often reads help files and struggles to find the best way to locate relevant materials or carry out his research tasks, and he listens carefully to colleagues when they show him how to use software tools and information resources because he feels that he is not an expert in this area. Once he has adopted a procedure, he tends to stick with it and does not look into new techniques unless he reaches a dead end. Ed does not have a laptop, but he has a computer at home that is connected to the institution's network. He is not keen on computerized systems that use busy screens and a lot of animation; he prefers clean, simple screens. He has a mobile phone but refrains from using it, and he is not a fan of gadgets like the iPod. Furthermore, he is a very private person and opts not to use services on the Web if these require authentication.

In his spare time, Ed enjoys hiking.

4.3.2 Persona B: Laura

Laura is a PhD student at a multidisciplinary research institute. She is 27 years old, one of the very few female students (and female academics in general) in her department. She lives with her 30-year-old boyfriend, and they plan to get married and start a family once she submits her thesis. They live in a small apartment not far from the institute. Her boyfriend just finished his veterinary studies and started working at a clinic.

"For me, dealing with high-energy physics also means working with the most interesting people on earth!"

Laura acquired her bachelor's and master's degrees elsewhere and came to this institution to do her PhD. As a matter of fact, her focus on the theoretical aspects of astroparticle physics and cosmology can be attributed to her PhD supervisor, whom she met at a conference where he introduced her to this field. After hearing his lecture, she decided that she would like to continue under his tutorship and in the same general area of research.

Laura was indeed admitted to the program that she aspired to and was granted a scholarship that enables her to devote herself to her studies full time. She spends many hours a day in her room at the institute. She has been to Harvard twice, for a month each time, and enjoyed her visits very much, both professionally and personally.

Laura spends at least two hours a day looking for scientific materials. She is quick to master new tools and takes advantage of every source of information, including Google, which can help her find the materials she needs. Because she is not yet familiar with all the players in the field, she often looks for advice from her supervisor or other faculty members so she can focus on relevant materials and assess the quality of materials she finds by herself. Knowing the right scientific terminology was an issue at the beginning, but she is improving gradually. She communicates with people who deal with the same topics in other countries, and as a result has some personal friends in the United States, several European countries, and Korea. She is a member of four social networks—My Space, LinkedIn, Plaxo, and Facebook—but her favourite is Facebook, which she uses to communicate with friends and with her two teenage brothers.

Right now, Laura is planning a trip to India, to attend a conference in Udaipur in the autumn. She is spending quite some time looking for information related to this trip. She attributes great importance to recommendations and reviews about hotels, sights, and festivals and plans to share her impressions with others on the Web once she gets back.

4.3.3 Persona C: Kevin

Kevin is retired. He is 72 years old, and two years ago he decided, reluctantly, to leave the academic institution in which he worked for over 40 years. He arrived as a young assistant professor and served in several roles during his career, including eight years as head of the department. He also spent periods of time at HEP centres—CERN, SLAC, and DESY.

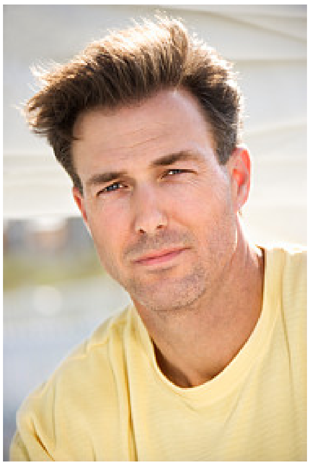
> "Looking back, I'm proud to have taken part in some of the most exciting scientific projects of my time."

Kevin lives with his wife in a small suburban house. They have a dog. Their routine includes a long evening walk, bridge games twice a week, and classical music concerts twice a month or so. They have a son—a scientist as well, a biomedical researcher—who lives in another country and comes every summer with his family. Their daughter lives close to them, with her partner and three small children. Kevin's wife is highly involved in taking care of these three grandchildren, and the couple often babysit for the children in the evening.

During his research career, Kevin saw many innovations in various areas. He managed to stay abreast of most new technologies, although, he admits, some recent changes in many non-scientific systems are too much for him. He cannot understand the reason people prefer communicating via social networks rather than meeting face-to-face or speaking on the phone. For him, computerized communication channels are made for delivering important information from one place to another. Also, he is not interested in the opinions of everybody; and in the context of his work, he does not see much advantage in opening up scientific evaluation to people who are not among the best in their areas. Furthermore, he suspects the motives of those anonymous writers of reviews and comments—are they 'hired' by the author? Are they competing with the author? He also finds the new systems more challenging from the design point of view: his eyesight is not as good as it used to be, and small fonts and busy screens make him uncomfortable.

Although he is retired, Kevin maintains close relationships with tens of colleagues, and his ex-students still consult him. He tries to follow recent posts and spends an hour every morning reading materials that relate to the many topics he dealt with as a theorist in particle cosmology, phenomenology of high-energy physics, and high-energy astrophysics. He knows that he can keep abreast only on a superficial level, so he focuses on abstracts and rarely reads an article from beginning to end. However, his great degree of accumulated knowledge and his ability to identify interesting trends make him a valuable source for consultation. Sometimes, when asked about specific issues, he looks for information that is not necessarily new. In the past, searching used to be as easy as breathing, but it has become more challenging, and he keeps going back to the help pages in order to rephrase his queries. He hates the fact that he cannot remember all the abbreviations of journal titles or the exact way in which his favourite information system requires one to write the names of institutions. Sometimes he wishes there was an easier way…

4.3.4 Persona D: Guy



Guy is a faculty member at a big academic institution and is a HEP experimentalist.

Guy is 48 years old, divorced, and lives alone. His two daughters (17 and 15, high-school students) live with their mother, and he sees them occasionally, when he has time and they are not too busy, either. He lives in the centre of a big city, about a 45-minute drive from his office. His life is rather hectic: besides his work, he is an activist in the Green party. Several evenings a week, he goes to meetings and demonstrations and is involved in other initiatives related to environmental issues.

"My goal is to leverage my knowledge of physics for solving environmental problems. The beauty of creation—as particle physics is slowly revealing—is about to be destroyed by human beings. We have to stop this process!"

Guy started out by studying medicine, but after three years at the university, he decided that he prefers physics and started all over again. His career, then, began a bit late, but he managed to move forward and is now a professor and a member of the ATLAS statistics forum at CERN. He often travels to CERN and is highly involved in the Large Hadron Collider project.

Guy teaches one course—he is considered a good teacher—and spends the rest of his time on research; he tries to avoid administrative tasks as much as possible, and his desk is piled with papers. As an experimentalist, he spends most of the time 'creating' science, as he puts it, and does not invest much in reading scientific literature. He is rather impatient and has no time to spend learning how to use software tools that are not absolutely mandatory for his work. Guy highly prefers simple, intuitive systems and is an avid user of Google and Amazon. He appreciates services that are tailored to his specific needs, such as recommendations that take his profile into account. He has an account on Facebook (with 71 'friends') but does not invest much time in communicating through this site. Guy has an iPhone and cannot imagine his life without it.

Besides speaking at professional conferences, Guy is a sought-after interviewee for TV shows and is often asked to present physics at popular-science events. At home, he has an interesting modern art collection and a big music collection, mostly jazz.

4.3.5 Persona E: Hiro

Hiro is a postdoctoral researcher at a prestigious institution. He was lucky to be accepted to work with a well-known team whose members are pleasant and easy to get along with, although he had to leave his country and move to another part of the world where the language spoken is completely foreign to him. He is 26 years old and single and left behind a large group of friends as well as his close family (parents and a younger brother). The first few months after the relocation were not easy, but now, about half a year later, he is more or less settled in.

"I'm privileged to deal with elementary particles, the cornerstones of the universe, and would never dream of giving up the experience, despite the competitive atmosphere and the uncertain personal future."

As an experimentalist, Hiro's work requires him to be familiar with advanced computer technologies and statistical tools and to have a thorough understanding of physical processes. Because Hiro is thinking of the future—he will need to start looking for a position at the end of his first postdoctoral year—he is very eager to come up with some interesting results that will enable him to apply for a job at a first-rate institution. He will also need his tutor to support his application for such a position. Meanwhile, he is working on several possible research directions and spends many hours a day at the lab.

Hiro rarely looks for new scholarly information. Most of the time, if pointed to something specific, he locates the relevant document as fast as he can and typically ends the search process there. He does not have time to read much, trying to focus on a breakthrough in his research, but sometimes he has to complement his knowledge with background information. Often such information relates to computations; he discovered that the best sources are theses that are scattered all over the Web. Finding the right one is challenging, but once he manages to do so, he gets the information he needs.

Hiro has a Facebook account and uses Skype and JAJAH, with a webcam, to communicate with his friends and family. Of course, he has very limited time to do so, and his location in another time zone makes the direct communication challenging, directing him back to e-mail. Sometimes he has the feeling that these communication tools, which seem as if they are bringing him closer to his 'natural' environment, make him miss home even more.

Right now Hiro is under much stress. He hopes that once he settles down in an institution he will be able to enjoy his work much more.

4.3.6 Persona F: George

George is an experimentalist, a researcher working at CERN. He is 32 years old and the proud father of a baby girl, six months old. He lives in a village on the French side of the border, about 35 km from CERN, because it is less expensive than living in Geneva, and he commutes by train and bus. He leaves home at 7:15 every morning and gets back around 7:30 in the evening, just in time to see his daughter before she goes to bed.

> "It's so much fun! Lots of technical challenges for me, but this is what I like doing."

George spent a year at CERN as a PhD student and then decided that he would like to work there in the future. Three years later, he managed to obtain a position at CERN, as a result of his intensive work as a postdoctoral student. Meanwhile, his wife completed her degree in psychology, but since their move to France, she has not been able to find a job because of the language barrier. Now she is a full-time mother and is making gradual progress in acquiring French.

George is a member of a team of researchers who are focused on designing and constructing a new apparatus as part of a large collaborative project with outside groups. George needs to invest considerable time in reading research articles, hardware and software specifications, technical reports, manuals, and more. Often he tries to find information that is new to the team, so he needs to conduct exploratory searches in a variety of resources. Sometimes he feels like a detective, following clues from one document to another, spending much time sorting out dozens of documents to reach the valuable relevant ones.

George is technologically savvy, of course, and can easily manage with new instruments and gadgets, typically using trial-and-error methods. On a personal level, he enjoys such new ‘toys’ and always has the latest model of every gadget. He already envisions the fun he will have playing with his daughter once she grows up a little—assembling complex structures and inventing new machines together.

On his way to work, George likes to read. He is fond of science fiction—has always liked it—and attributes his fascination with particle physics to the books that he read as a child. On his way back home, he is typically too tired to read and prefers listening to rock music on his iPod. In the evenings, he likes watching TV, mainly football and basketball games, and on the weekends, he sometimes plays basketball with the locals in his village. On the weekends, he also goes shopping with his wife, and they typically have lunch in a nice restaurant. Since his daughter’s birth, they have hardly gone out in the evening, but he likes doing that. As far as he is concerned, his life is now on track and he plans to remain at CERN for the foreseeable future.

4.4 Evaluation of the Personas

To ascertain that indeed the personas represent groups of HEP researchers and that these groups cover the range of HEP community members, the descriptions of the personas were first sent to the team of HEP information specialists, who provided their input.

The reply from the HEP information specialists’ group was written by Travis C. Brooks, the manager of information systems and SPIRES, as follows:

> First off the Personas are great! They really evoke HEP users in a way that I think the technical wonks are not capable of thinking when developing. The utility of these are clear to me already in the sense of remembering that one’s users are not oneself. However, we certainly need to take these forward and investigate how these users will feel about INSPIRE and what they will need from it. What

services will INSPIRE need to capture all of these users and serve them well?

I think my first step is to tell you which Personas I feel are most important. I agree that I would not part with any of them, but I also imagine the scope must be limited. So I rank them in order of importance, 1 being the most useful and evocative.

1:Ed
2:Laura
3:Kevin
4:Guy
5:Hiro
6:George

I think this roughly satisfies the need to have one from all the various demographics. Laura is pretty similar to George in many ways, and I'd prefer her to him. Hiro is rather like Laura + Guy I think.

It would be a shame to lose any of them, but George can probably go. If you must trim another, Hiro is ok to lose. I would really worry about losing more than that. (T. C. Brooks, pers. comm.)

This ranking suggests more emphasis on theorists—an emphasis that matches the profile of the theorists portrayed by the survey results and the interviews, as discussed in Chapter 3. However, because guidelines for using personas for user-interface design recommend three to four personas, this thesis refers to two groups of personas; one group consists of the first four personas (three of whom are theorists), and the other group, of all six (three theorists and three experimentalists).

Two other individuals were asked to read the description of the personas and provide their feedback. Prof. Ady Stern, a condensed-matter physicist from the Weizmann Institute of Science, writes: ‘I read the descriptions of the personas that you developed for your study. I found this way of approaching the subject to be very creative. I am not part of the HEP community myself, but it is a community I know very well as a condensed matter physicist. I found the personas you developed to be a faithful representation of the HEP community I know, and to illuminate precisely the personality aspects that are relevant for the way they search for information. It is a very insightful piece of work'. (A. Stern, pers. comm.) Prof. Yosef Nir, dean of physics at the Weizmann Institute of Science and a member of the HEP community, responded as follows in an e-mail message: 'I read with much interest the description of the six HEP personas. It is excellent: On one hand, I could (almost) identify the real persons behind the personas, and, on the other hand, the six personas give a faithful and comprehensive description of the HEP world'. (Y. Nir, pers. comm.)

4.5 Conclusions

The six personas, developed on the basis of qualitative and quantitative research and corroborated by the HEP information specialists and two other researchers—one a community member and the other a member of a related scientific community—contribute to the design and evaluation of the information-seeking model, described in Chapter 5 and the user interface design described in Chapter 6.

The HEP information specialists ranked the personas on the basis of each one's relevance to the design of the future information system as extrapolated from the personas' current involvement as 'users' of the existing HEP information systems. Later in this work (Chapter 6), a subgroup of the four most relevant personas is addressed separately from the whole group so that the conclusions derived from assessing the behaviour of the more relevant subgroup can be compared with the conclusions derived from assessing the whole group.

Chapter 5 Information-Seeking Model of HEP Researchers

On the basis of the survey of existing information-seeking models, described in Chapter 2 of this thesis, and the empirical research described in Chapter 3, the author developed a new information-seeking model, presented in this chapter. The model provides a comprehensive abstract representation of scholarly information seeking, including an active search process, of a specific scientific community—high-energy physics (HEP) researchers. The special characteristics of both the sought materials and the researchers who are seeking them have been taken into account in designing this model.

5.1 The Scholarly Materials

Scholarly literature, by nature, has always relied on a network of links. Regardless of the World Wide Web, academic research is anchored in the body of human knowledge; the work builds on prior research and leads to successive investigations. An academic publication always cites previous publications, enabling a scholar to go through a sequence of works—each citing the previous one—and trace the way in which a theory has evolved over time.

When doing research, scholars strive to set their work in the context of the body of literature. They look for previous publications that support their research hypothesis and for publications that describe similar initiatives that are taking place elsewhere at the same time and might support or negate their hypothesis. Scholars employ various information-seeking techniques (Bates 2002), the most obvious of which is searching—an active, directed mode of obtaining information. Another method that has always been used heavily is reliance on the advice of others, primarily colleagues, tutors, research collaborators, and librarians. Scholars have also obtained useful leads at conferences, workshops, and seminars.

However, regardless of how a specific publication is found, it can be seen as a starting point: from that publication, the scholar can reach other publications that may also be relevant and might, at times, be even more relevant than the initial publication. Connections to other publications can be based on citations and other types of relationships—publications written by the same author, dealing with the same topic, presented at the same conference, or developed at the same institution or department. Such associations are based on common attributes: for example, two publications having the same author name are considered related. Automated systems can easily detect such connections because of the structure of the metadata—for example, by comparing the content of the author name field—or even detect evidence in the text, such as a heading called *References*, indicating that the list following it is a reference list.

Another type of association between publications can exist because of the way in which researchers use these publications. For example, all the publications that appear in the reference list of an article or all the publications that a scholar looks at during a single information-seeking session can be regarded as associated. A researcher—or an automated system, for that matter—cannot infer such associations from the publications themselves; rather, the researcher or system identifies the associations from data that is available at a later time, such as citations in later publications or an analysis of usage data (usually by an automated system).

The associations that have been discussed up to now can be broken into two groups: associations based on data or metadata that publications have in common, and associations henceforth referred to as *conceptual links*—links that a researcher or system makes between publications *ex post facto*. Many of these conceptual links could not be easily traced in the pre-automated information-seeking environment. The creation of citation indexes, for example, required great effort. With today's technological advances, the online accessibility of full text, and the availability of usage data, information systems can turn conceptual links into actual hypertext links (such as a link from an article to another one that cites it). The current technological challenge relates to the comprehensiveness of the information that such systems cover. For example, researchers consider the citation index in SPIRES very reliable within the HEP field. However, cross-disciplinary citations are not well covered because SPIRES includes only pure HEP data and excludes citations by scholars in other fields, even related ones. Usage data, too, needs to be comprehensive in order to provide a robust basis for an analysis of user behaviour.

Before the introduction of automated library systems, researchers could follow connections between documents by working through reference lists, looking at items located near a publication of interest on the physical bookshelf, looking for materials published by the same author or authors, or relying on explicit recommendations by others. Automated systems—library catalogues at first, abstracting and indexing databases later, and, in the last decade, Web search engines and other information systems—shifted the information-seeking process to a greater reliance on searching. As a result, scholars employ various searching techniques, aiming to gather as many relevant materials as possible through searching. However, searching as conducted today requires that the searcher have some knowledge about the desired material. In fact, one can regard searching as a hypothesis about the kind of keywords that would be found in or assigned to the sought material (Robertson 2010). Therefore, a search can lead the searcher only to documents that correspond with the searcher's prior knowledge, and it is through the expansion of the search to closely or loosely related materials that the searcher's horizons can be broadened.

This work suggests that a complete information-seeking environment should not be confined to searching alone. Although good searching is crucial, a model that relates to information seeking—and the system that is derived from such a model—should take into account the various associations that exist between scholarly materials. A model that does not describe a combination of searching and following links that are defined on the basis of these associations is not likely to portray the true essence of information seeking in the scholarly environment.

In a schematic representation of this view of information seeking, the scholarly materials are linked objects. In Figure 22, the document in focus (the outlined square) in A is related to other publications in various ways; each colour indicates a type of relationship. When a researcher moves to a related document (in B), that document becomes the one in focus (in C), and the network of links changes to reflect the relationships of the other materials to that document. The researcher can keep moving from one document to another, each time following the new paths that become visible when the focus is transferred to a different document.

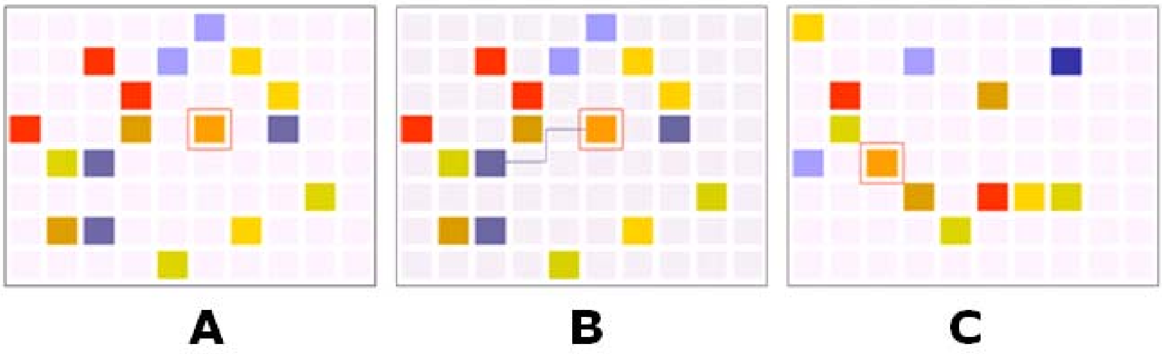


Figure 22: A representation of scholarly materials as linked objects.
The squares represent documents in the information landscape. Each document is related to other documents through various associations, indicated by the different colours.

An interaction of this kind—moving from one document to another—can be viewed as adhering to the berrypicking model described by Bates (1989) (see 2.2.2.5). However, the move from one document to another is not necessarily done by searching (formulating a new search or modifying an existing search); rather, it can be done by following leads. Such a mode of interaction complements straightforward searching.

A relationship between documents can exist on more than one dimension. For example, although a published article is the final product, other versions of the article (such as preprints or the version posted on the author's home page), supporting data (such as datasets and reports), and related data (such as lecture notes and presentation slides) are all conceptually connected to the final product. Often a dataset or a preliminary version of a published article can be of great interest, and notes from a talk given by the author on the same topic can shed light on the content of the article.

Another dimension of document-document relationships can exist between a scholarly document and a reference document—for example, an association between the name of a chemical appearing in the text of a document and a description of that chemical in a reference source, such as an online encyclopaedia. Although in this case the association is based on words appearing in the text, it is not derived from the structure of the document (its metadata or some other indication in the text such as a heading). Similar to the way in which people used to read books in foreign languages before the computerized age, with the aid of a physical dictionary, automated systems can provide links from any word in a text to a dictionary or other reference tool. Furthermore, the links are not necessarily provided by the information system. Today they can be generated by a system, such as an online dictionary, that is independent of the information system. The online dictionary or reference tool detects words on the screen and, when so requested by the user, brings up a reference document.

Moving from one data item to another in the cases just described is different by nature from moving from one published work to another but can be described in the context of the same information-seeking model.

5.2 The User

The suggested model describes the information-seeking behaviour of HEP scientists. The model is based on the characteristics of the community members that are described in Chapter 3 and that served to create the personas described in Chapter 4. Nevertheless, this model is likely to be applicable to other scientific communities.

5.3 Existing Models Applied to HEP

Most models of information seeking, which are general in nature, can be applied to the HEP community's practices. However, such models are either too general to provide valuable insight or do not cover the entire information-seeking behaviour of this specific community. Before a comprehensive model is suggested for the HEP community, some existing models (briefly described in Chapter 2) are examined in relation to this community.

Bates (2002) suggests a model of modes of information seeking, all of which can be applied to the HEP community, and identifies a matrix whose rows represent directed and undirected modes, respectively, and whose columns represent active and passive information-seeking modes (see 2.2.2.5). Bates defines the terms as follows:

> 'Directed' and 'Undirected' refer, respectively, to whether an individual seeks particular information that can be specified to some degree, or is more or less randomly exposing themselves to

information. ‘Active’ and ‘Passive’ refer, respectively, to whether the individual does anything actively to acquire information, or is passively available to absorb information, but does not seek it out. (Bates 2002)

Bates also labels each cell in the matrix; active, directed mode is ‘searching’; passive, directed mode is ‘monitoring’; active, undirected mode is ‘browsing’; and passive, undirected mode is ‘being aware’.

Figure 23 describes the modes of information seeking in the context of the HEP community. However, in terms of the information-seeking behaviour of HEP researchers, the general, well-defined boundaries between the modes are blurred.

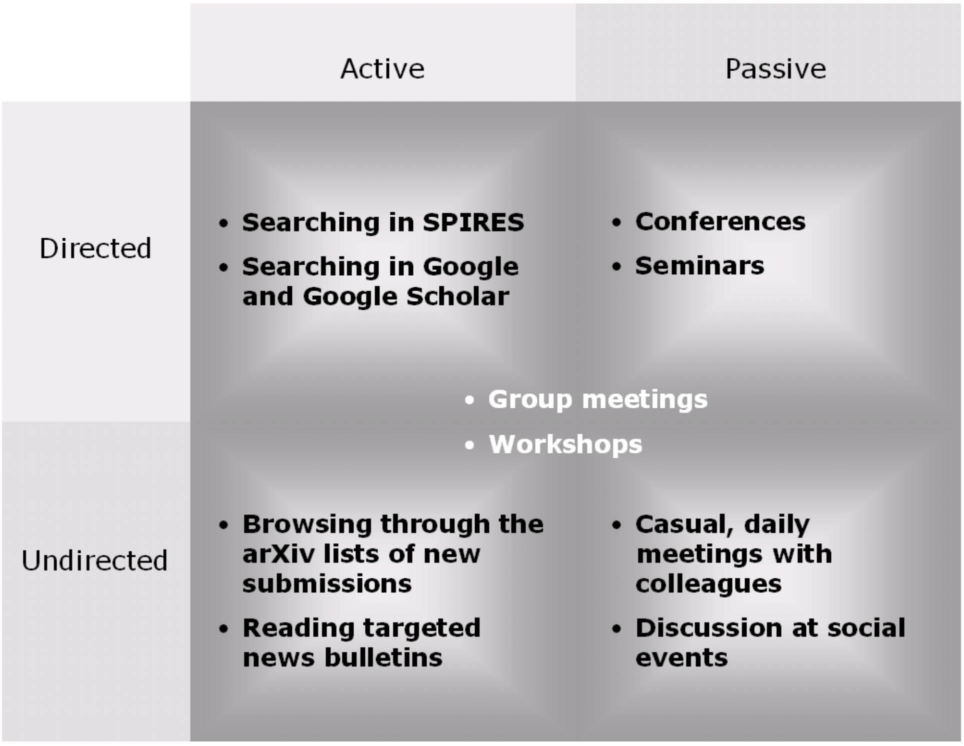


Figure 23: The Bates model applied to HEP

The ways in which HEP researchers actively acquire information can be categorized as directed—when researchers state explicitly in a search query what they are seeking in an information system—or undirected—when they examine new submissions, typically on a daily basis, and occasionally browse through other resources. In addition, researchers are always receptive to information and rely on a variety of information forums, either those that the researcher intentionally selects, such as conferences and seminars,[17] or those that occur spontaneously, such as casual meetings with colleagues or discussions at social gatherings. Information is also obtained in group meetings and workshops, which are at the intersection between directed and undirected and active and passive information-seeking behaviour. Furthermore, as one event leads to another, an information-seeking path may evolve; it might start as an undirected, passive event—for example, a casual conversation during a lunch break—and lead to a completely different mode—for example, an active search for materials regarding a topic mentioned at lunch.

The HEP community's information-seeking practices can also be mapped to Belkin et al.'s (1994) four 'modes', or dimensions, of an information seeking strategy: method of interaction, goal of interaction, mode of retrieval, and resource considered (see 2.2.2.6). Figure 24 and Figure 25 illustrate Belkin's dimensions in reference to the two most common activities of HEP researchers—active searching for materials in SPIRES and browsing through a daily list of new submissions in arXiv.

[17] Interviewees described professional meetings as follows:

- A workshop is a multi-institutional, typically international event. Unlike a conference, a workshop is set around a specific research topic that is of interest to all attendees. It involves presentations, setting the context, but the main objective of a workshop is to have in-depth discussions and work together on specific issues.
- A seminar is a departmental or institutional meeting where a researcher—either from the institution or a visitor—presents a topic. It could be a weekly or an ad hoc event.
- A group meeting is an informal meeting of a research group that assembles on a regular basis. In this forum, colleagues talk about their work and share their news. At times they may present new findings or discuss works of others.

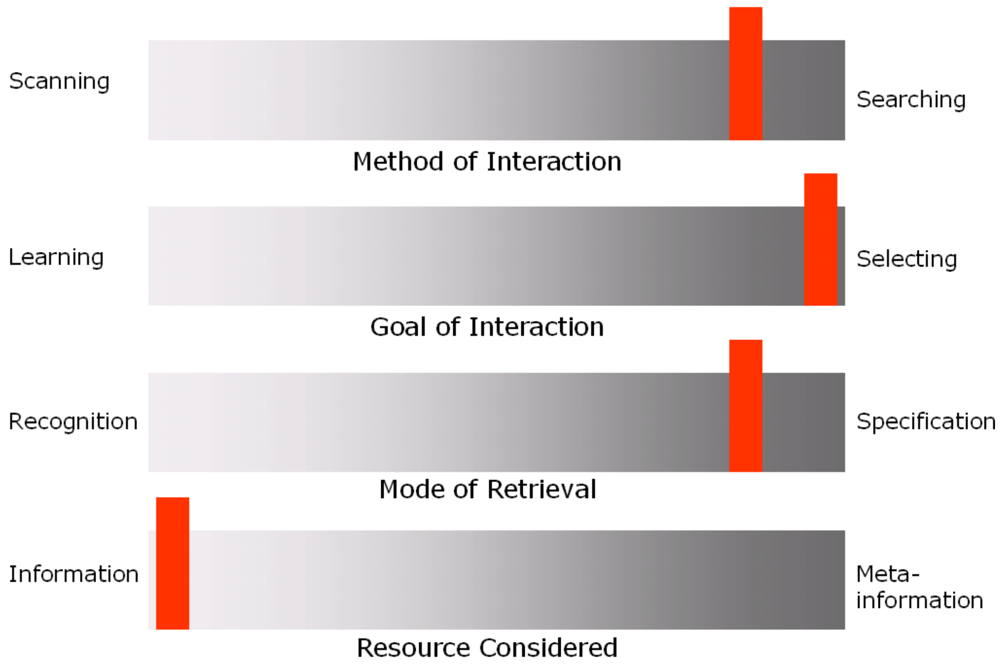


Figure 24: The four dimensions of the action of searching for content in SPIRES. The four dimensions were proposed by Belkin et al. (1994). The red bars represent the location of the action along each dimension.

As seen in the figure, searching for materials in SPIRES—a strategy employed by most HEP researchers—constitutes intentional searching with the goal of selecting materials. A researcher specifies the required materials to a greater or lesser degree of completeness, depending on the researcher's information needs and prior knowledge, and, in most cases, is interested in the document itself, not the metadata describing it. Nevertheless, the researcher is likely to scan lists of results, mainly when the search is not for a specific item and hence could yield many results. In this case, selecting the appropriate material involves recognition: looking at the result list, the researcher picks up the materials that seem relevant to the specific need.

Monitoring daily submissions in arXiv (Figure 25) consists first of a scanning interaction (even though the list is prefiltered according to the researcher's preferences). The goal of the interaction extends from learning to selecting: although most of the interaction is aimed at enabling the researcher to keep abreast of new developments, some submissions may attract the attention of the researcher and trigger a selection action. The mode of retrieval is again confined to one kind of interaction—the researcher decides whether a document is of interest by recognition. When it comes to considering a resource, the interaction is once again less defined. In most cases, the researcher just looks through the list and is satisfied by the metadata, including the abstract; however, in some cases the researcher selects a document for the purpose of reading it immediately or at a later time.

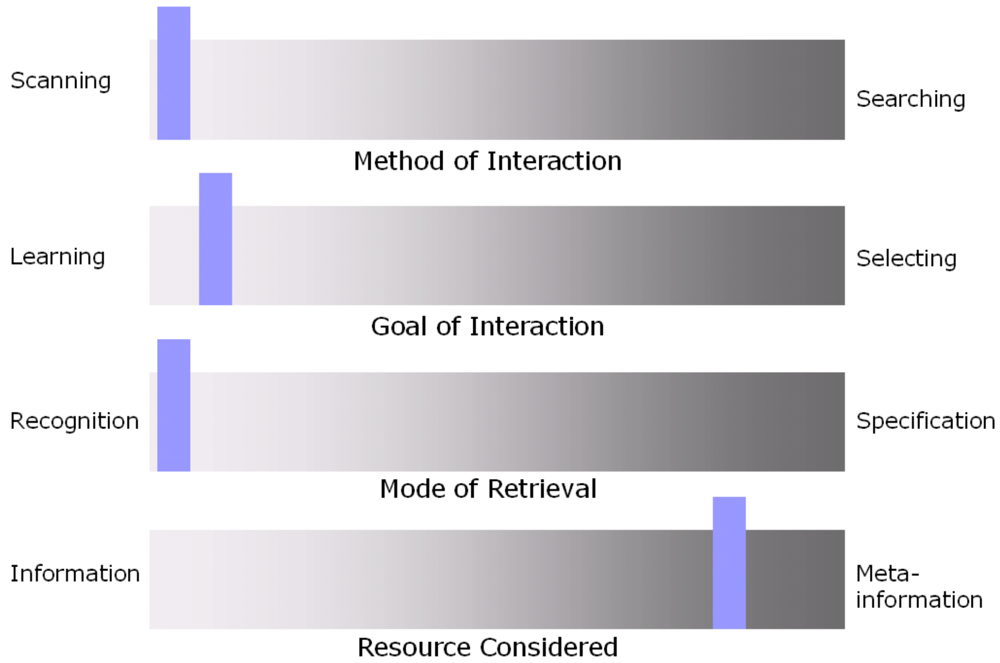


Figure 25: The four dimensions of browsing through new arXiv submissions. The four dimensions were proposed by Belkin et al. (1994). The blue bars represent the location of the action along each dimension.

Wilson's (1981) information-behaviour model (see 2.2.2.1) (Figure 26), though applicable to the HEP researchers' behaviour, has several limitations that cause it to fall short of covering the range of behaviour that is portrayed in this study. First, the model depicts a 'need' as a trigger for the information-seeking process (*1* in Figure 26)—'information-seeking behaviour results from the recognition of some need, perceived by the user' (Wilson 1981). If one defines 'information need' as a state of mind, one can conclude that HEP scientists are always in a situation in which they have an information need, albeit not a specific one; on the basis of this definition, Wilson's model can be deemed applicable to the information-seeking behaviour of the HEP scientists. However, if 'information need' is defined as a specific, conscious gap in a particular researcher's knowledge at a given moment, the model, though applicable to that researcher's directed information-seeking behaviour, does not describe the entire range of information behaviour of HEP researchers, such as their reliance on constant updates (mainly finding out about new submissions to arXiv) regardless of a specific need at a given moment.

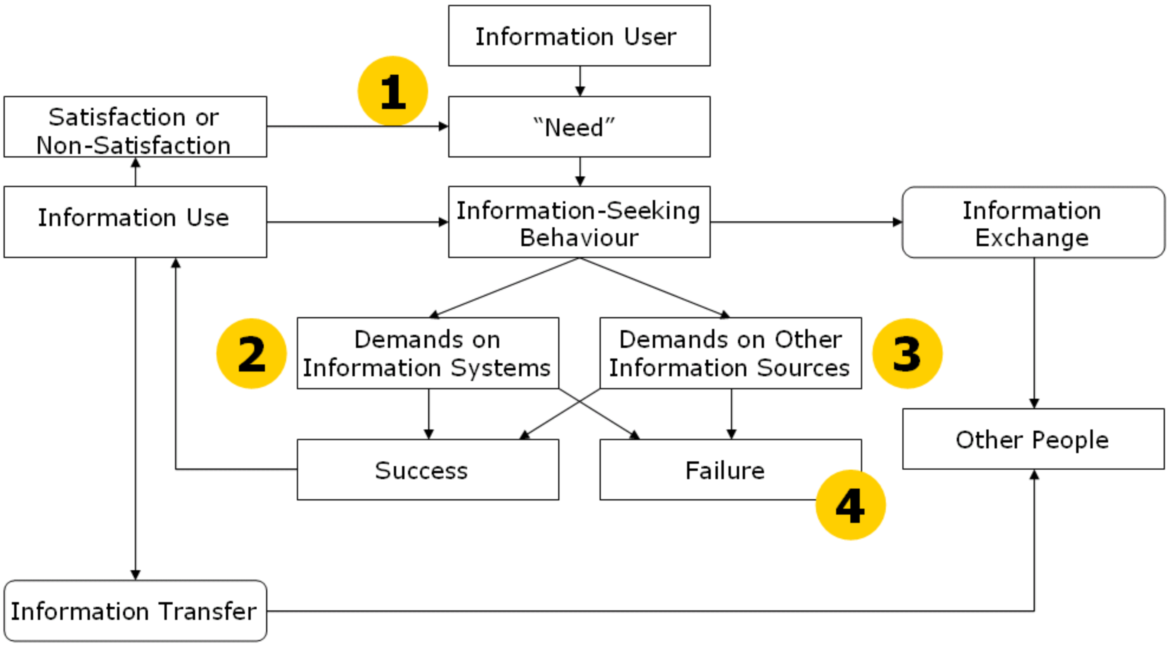


Figure 26: Wilson's information behaviour model.
Modified from Wilson (1981), Figure 1.

Referring to a stage as 'demands on information systems' (*2* in Figure 26) is perhaps too great a generalization, as it fails to reflect the various kinds of demands—the search for a known item as opposed to an exploratory search, for example—that can lead to the use of multiple systems in different ways. Wilson (1981) treats the different kinds of demands in his second model (Figure 27).

In addition to demands on information systems, Wilson's (1981) information-behaviour model includes demands on 'other' information sources (*3* in Figure 26), which he defines as 'systems which may perform information functions in addition to a primary, non-information function (such as estate agents' offices or car sales agencies, both of which are concerned with selling, but which may be used to obtain information on current prices, areas of 'suitable' housing, or details of cars that hold their secondhand value)' (Wilson 1981). However, these demands do not seem to apply to HEP researchers today; perhaps these demands are rooted in the reality of the period during which Wilson developed his model. Although HEP researchers sometimes need information that is not found in scholarly information systems—for example, information about instruments or companies—the way they obtain it is not through mediators but through one of the information systems they use—in this case, typically Google, which would direct them to information found on the Web.

Furthermore, Wilson's (1981) model describes failure as a dead end (*4* in Figure 26). In today's reality, there is always a way to continue, even, in extreme cases, by adjusting the information need.

On the other hand, the model appropriately depicts the human interaction that characterizes the HEP community—the exchange of information and the reliance on cross-institutional human networking for obtaining information.

Wilson's second model from 1981 (Figure 27) focuses on the information-seeking behaviour of users. In this model, the information need is described as a secondary need that emerges from more basic needs—physiological, affective, or cognitive—each of which is set in the context of a person's personal life and global environment. The same context, according to Wilson, will be the source for the barriers that may impede the person's search for information. In this model, Wilson also incorporates the feature set from Ellis (1989, as described by Wilson [1999] and Ingwersen and Järvelin [2005], among others), which complements his model in regard to information-searching behaviour (for a discussion of Ellis's feature set, see 2.2.2.3).

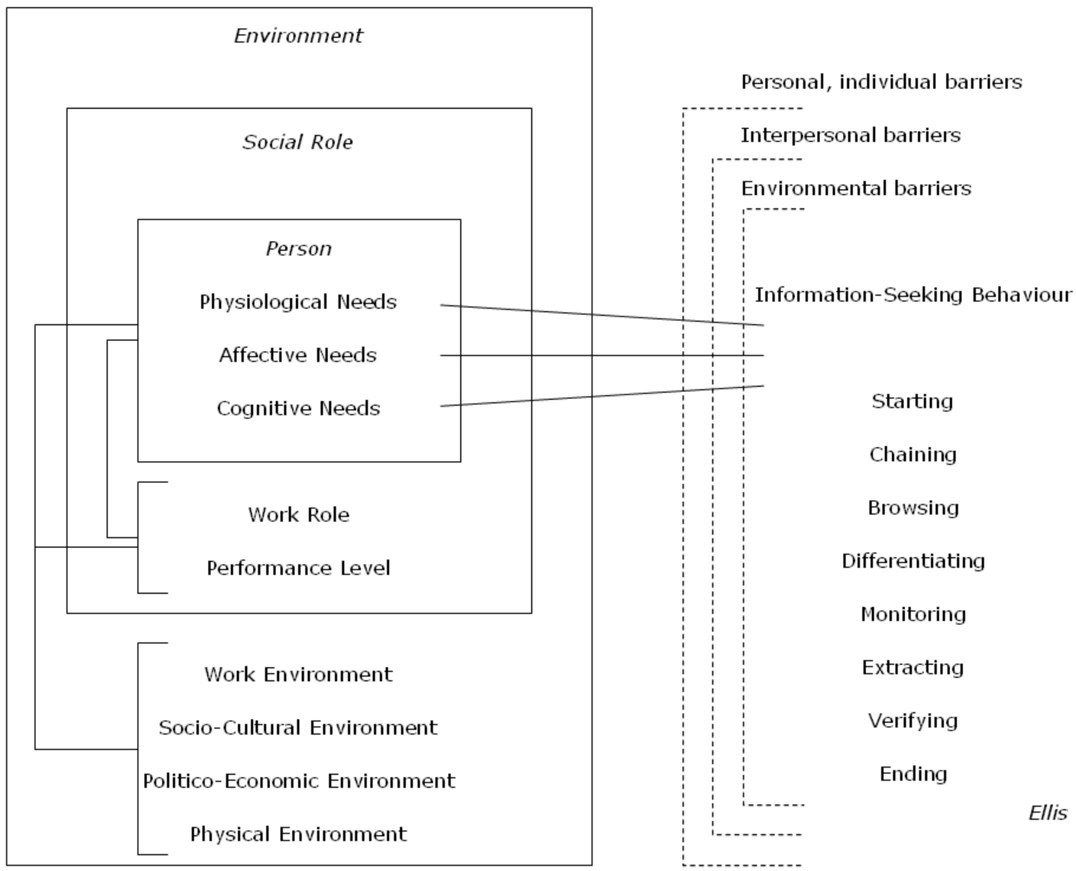


Figure 27: Wilson's information-seeking behaviour model.
Modified from Wilson (1999), Figure 2.

Because the focus of this thesis is on a well-defined community acting in a specific, work-related context, the comprehensiveness of the model presented in Figure 27 is of little use in the present study. However, if one modifies the model to describe the more specific context of the HEP researchers (for example, by eliminating the role of physiological and affective needs, which seem negligible for the HEP researchers' information-seeking behaviour), one can, indeed, apply the model to this study.

In 1996, Wilson revised his earlier information behaviour model (presented in Figure 26) to propose the use of theories from various other disciplines—decision making, psychology, innovation, health communication, and consumer research—in the analysis of information behaviour (Figure 28). The model depicts information seeking as being activated in a context (*1* in Figure 28) and having various modes (*2* in Figure 28). These modes (*3* in Figure 28) are very relevant to the modes characterizing the HEP community: passive attention, passive search, active search, and ongoing search (in the earlier model, presented in Figure 26, the active search was the focus of attention). Information seeking, in this newer model, results in information processing and use (*4* in Figure 28), which in turn leads to an examination of the information need's fulfilment (based on the user's context) (*5* in Figure 28).

This model sheds light on certain aspects of information seeking and incorporates three theories that are adapted from other research fields:

- The theory of stress/copying can explain why some information needs do not invoke information seeking.
- The theory of risk/reward may help explain the tendency of individuals to prefer some sources of information over others.
- The theory of social learning may help explain how people learn information-seeking behaviour through observational learning.

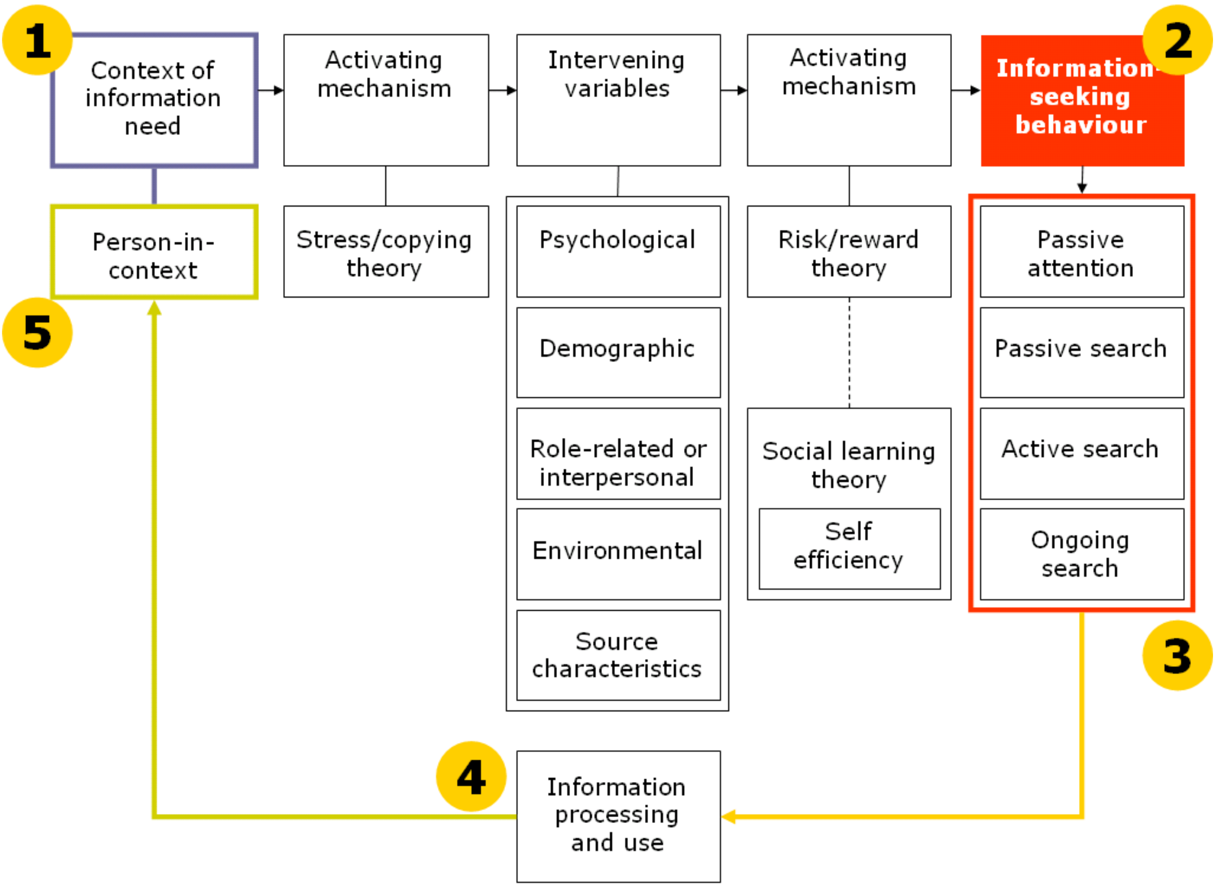


Figure 28: Wilson's 1996 model of information behaviour.
Modified from Wilson (1999), Figure 7.

Wilson (1999) notes that this model 'remains one of macro-behaviour, but its expansion and the inclusion of other theoretical models of behaviour makes it a richer source of hypotheses and further research than Wilson's earlier model' (Wilson 1999).

The behavioural model of information-seeking patterns suggested by Ellis (1989, as described by Wilson [1999] and Ingwersen and Järvelin [2005], among others) (see 2.2.2.3) can be applied to information-seeking activities of HEP scientists as well as those of other communities. Ellis does not specify the exact order and relationships within the set of eight features of information-seeking behaviour that he proposes—*starting*, *chaining*, *browsing*, *differentiating*, *monitoring*, *extracting*, *verifying*, and *ending*—with the exception of *starting* and *ending* (see Wilson's sequencing of Ellis's feature set in Figure 29).

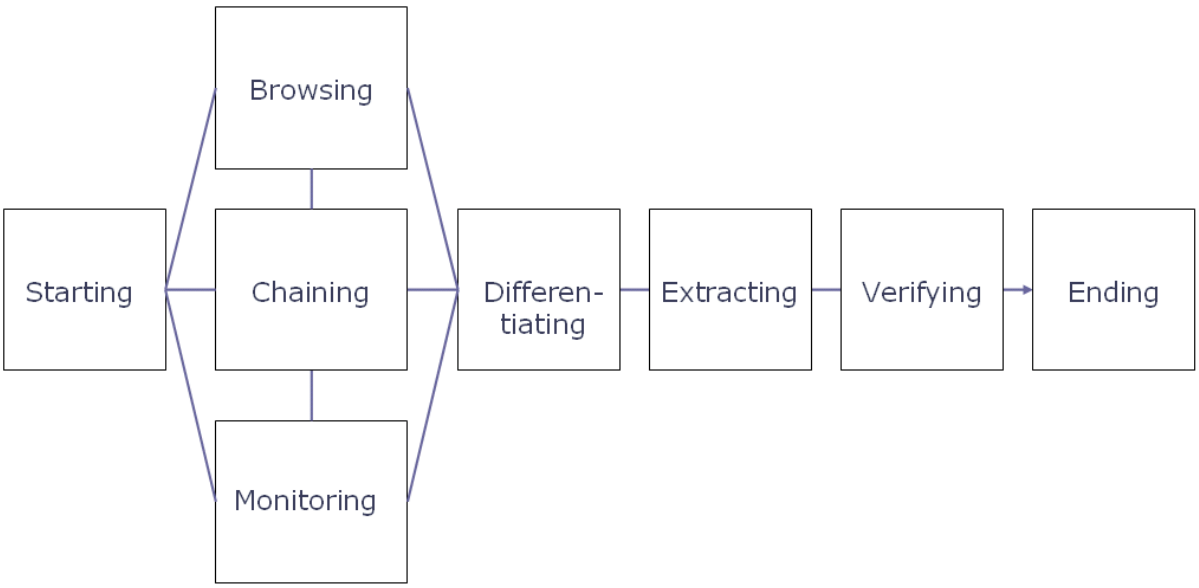


Figure 29: Wilson's sequence of Ellis's feature set of information-seeking behaviour. Modified from Wilson (1999), Figure 5.

However, a different sequence, presented in Figure 30, is arguably a more appropriate way of describing the information-seeking processes of HEP scientists. In this sequence, three entry points are defined. The s*tarting* feature described by Ellis is mapped to active, directed searching, while the *monitoring* feature is mapped to the scanning of information, such as checking lists of new submissions in arXiv. An example of the *browsing* feature might be looking through a table of contents (TOC) of an e-journal (a behaviour that is not very common among HEP researchers). *Differentiating* can be regarded as refining result lists; *extracting*, as focusing on an item and using it—for example, downloading the full text and reading it—and *chaining*, as the process of following links to previous articles, later articles, or related articles.

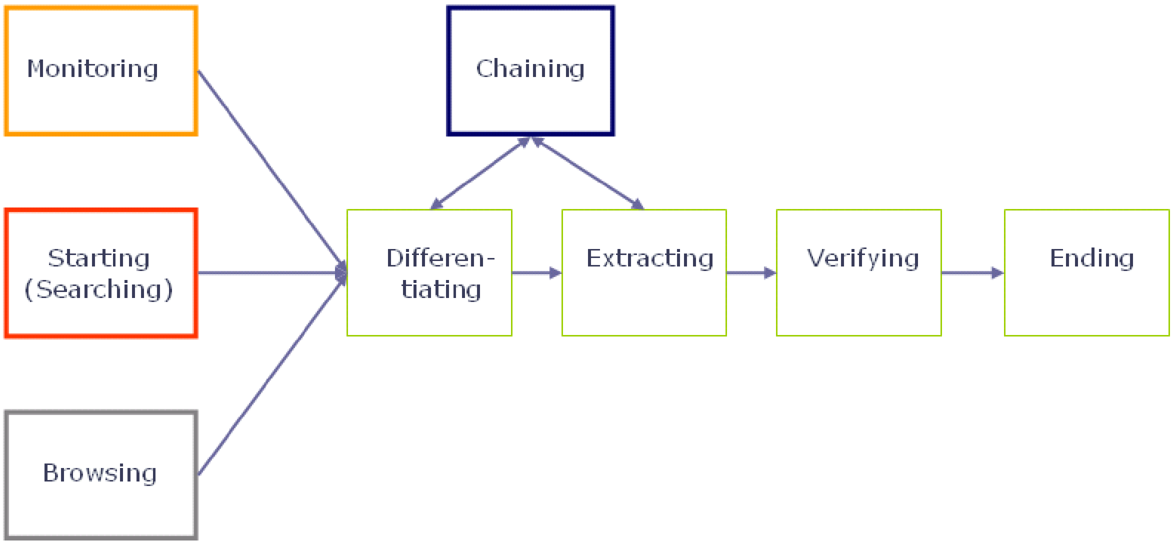


Figure 30: Suggested sequence of Ellis's feature set

In Figure 31, the features defined by Ellis are applied to the information-seeking activities of HEP researchers.

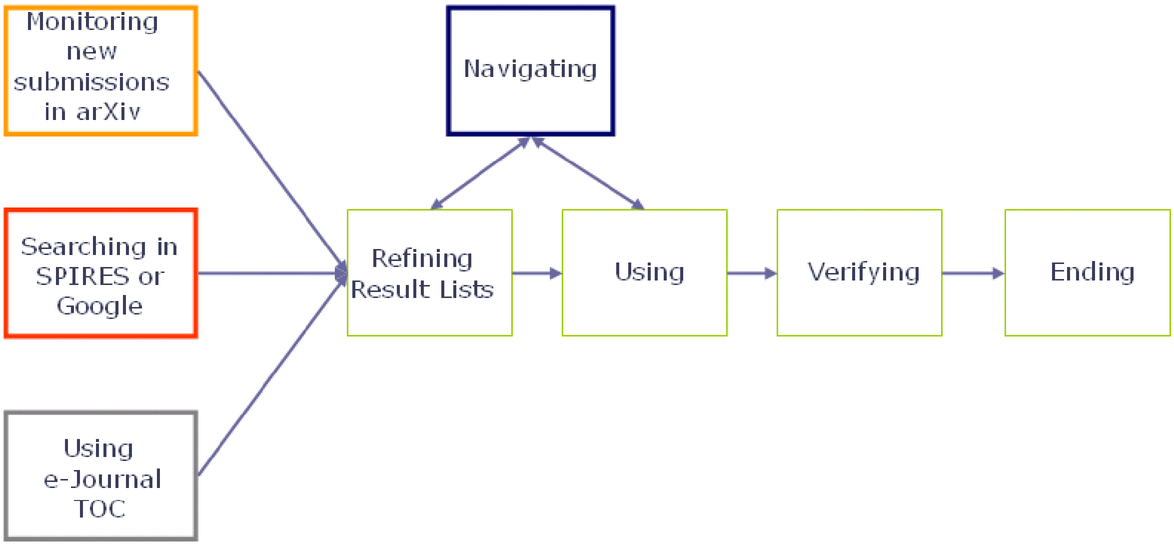


Figure 31: Ellis's feature set adapted to HEP practices

Even with these variations in the sequence of Ellis's feature set, the fixed beginning and end of the process render this model questionable for the HEP community. More relationships between features would be needed to make the flow more cyclical and thus more reflective of the information-seeking flow of HEP researchers (see 5.4).

Marchionini (1995) (see 2.2.2.7) defines an information-seeking process that is also composed of subprocesses or functions, some of which are very similar to Ellis's features; however, in Marchionini's model, the subprocesses are sequenced as described in Figure 32.

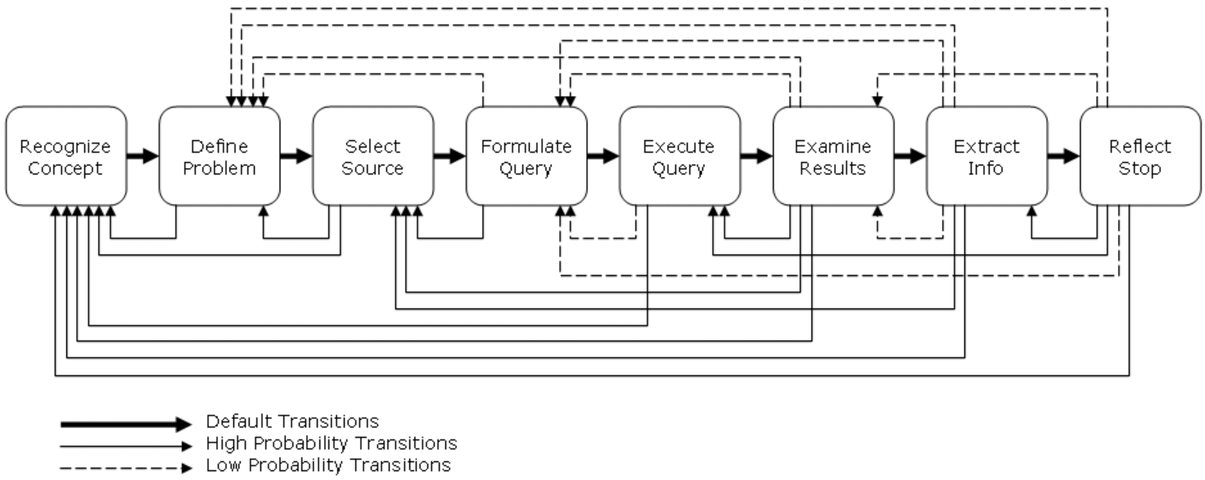


Figure 32: Marchionini's information-seeking model.
Modified from Marchionini (1995), Figure 3.3.

Marchionini describes this process as ‘both systematic and opportunistic’ (Marchionini 1995, 49). Subprocesses ‘are better considered as…activity modules that may be called into action recursively at any time, that may be continuously active…, that are temporarily frozen while others proceed, and that may make calls to other subprocesses. Thus, the information-seeking process can proceed along parallel lines of progress and take advantage of opportunities arising from intermediate or random results’ (Marchionini 1995, 49-51).

The subprocess of choosing a search system (*Select Source* in Figure 32) resonates well with the evidence gathered about the HEP researchers’ behaviour: ‘Choosing a search system is dependent on the information seeker’s previous experience with the task domain, the scope of his or her personal information infrastructure, and the expectations about the answer that may have been formed while defining the problem and the task’ (Marchionini 1995, 52).

While the generalized subprocesses and extensive and flexible relationships described by Marchionini (1995) make the mapping of almost any directed information-seeking process possible, the model falls short in covering the non-directed processes and does not present a clear picture that characterizes a specific community such as the HEP community.

5.4 HEP Information-Seeking Model

Although existing models are applicable to some degree to the information-seeking practices of the HEP community members, such models do not adequately describe the full spectrum of the information-seeking behaviour of the HEP scientists. This work, therefore, aims to suggest a new model that leverages existing models and complements them to provide a comprehensive and accurate portrayal of active information seeking in the context of the HEP community.[18]

The focus is on active information seeking rather than the entire range of information-seeking behaviour because the active behaviour consists of specific, describable actions and hence lends itself to modelling in a clear, comprehensive manner. The inclusion of passive information-seeking activities in the model—activities such as participation in conferences, seminars, group meetings, and workshops; casual meetings at the workplace and discussion at social events; and the supervision of students, to name just a few—is likely to make the model too complex and less purposeful . Furthermore, because this study aims to improve the active information-seeking process of a scientific community, passive behaviour is not a fundamental aspect of the discussion.

This work describes both directed and undirected active information seeking and suggests an inclusive model (Figure 33). Although the model has two different starting points (*a* and *b*), there is no clear end point, because information seeking is a constant state for HEP researchers. They may satisfy a specific need and exit the process at any stage, yet they are always information receptive and are likely to return to the process before long. A later section of this work (5.5) focuses on information searching, which is considered part of the overall information-seeking process.

---

[18] The author first presented this model on June 10, 2010, at the ELAG conference in Helsinki as part of a talk entitled Meaningful Relevance Ranking for Heterogeneous Scholarly Materials.

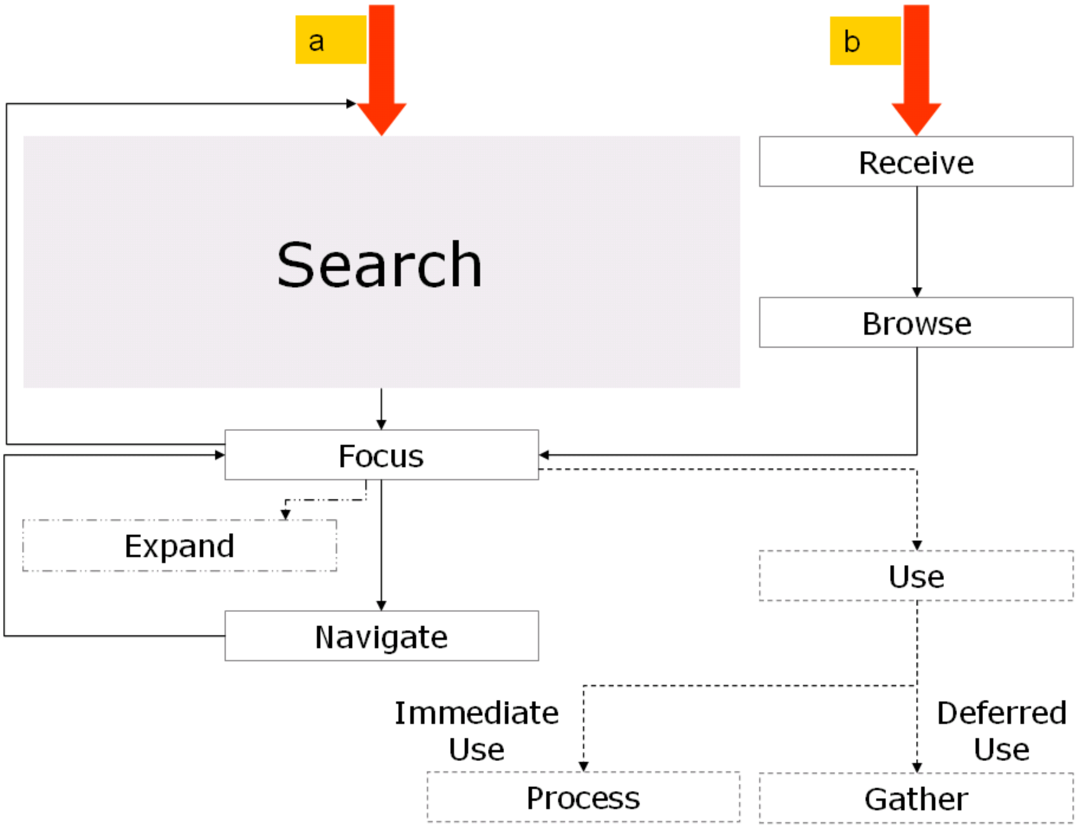


Figure 33: Information-seeking model of HEP researchers

As illustrated in Figure 33, the model describes two processes of active information seeking: directed—that is, searching for information that can be described to some extent—which starts at point *a*, and undirected—the scanning of information without having a specific information need in mind—which starts at point *b*. The nodes represent the interaction of the searcher with either an automated system or real-world elements (such as a physical library or a human being) and are labelled as actions (for instance, Receive, Browse, and Focus), while the arrows indicate the course of events. Actions that are at the core of the process are framed with solid lines; actions that relate to using a document or to checking additional information about a document, which are therefore dead-end sub-processes, are indicated by dotted lines. Both the directed and the undirected information-seeking processes typically involve the researcher's focusing on a specific document. The general model expressed in Figure 33 describes the Focus action and the actions that follow it that are relevant to both processes (for a model of the search process, see 5.5).

The undirected process starts with the Receive action: the researcher obtains a list, either by accessing a dedicated Web page or by receiving an automated update (an RSS feed or an e-mail message, for instance). Regarding the HEP community, all these updating options are provided as arXiv services. Such a list can also be obtained from a person—a supervisor, an instructor, or a librarian. A list of references at the end of an article can be considered relevant to this process, too. However, researchers would typically not regard a list of references as an independent source of information, detached from its context; they would use the list as means to link to another document that caught their attention while they were reading the document that contains the list.

The next action is Browse: the researcher goes through the list and examines the items on it. In most cases, only the metadata (including the abstract) is of interest to the user, but in some cases, the user wishes to examine a document more closely and focuses on it. The Focus action is the intersection of the two paths—that of the directed information seeking and that of the undirected information seeking.

Once the researcher focuses on a document, he or she may wish to check additional information that relates to textual elements and figures given in the text. Conceptually, the goal of such behaviour is the ‘expansion’ of the document to include, for example, an explanation about a chemical, a map of a place, or a translation of a word. Other ways of expanding a document might bring up earlier versions of the article or a dataset that resulted in a specific graph.

At this point, the researcher can also decide to use the document; that is, obtain the full text (or other media) and either process it immediately (read it or cite it) or gather it along with other documents to process it at a later time. Regardless of usage—immediate or deferred—the researcher may choose to navigate to other documents, following links of various types. While navigating, the researcher focuses on one document at a time. The process is iterative, and at any point, the researcher can choose to continue to navigate or to invoke a new search, possibly while using metadata elements of the material in focus as search terms.

In the HEP context, the process described in Figure 33 is carried out by means of several information systems. For example, researchers often start their day by monitoring arXiv for the new submissions added the day before. A researcher may then choose the New Submission page of the field of interest (for example, High Energy Physics—Theory) (Figure 34).

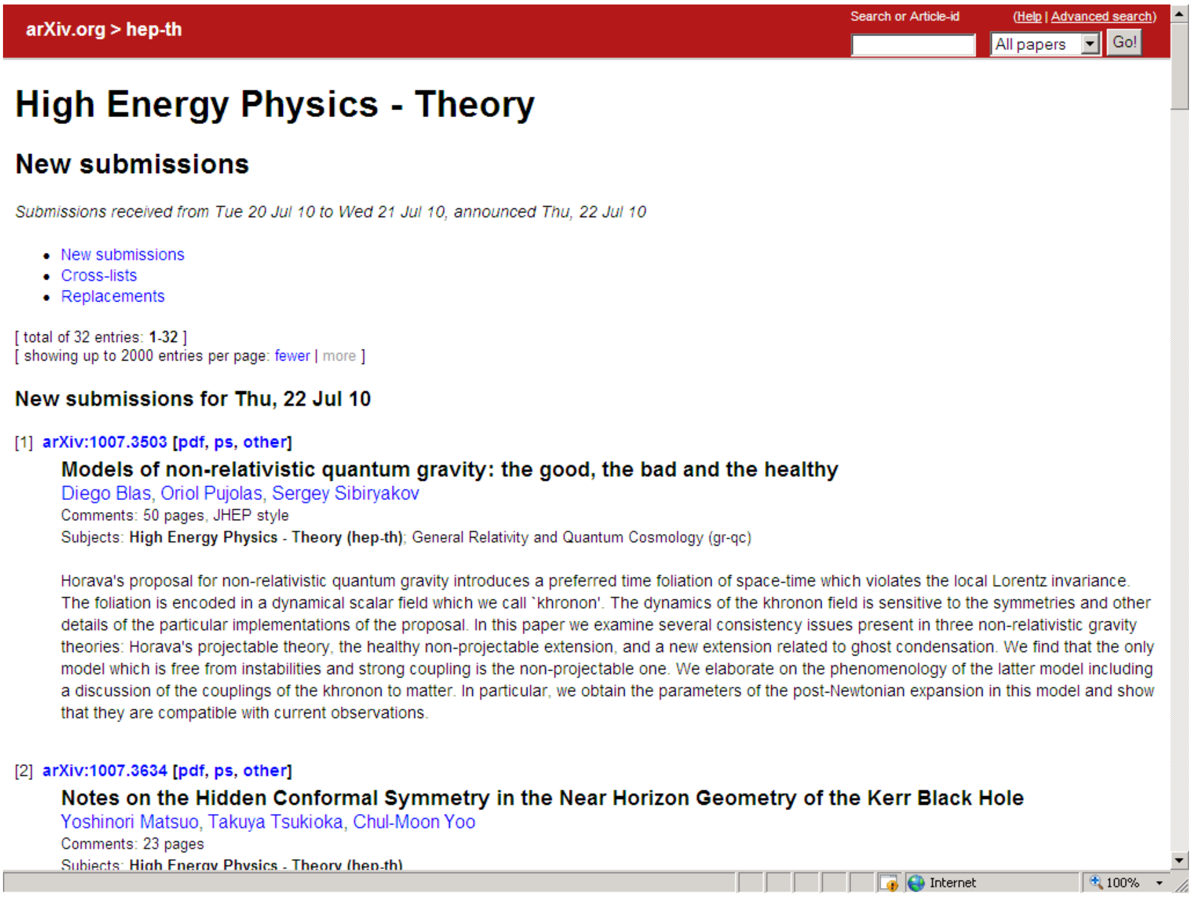


Figure 34: A page of new submissions

The researcher would browse the list and perhaps decide to focus on a specific item and look at more details or download the document. Once a specific document is in focus, the researcher can start navigating—for example, click an author's name to look for other articles in arXiv by that author, or, after reading the article, link to one of the articles that it references. It could well be that the researcher decides to invoke a search for other articles that discuss a topic raised in this article. In this case, researchers are likely to use another information system—typically SPIRES but possibly Google or Google Scholar—and launch a new search regarding that topic.

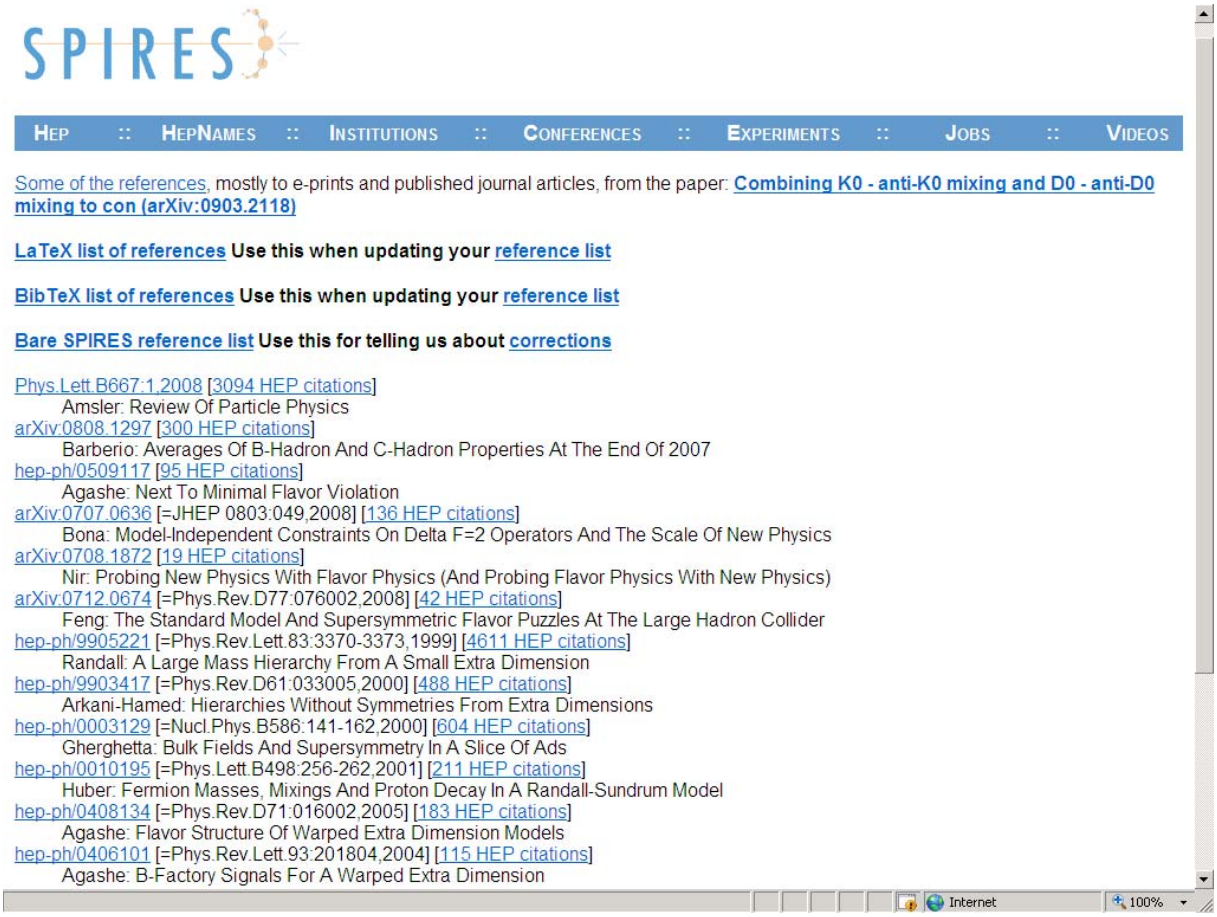
Once the researcher focuses on an item in SPIRES, more navigation options become available. For example, the researcher may navigate to articles citing the article in focus or to articles sharing the same keywords (Figure 35). In addition, the researcher can link to the full text of the article in arXiv or from the publisher's site.

14) **Testing New Indirect CP Violation.**
Yuval Grossman, (Cornell U., CIHEP) , Yosef Nir, Gilad Perez, (Weizmann Inst.) . Apr 2009. (Published Aug 14, 2009). 5pp.
Published in **Phys.Rev.Lett.103:071602,2009**.
e-Print: **arXiv:0904.0305** [hep-ph]

References | LaTeX(US) | LaTeX(EU) | Harvmac | BibTeX | Keywords | Cited 21 times
Abstract and Postscript and PDF from arXiv.org (mirrors: au br cn de es fr il in it jp kr ru tw uk za aps lanl)
Journal Server [doi:10.1103/PhysRevLett.103.071602]
Bookmarkable link to this information

Figure 35: A metadata record displayed in SPIRES.
The display enables researchers to navigate to other records.

SPIRES was one of the first information systems to present the number of articles citing the article in focus and to provide links to these articles. The citation information is displayed in several contexts. For example, if the researcher chooses to see the references, SPIRES presents the list of references, each followed by a link to a list of citations for that reference (Figure 36), thus enabling the researcher to follow a scientific idea back to its origin or forward to the various shapes and flavours that it has assumed over time.


SPIRES

HEP :: HEPNAMES :: INSTITUTIONS :: CONFERENCES :: EXPERIMENTS :: JOBS :: VIDEOS

Some of the references, mostly to e-prints and published journal articles, from the paper: Combining K0 - anti-K0 mixing and D0 - anti-D0 mixing to con (arXiv:0903.2118)

LaTeX list of references Use this when updating your reference list

BibTeX list of references Use this when updating your reference list

Bare SPIRES reference list Use this for telling us about corrections

Phys.Lett.B667:1,2008 [3094 HEP citations]
Amsler: Review Of Particle Physics
arXiv:0808.1297 [300 HEP citations]
Barberio: Averages Of B-Hadron And C-Hadron Properties At The End Of 2007
hep-ph/0509117 [95 HEP citations]
Agashe: Next To Minimal Flavor Violation
arXiv:0707.0636 [=JHEP 0803:049,2008] [136 HEP citations]
Bona: Model-Independent Constraints On Delta F=2 Operators And The Scale Of New Physics
arXiv:0708.1872 [19 HEP citations]
Nir: Probing New Physics With Flavor Physics (And Probing Flavor Physics With New Physics)
arXiv:0712.0674 [=Phys.Rev.D77:076002,2008] [42 HEP citations]
Feng: The Standard Model And Supersymmetric Flavor Puzzles At The Large Hadron Collider
hep-ph/9905221 [=Phys.Rev.Lett.83:3370-3373,1999] [4611 HEP citations]
Randall: A Large Mass Hierarchy From A Small Extra Dimension
hep-ph/9903417 [=Phys.Rev.D61:033005,2000] [488 HEP citations]
Arkani-Hamed: Hierarchies Without Symmetries From Extra Dimensions
hep-ph/0003129 [=Nucl.Phys.B586:141-162,2000] [604 HEP citations]
Gherghetta: Bulk Fields And Supersymmetry In A Slice Of Ads
hep-ph/0010195 [=Phys.Lett.B498:256-262,2001] [211 HEP citations]
Huber: Fermion Masses, Mixings And Proton Decay In A Randall-Sundrum Model
hep-ph/0408134 [=Phys.Rev.D71:016002,2005] [183 HEP citations]
Agashe: Flavor Structure Of Warped Extra Dimension Models
hep-ph/0406101 [=Phys.Rev.Lett.93:201804,2004] [115 HEP citations]
Agashe: B-Factory Signals For A Warped Extra Dimension

Figure 36: Reference list in SRIRES

5.5 HEP Information-Searching Model

As explained earlier, information searching is an active, directed process that is part of the more general activity of information seeking. When one of the interviewees was asked to describe the kind of active searches that he typically carries out and to list his goals in such searches from the most common to the least common, he provided the following list:

1. See if someone already had the same idea I have and what happened to the idea
2. Find evidence to confirm (or reject) an idea that came to mind
3. Find evidence to support or dispute someone else's theory
4. Find information about a person or a department

5. Look at the evolution of an idea—either backward (where it came from) or forward (how it was received in later publications and what it led to)
6. See what's new in the work of a specific person (or team) because I know that they are doing interesting or relevant work
7. Satisfy my need to know what's new, in general, not necessarily related to my field

A model of information searching that addresses all these needs is described in Figure 37. Although most of the discussion about this model relates to searching in automated information systems, the interaction described applies also to querying a human being. Once again, this model—which portrays information-searching behaviour of the HEP community—is general enough to apply to other scientific communities.

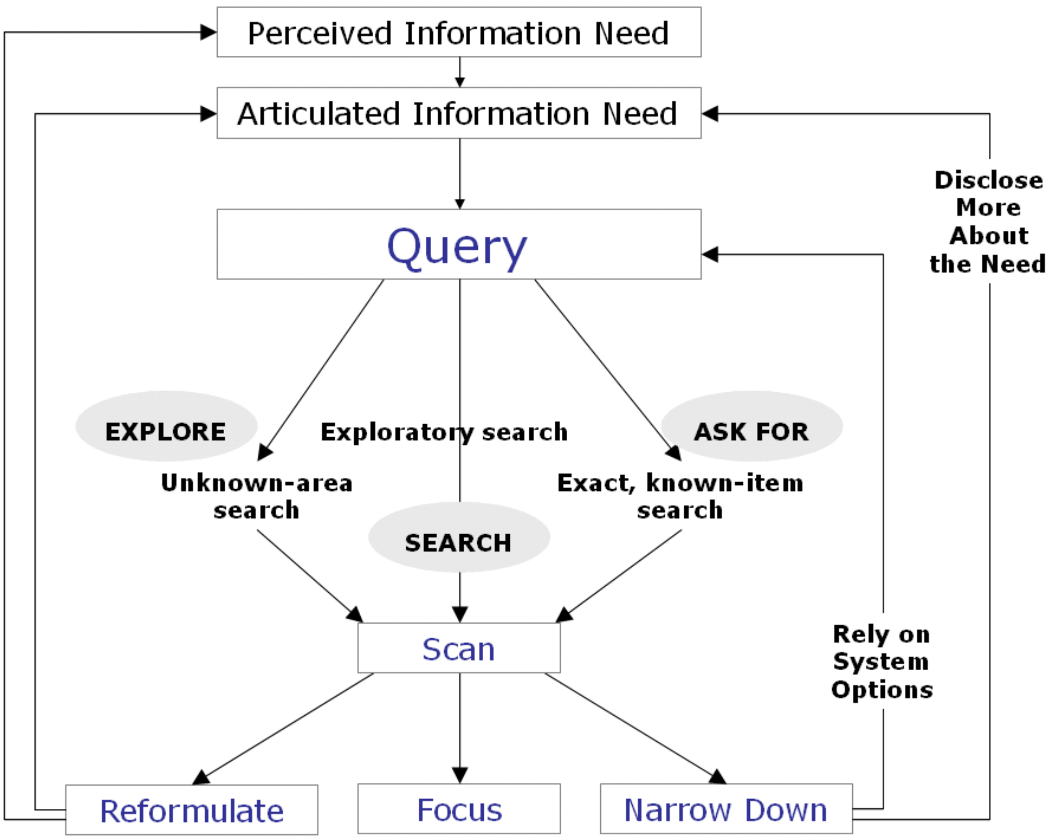


Figure 37: A model of information searching

Unlike the undirected process, searching starts with a *perceived information need*. The researcher has an idea of the material required, and in order to query the information system, the researcher converts the perceived information need to an *articulated information need*.

As discussed earlier in this chapter (5.1), a query may be regarded as representing the researcher's hypothesis about the specific keywords that are found in the required documents or attached to them as metadata. For example, a researcher may be trying to find a substance that is both a semiconductor and a superconductor. An option that crosses the researcher's mind is silicon-germanium. To check whether or not silicon-germanium matches those criteria, the researcher hypothesizes that an article that includes the words *experiment*, *silicon-germanium*, and *superconductivity* would provide the required information.

A successful conversion of the researcher's perceived information need to an articulated information need depends not only on the researcher's ability to use the appropriate terms in the query but also on the researcher's expertise in querying the information system—a knowledge of the query syntax and an awareness of other options that the system might offer, such as prefilters. From a more general perspective, one could say that the Query action in this model applies also to a natural-language question posed to a human.

The distinction between a perceived information need and an articulated information need is crucial to the information-searching process. A failure of the system to return the results that a researcher is looking for can be attributed to the nature of the need itself (e.g., the information to satisfy it may not exist), the user's failure in describing the need in a suitable way, or other factors.

The Query action can have three modes: Explore, Search, and Ask For, depending on the information needs and the way they are expressed in queries. Queries vary from being very precise—when researchers describe a specific item—to very vague, typically when researchers look for information about an unfamiliar field and thus cannot define the information need clearly. For example, in an exact search (a search for a specific article that the researcher knows about), a researcher is likely to enter information such as the article's title or the names of the authors. In this case, the researcher's mode of searching is Ask For, as opposed to Search or Explore, even though the technical process is similar for the three modes. In the Ask For mode of searching, if the document is well defined by the user, the search result is likely to be only that item, or, depending on the way the query was articulated, very few items—for example, the article itself, several reviews of it, and articles citing it. Furthermore, a good relevance-ranking algorithm would position the article as the first item on the list.

The Explore mode can be exemplified by a search for information in a field outside the researcher's area of expertise—a HEP researcher looking, for instance, for information related to the philosophy of physics, particularly information that the researcher can use to broaden the scope of an article about physical cosmology and the origin of the universe. However, lacking knowledge of the philosophy of physics, the researcher is unfamiliar with the most appropriate search terms and with the kind of materials that the query results might offer.

Most queries, however, are somewhere between these two extremes: researchers are familiar enough with their field of interest to clearly define their information needs yet do not know of a specific item that would be the ultimate result. The variations of the possible queries in this case relate to the amount of information that the researchers know about the topic or decide to provide when they convert their perceived information needs to articulated information needs. The more information they know and provide, the closer a query is to asking for a specific item. This mode of searching, which corresponds to what we intuitively perceive as searching, can be described as Search mode. The balance between formulating a query with too little of the hypothesized information—which can lead to too many results—and including too much of the hypothesized information—which can eliminate relevant results—is of great concern to most researchers and triggers a 'trial-and-error' mode of articulating the information need.

Once a query is submitted, the system displays a result list to the researcher. Even before scanning the results, the researcher obtains valuable information that the system has provided: the number of items on the list and the system's suggestions that relate to the query (such as Did You Mean…?) or relate to the result list. The latter might include, for example, post-search groupings (facets or clusters[19]) and suggestions for new searches. All this information enables the researcher to know right away if the result list is worth exploring in more depth.

[19] Facets and clusters are discussed at length in Chapter 6.

For example, a lack of results may signify that either the information does not exist, there is an error in the query, or that the articulated information need should be reconsidered. Too many results may signify that the query was too broad. In this case, a researcher who does not have a specific item in mind might miss relevant material by assuming that the sorting order placed the desired items at the top of the list. The system's Did You Mean…? suggestion may draw the researcher's attention to a misspelled name or a variation in a term. Topics and other information (date ranges, authors, types of materials, languages, publishers, journal titles, and more) that serve as post-search groupings provide a brief summary of the result list: by looking at the terms displayed in these groups, the researcher can see the major characteristics that the items on the list have in common.

Researchers typically scan the first items in a result list before taking an action. Even a brief look at these items usually provides enough clues as to whether the researcher is on the right track, especially when the list is sorted by relevance. Often the first item, or one of the first items, is the requested item—mainly when the query mode is Ask For. If none of the first items seems relevant, researchers typically re-evaluate their query.

After analysing the first screen, a researcher chooses one of the following options:

- Focus: If the result list is satisfactory and there are results that seem relevant, at least at first glance, the researcher may focus on a specific item.
- Narrow down: If there are too many results, the researcher may choose to narrow down the list so that it shows only the items that are more relevant.
- Reformulate: If there are no results or the results do not seem relevant, the researcher may decide to reformulate the query.

Narrowing down can be carried out in two ways: either the researcher takes advantage of the system's options—perhaps by clicking a facet, such as a specific date range, topic, or journal name, or by choosing to see only materials available online—or the researcher decides to modify the articulated information need by providing more information. For example, a researcher looking for aspects of the connection between direct and indirect dark-matter detection may start a search with a general query, such as *dark matter detection*, to make certain that no materials that relate to dark matter detection are ignored. However, having looked at the long result list, the researcher may decide to add terms to make the query more specific about the information need (for example, *direct AND indirect*), thus narrowing down the result list considerably. Technically, both methods—using the system's options and modifying the query to be more specific—yield another search, and the system displays a new result list.

However, when the researcher decides that the query was not adequately formulated, he or she needs to reformulate it. Reformulation can be minor or can be suggested by the system—when a name or term has been misspelled, for instance—but there could be a case where the researcher needs to consider the information need again. For example, after looking for relationships between two phenomena and receiving no documents that match a query that addresses both, the researcher is likely to conclude that there is no evidence that these two phenomena are related to each other and hence the information need must be redefined. In such cases, the researcher may modify the perceived information need and start the process again.

Once a researcher focuses on a retrieved item, the process described in the information-seeking model (Figure 33) takes place.

An example of actual search processes, as described in an e-mail message from one of the interviewees, can be seen in the context of the information-searching model, as follows:

Ask For:

> I was looking for a paper I remembered seeing. I did not remember the exact title, but I knew who the two authors were, and I remembered that the article was published last year. I entered the author names and the year of publication and received two articles that these authors published last year. The article I was looking for was the second one.

Explore:

> When I was a member of the committee on appointments and promotions and had to learn about a candidate, I often tried to understand the “world map" of fields of research about which I know next to nothing (e.g., life science research). I looked in Google for the name of the candidate or the name of someone that wrote a recommendation for this candidate, and from there I continued by clicking various sites that looked relevant.
>
> I also often visited the site of a conference in which the candidate gave a presentation and looked at the topics, the names of the speakers, etc, and went on (via Google) with that.

Search:

> Next week I need to give a lecture to a large audience. The topic is lepton flavour violation—not my main field of research—so I am not very knowledgeable about the literature in this field. I went to SPIRES and looked for reviews, entering *lepton AND flavour AND violation* as query terms, and I received about 180 results. The results were sorted by date, and because I was interested in the newer reviews, I quickly scanned the list from the top, easily identifying the articles that would be relevant for my talk.

As explained earlier, the model of information seeking and, specifically, information searching does not depict a distinct beginning and end, as opposed to models such as that of Ellis (1989). Although at times researchers have very specific information needs that can be satisfied and thus the information-seeking process may seem complete, most information needs are ongoing and last for long periods of time. Researchers are always exposed to new information through passive channels, which trigger further active searching for related materials.

5.6 Testing the Model via the Personas

Typical information-seeking behaviour of the six personas can be described using the model. The descriptions provided in this section illustrate both directed information-seeking activities—namely, searching and navigating—and undirected information-seeking activities, particularly ways in which scientists keep abreast of new developments in their field. Seven user scenarios were defined and demonstrated by means of the information-seeking model and its subcomponent, the information-searching model.

The colour scheme used to illustrate the information-seeking and information-searching processes is described in Figure 38.

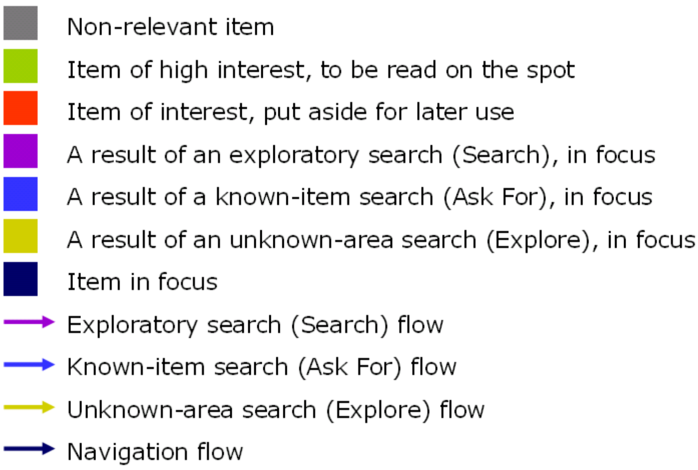


Figure 38: Legend for figures 39 through 62

5.6.1 Scenario 1

Step I (Figure 39): Ed is scanning the new submissions in arXiv (1). He sees an article that may be of relevance to research that he had in mind (2) and reads the abstract (3). Ed thinks that the article is worth reading (4); however, he does not have time to read it on the spot. He downloads the file and saves it in his Future Readings directory (5).

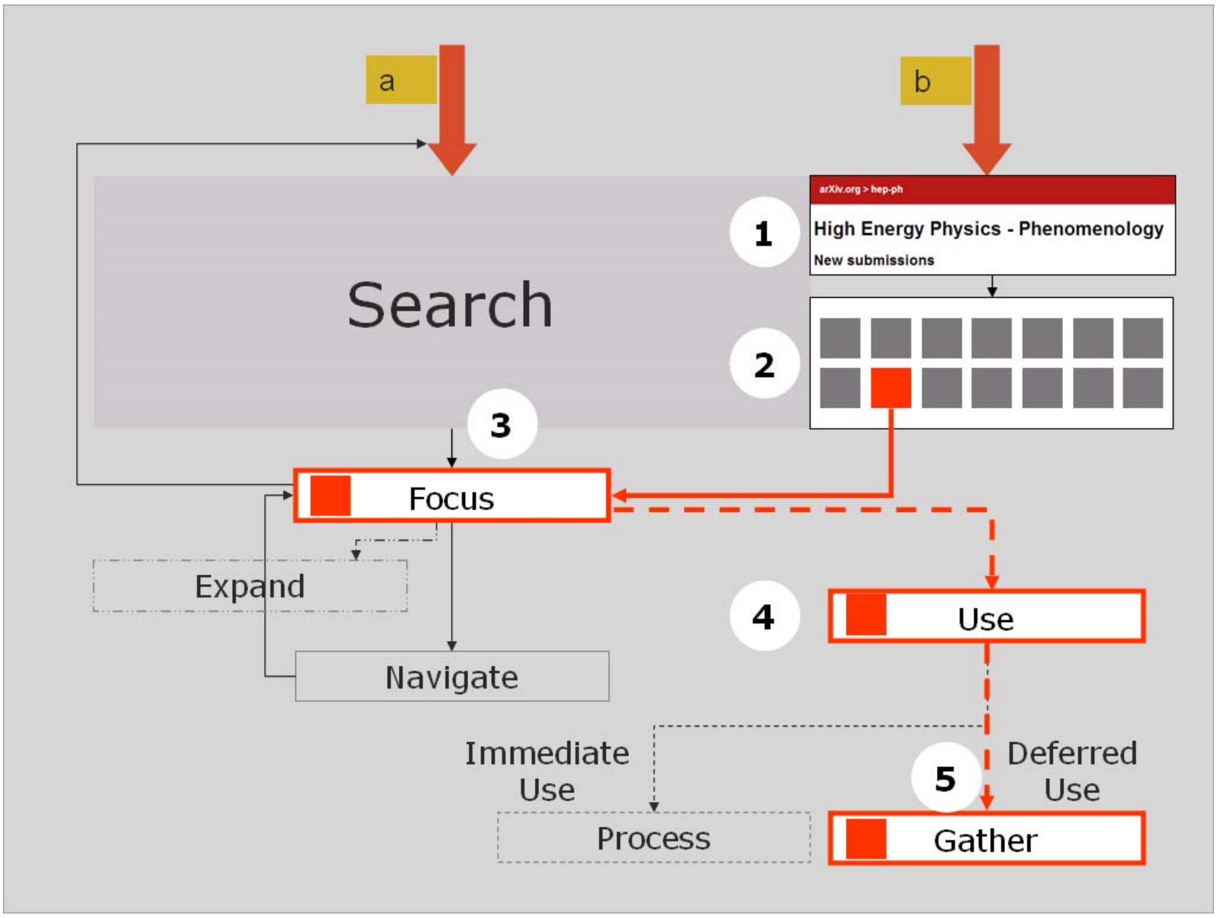


Figure 39: The information-seeking model applied to scenario 1, step I

Step II (Figure 40): Ed continues to browse through the arXiv list of new submissions (2) and is drawn to an article that is related to his current research (3). He reads the abstract and decides to read the whole article (4, 5). After reading the article, which is written by a group he knows well, he remembers that last year, one member of that team published an article on a related subject. Ed wants to search for that article. (6)

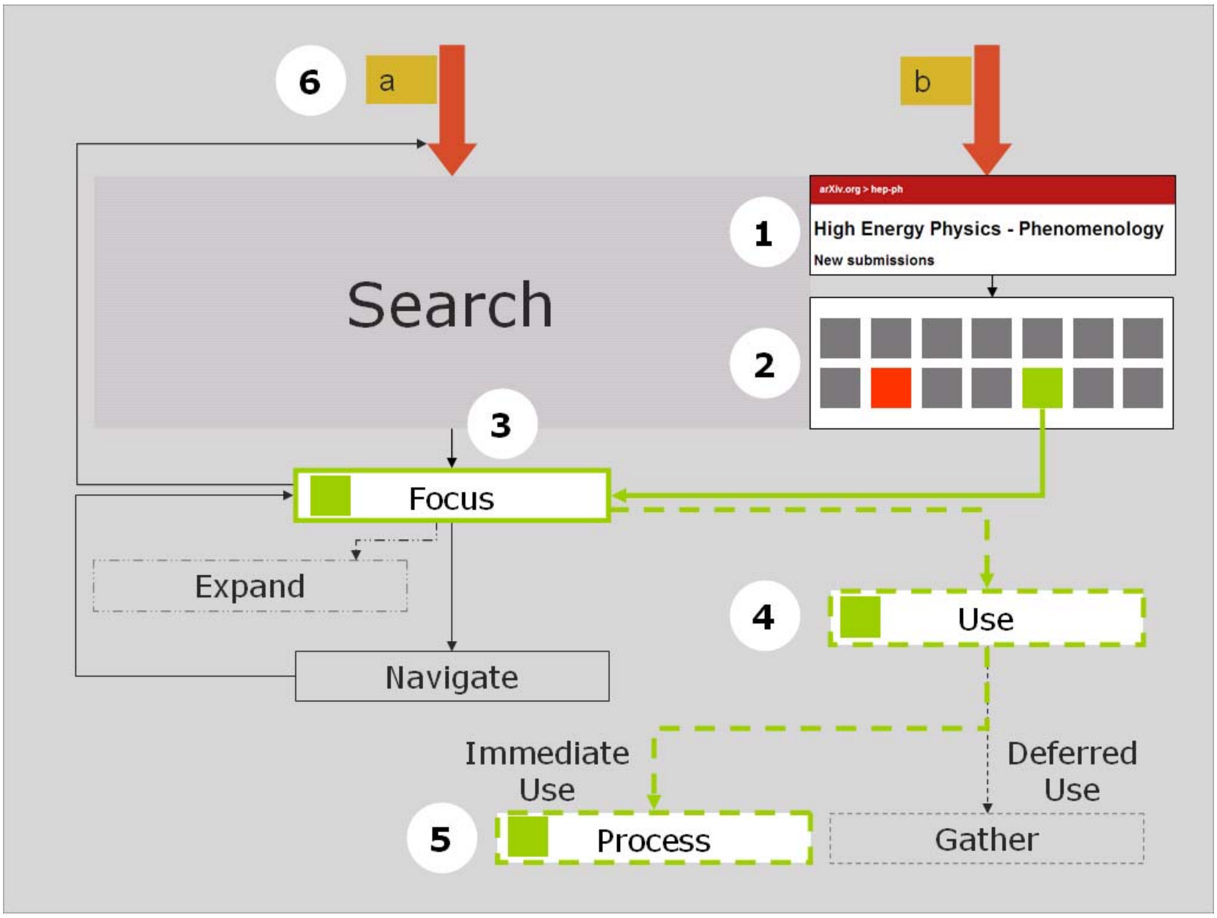


Figure 40: The information-seeking model applied to scenario 1, step II

Step III (Figure 41): Ed now turns to SPIRES (1), where he searches for the article by the person whom he remembers from the group that authored the article he just read (2). The system displays four articles (3); one of them is the article that Ed had in mind, and Ed focuses on it (4).

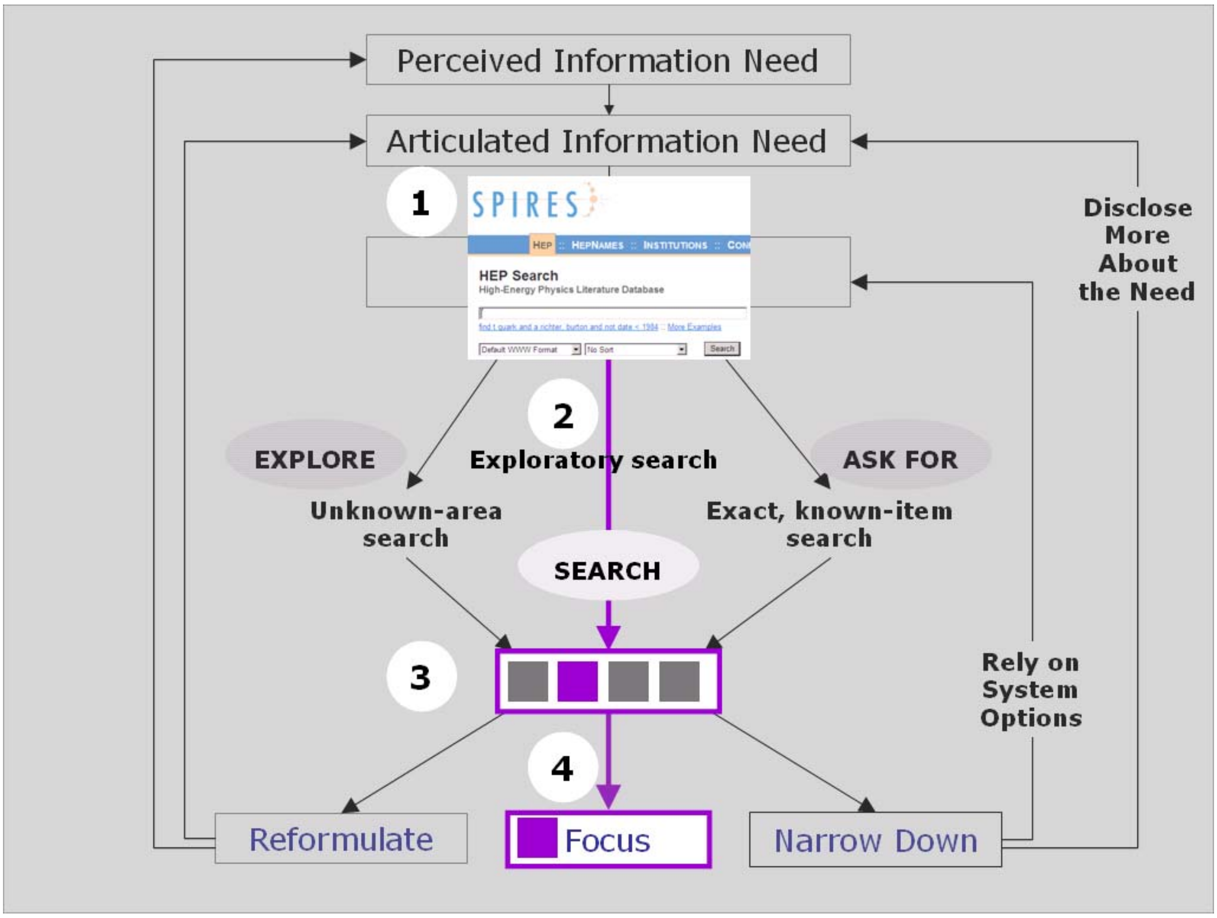


Figure 41: The information-searching model applied to scenario 1, step III

5.6.2 Scenario 2

Step I (Figure 42): One of the PhD candidates that Ed supervises asks for Ed's advice regarding a specific idea that she has—showing a relationship between two phenomena that, at the outset, do not seem to be connected. Ed is not sure that the student is indeed on the right track but thinks that checking whether any such idea has been previously explored would be worthwhile. Together with the student, he uses SPIRES (1) to search for articles that discuss both phenomena (2). The system does not find any results. (3) Ed suggests a variation to the query, making it more general (4).

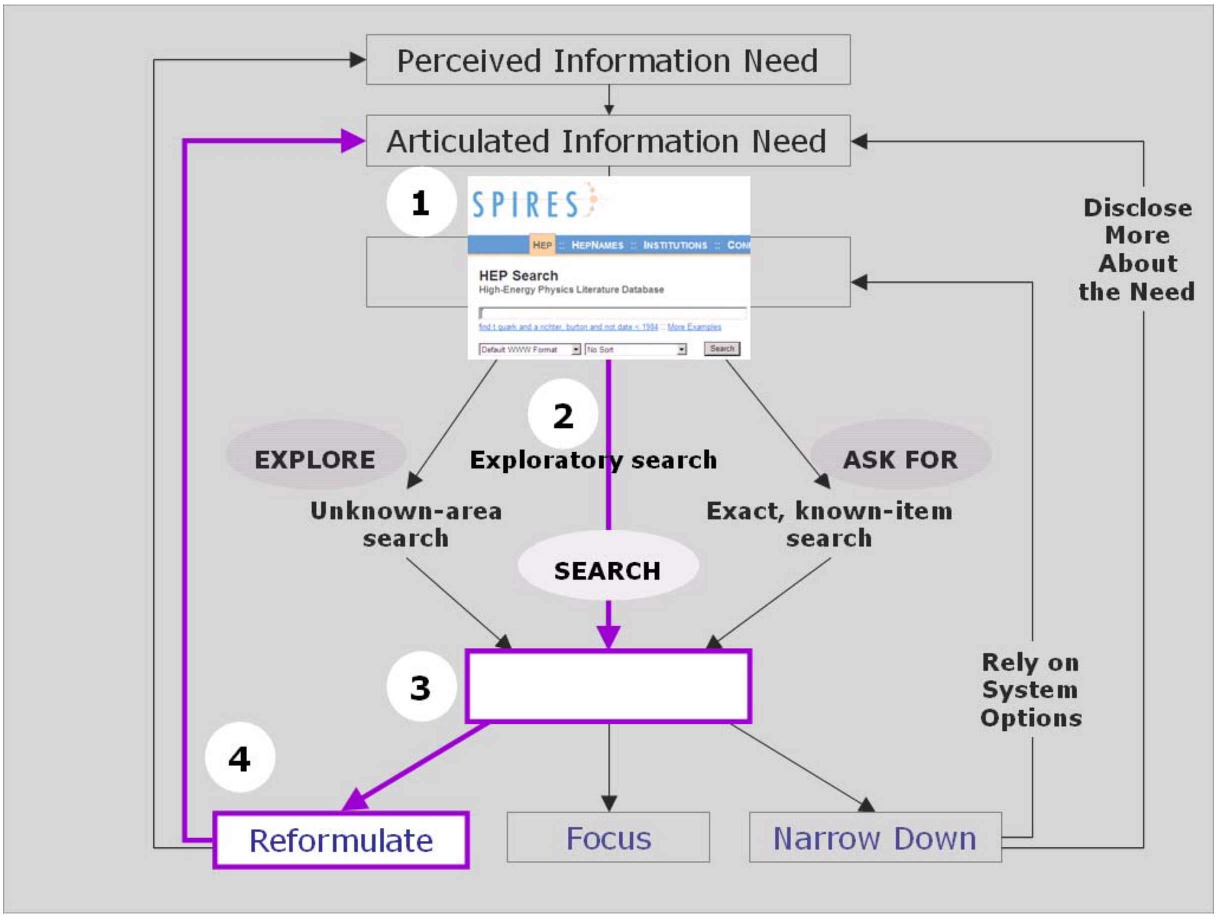


Figure 42: The information-searching model applied to scenario 2, step I

Step II (Figure 43): With the revised query, several hundred articles show up (1). Scanning the list, Ed realizes that another term added to the query may narrow down the list and help the student focus on the items that deal with the subject that she is exploring. Ed modifies the query by adding that term (2) and launches another search (3).

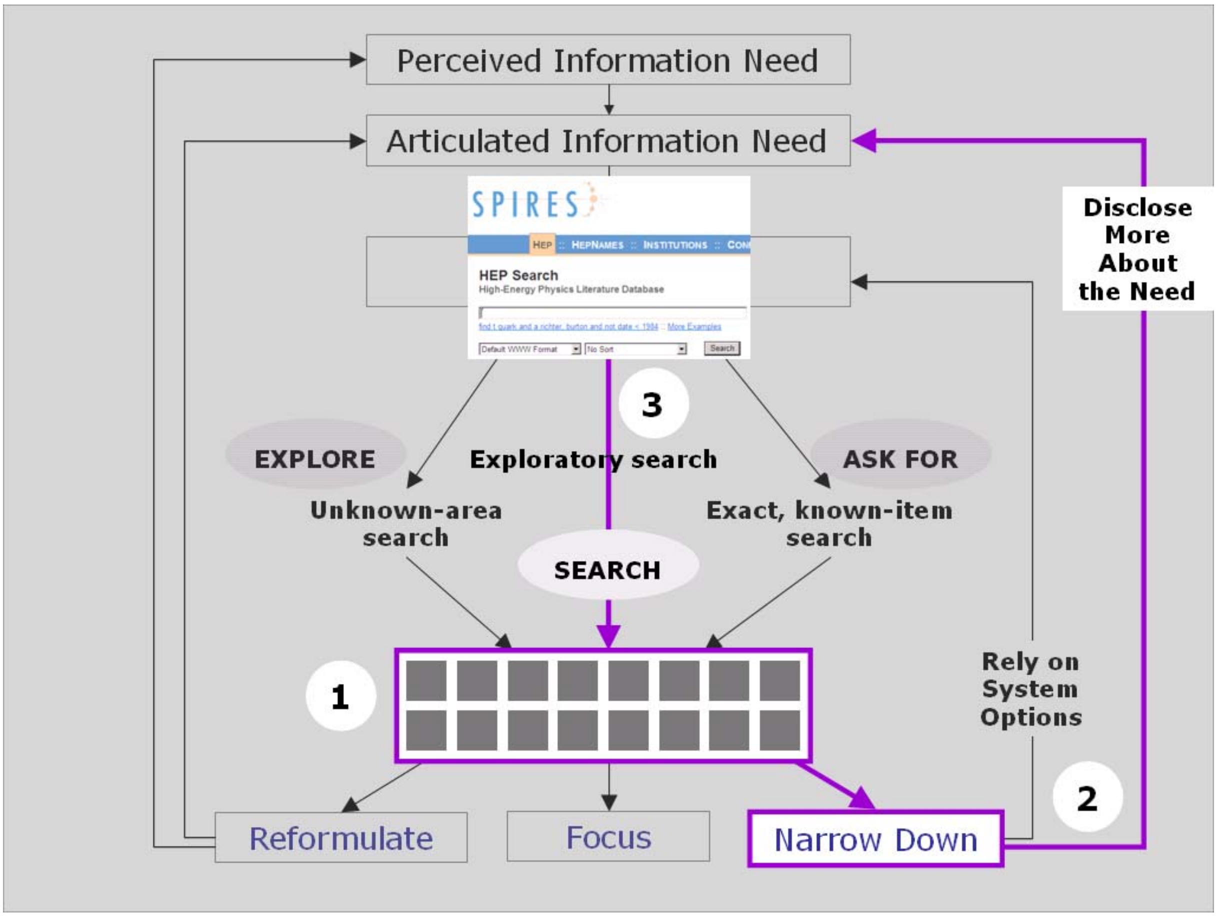


Figure 43: The information-searching model applied to scenario 2, step II

Step III (Figure 44): Having added the new term to the query (1), Ed is now presented with several dozen results (2), most of which look relevant at first glance. The student can now examine these results (3) one by one and see whether other researchers have investigated the relationship between the two phenomena and what conclusions were reached, if any.

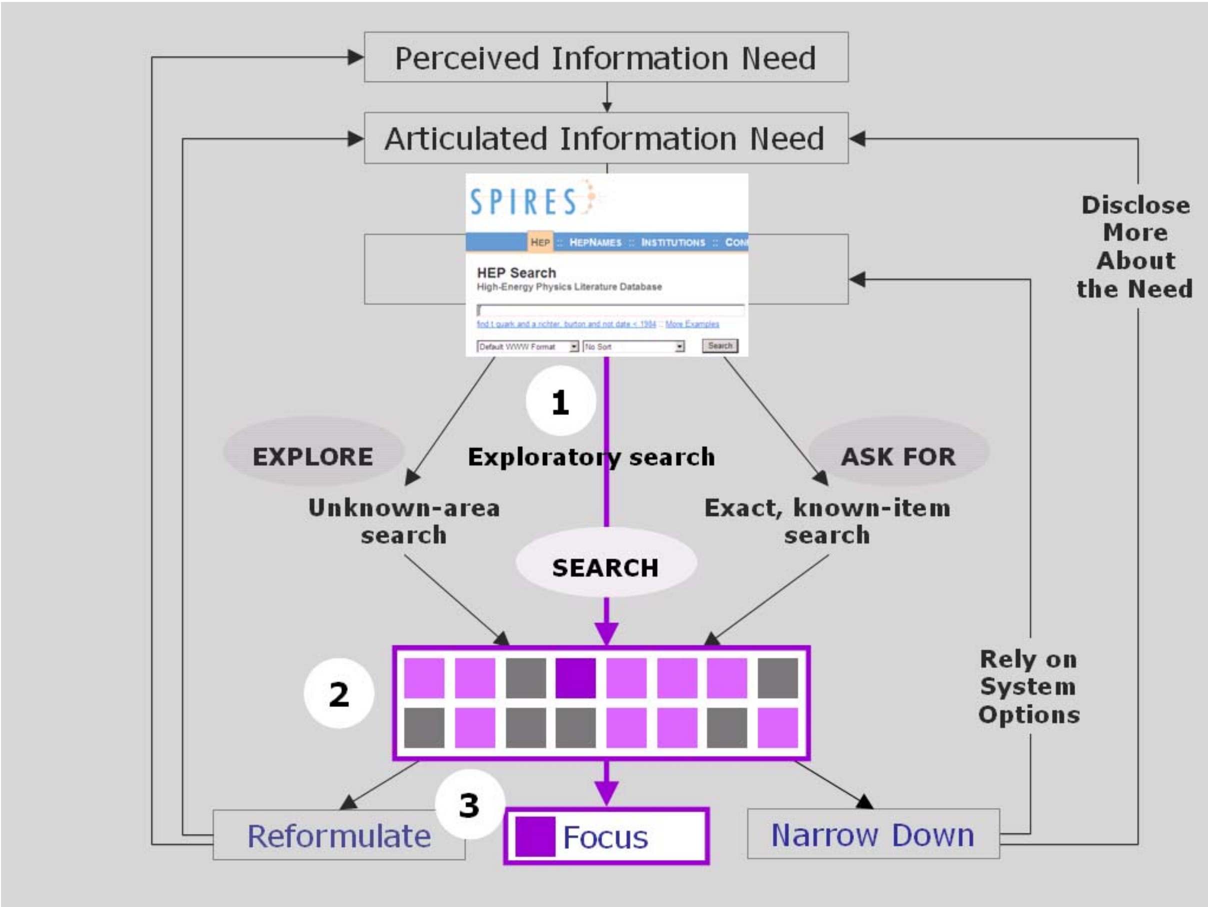


Figure 44: The information-searching model applied to scenario 2, step III

5.6.3 Scenario 3

Step I (Figure 45): Discussing a new direction for Laura's research, her supervisor mentioned a good article that Laura should read. The supervisor sent Laura an e-mail message with the name of the author and the title of the article. For this type of search, Laura prefers Google, so she pastes the article title into the Google search box (1) and launches a search (2). Google displays a short list (3), in which the specified article is the first item on the list (the list also includes the author's version of the article on the author's home page and other articles that refer to the specified article). By clicking the article title, Laura accesses the article in arXiv (4), where she can obtain the article's full text.

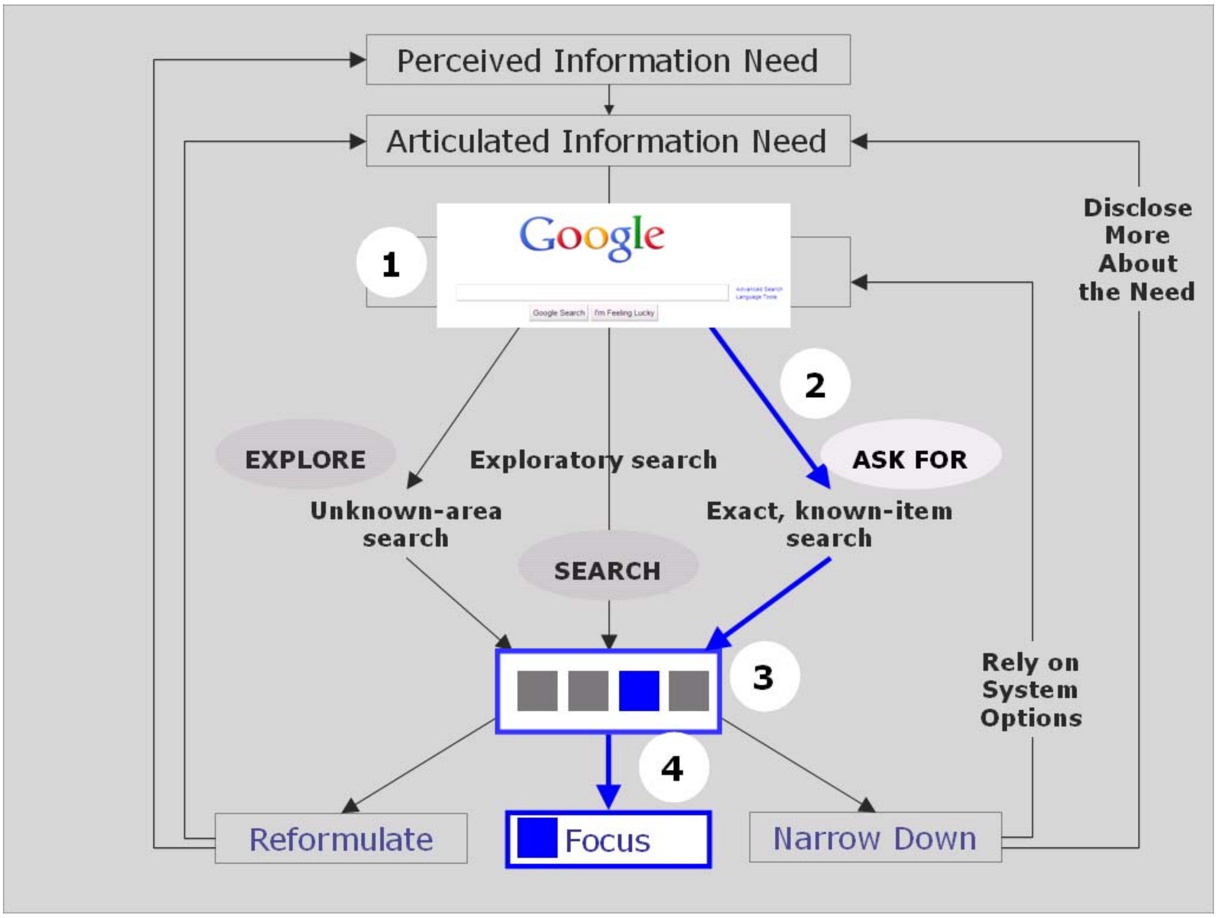


Figure 45: The information-searching model applied to scenario 3, step I

Step II (Figure 46): Laura downloads the PDF version of the article and reads it. She realizes that she should read at least one of the articles cited by the article that she just read so that she will fully understand the theory under discussion. Because there are no hypertext links to the references, she decides to turn to SPIRES rather than search for the references one by one in Google. Laura searches for the article recommended by her supervisor in SPIRES (1) (2) and, there, too, receives a short result list (3). She focuses on the article (4).

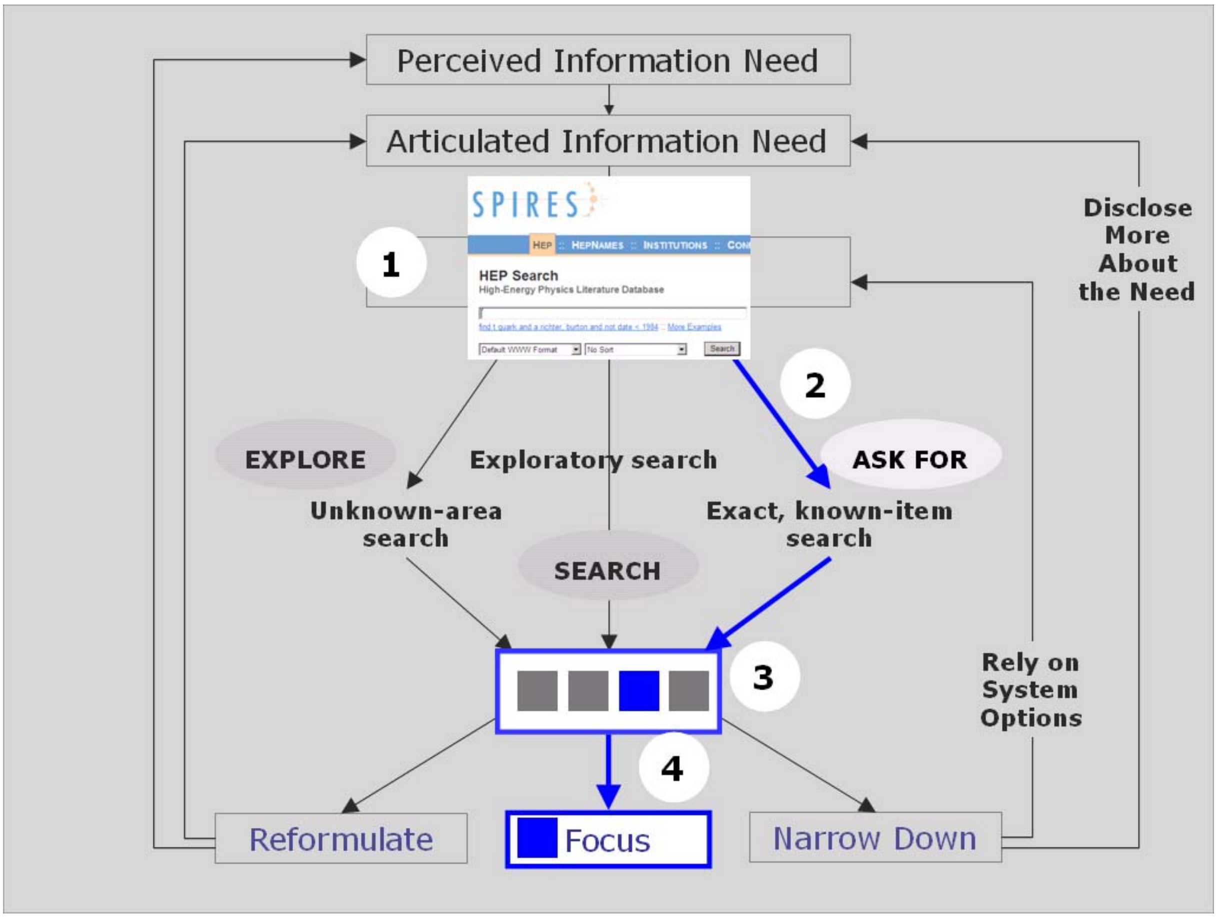


Figure 46: The information-searching model applied to scenario 3, step II

Step III (Figure 47): After focusing on the article recommended by her supervisor (1), Laura uses the References option in SPIRES, which provides hypertext references, to navigate (2) to one of the cited articles (3).

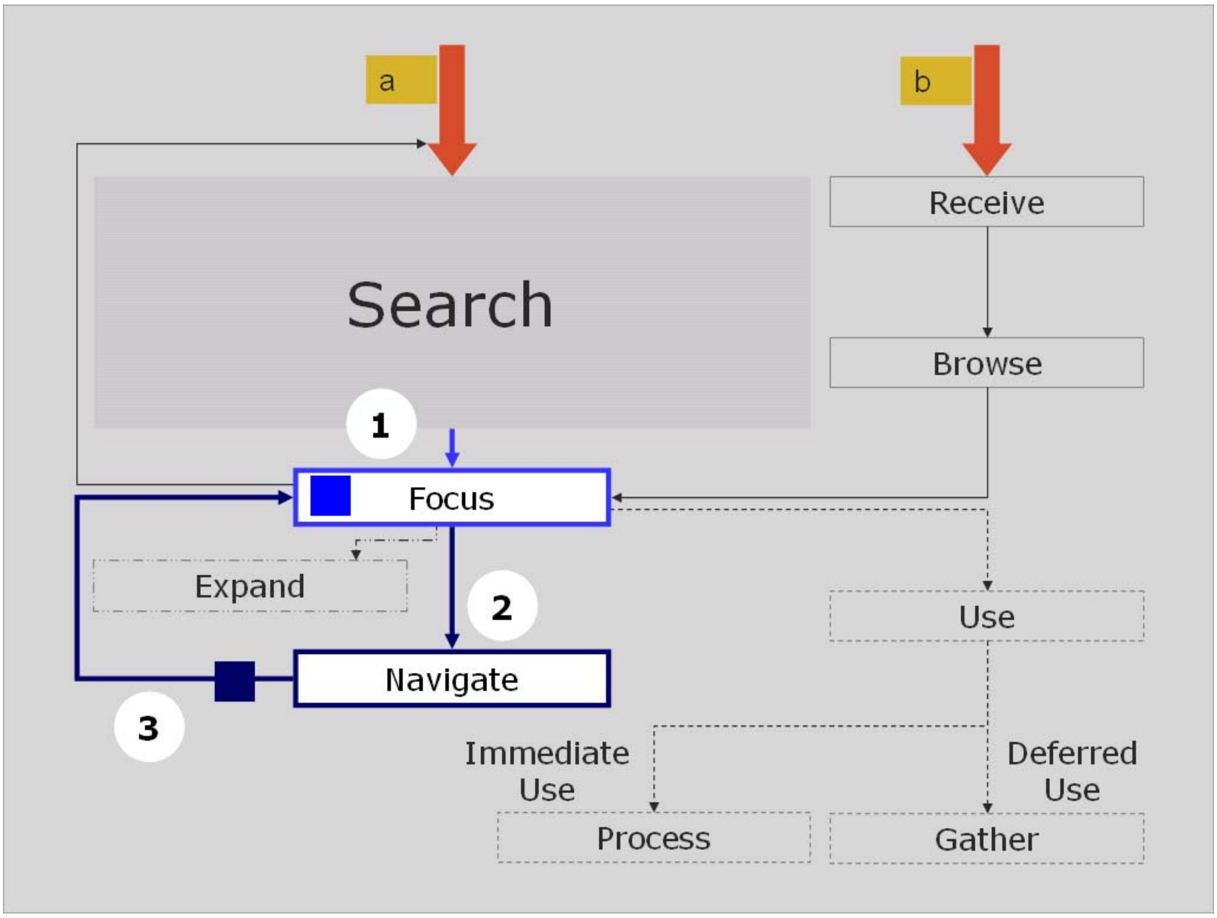


Figure 47: The information-seeking model applied to scenario 3, step III

Step IV (Figure 48): Once the cited article is in focus (1), Laura reads it and decides to follow the path that the theory presented in this article took as it became accepted by other researchers. This time she uses the Citations option in SPIRES and navigates (2) to a third article (3).

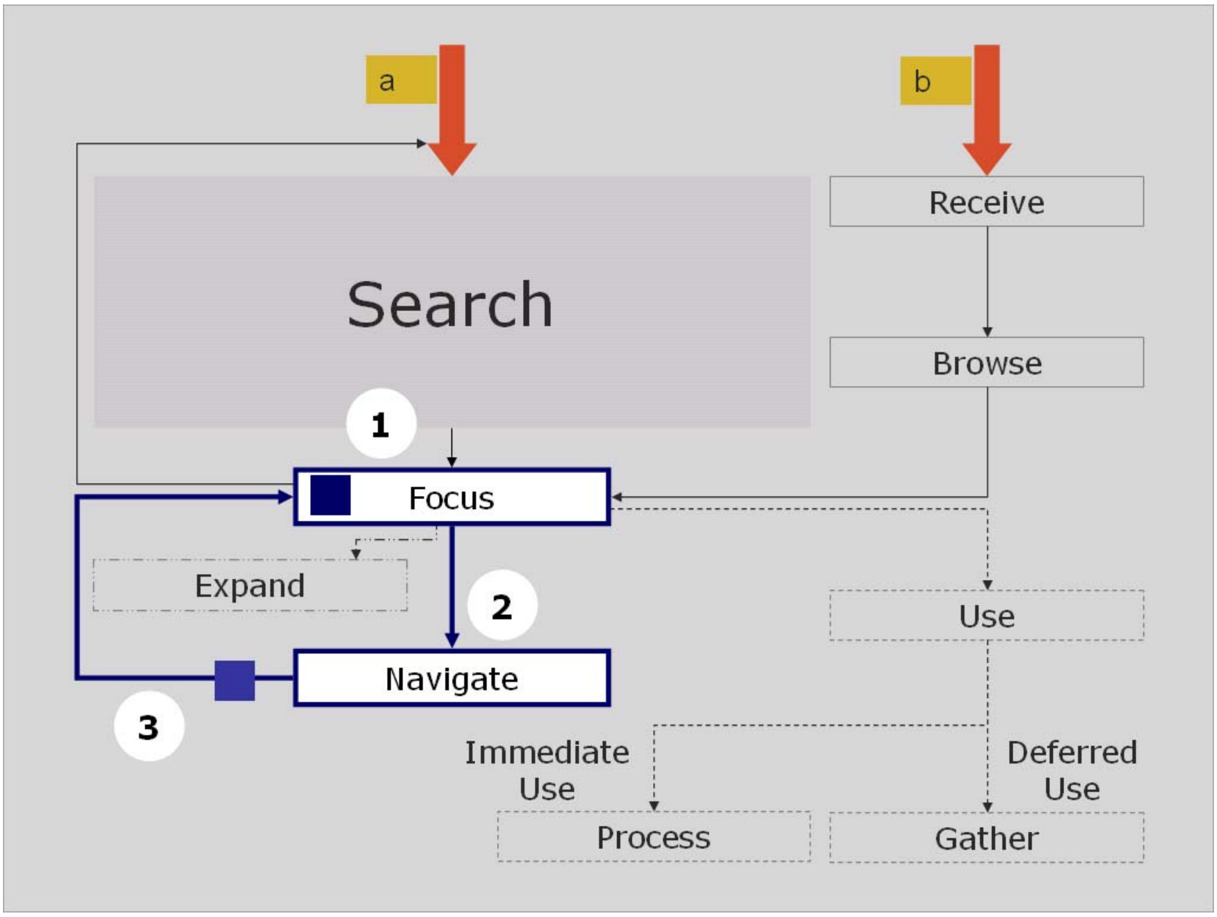


Figure 48: The information-seeking model applied to scenario 3, step IV

5.6.4 Scenario 4

Kevin is asked by an ex-colleague to comment on a new development in the field. The new development is discussed in a paper that was submitted to arXiv a couple of weeks earlier, which Kevin noticed but did not read; the arXiv ID is noted in the ex-colleague's e-mail message.

Step I (Figure 49): Kevin uses the arXiv ID to search for the article in arXiv (1) (2)—obtaining exactly one search result (3)—and reads the article (4). He knows one of the authors and remembers that that author was involved in another project that on the face of it seems to be dealing with a theory similar to the one presented in the article. Kevin does not remember seeing the published results of the earlier project, but he may have missed them. The way to check whether there were such results, he thinks, is to look at the publications of that author in the last five years or so.

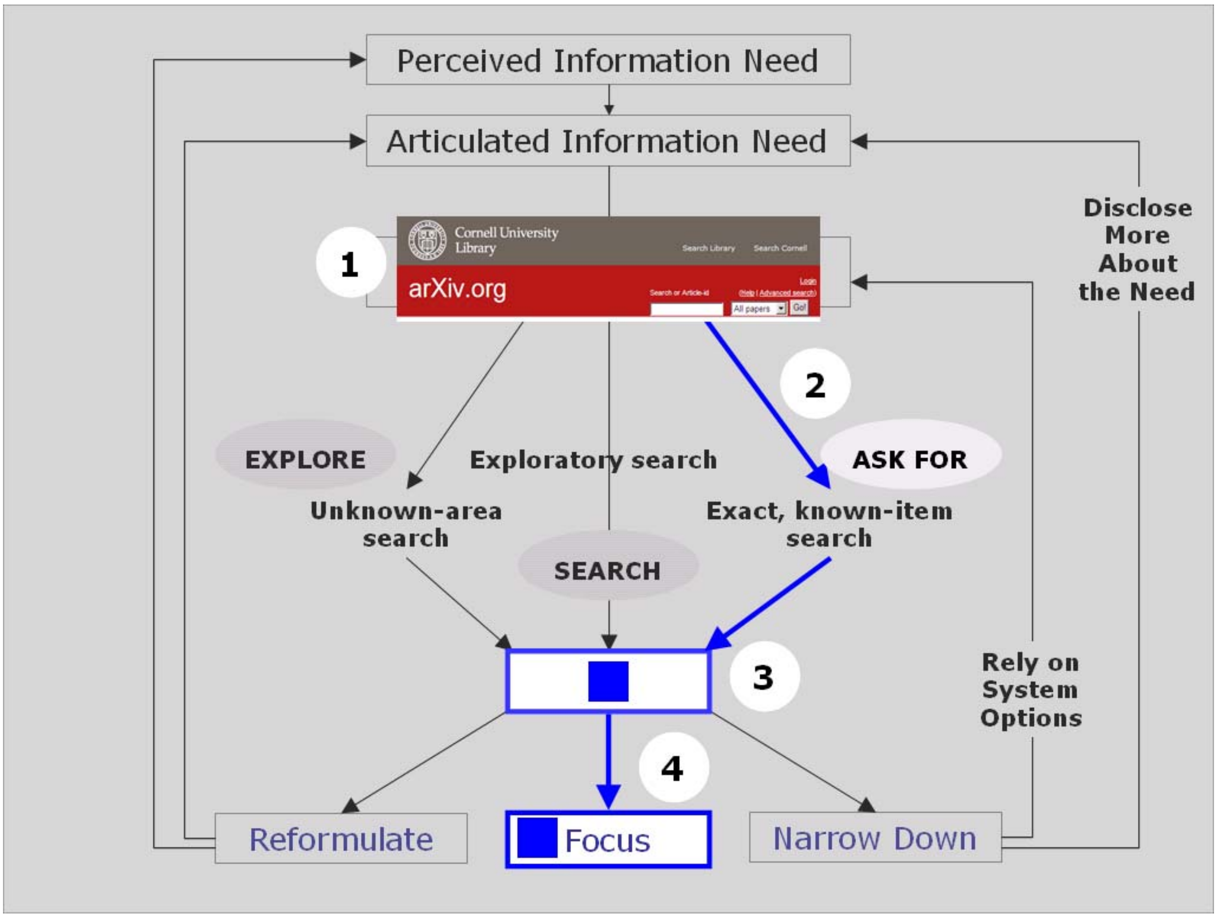


Figure 49: The information-searching model applied to scenario 4, step I

Step II (Figure 50): Kevin launches a new search in arXiv by clicking the name of the author whose project results he is seeking (1).

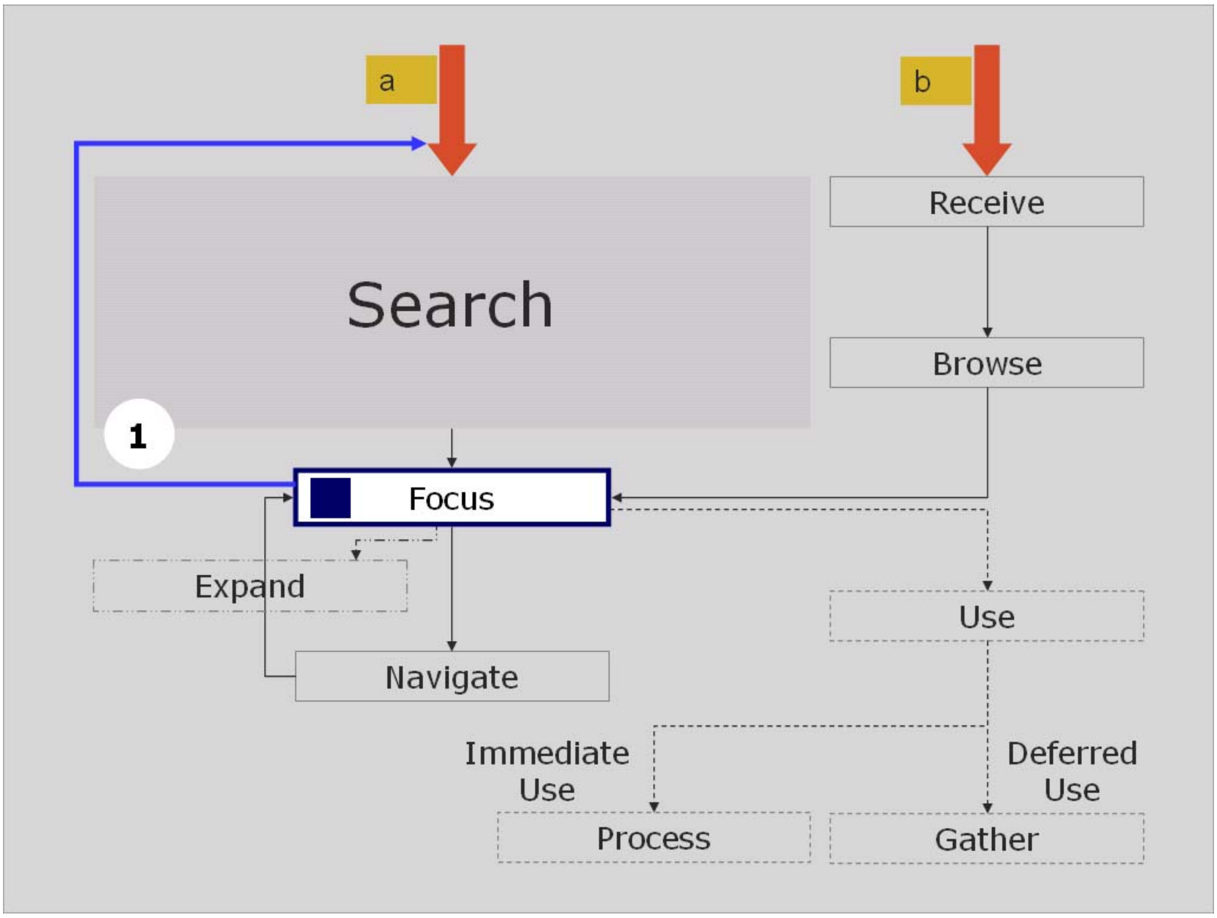


Figure 50: The information-seeking model applied to scenario 4, step II

Step III (Figure 51): arXiv searches for the publications of the author whose work Kevin is seeking (1) and displays the list (2). Kevin scans the list but does not find any article on the topic that he remembered.

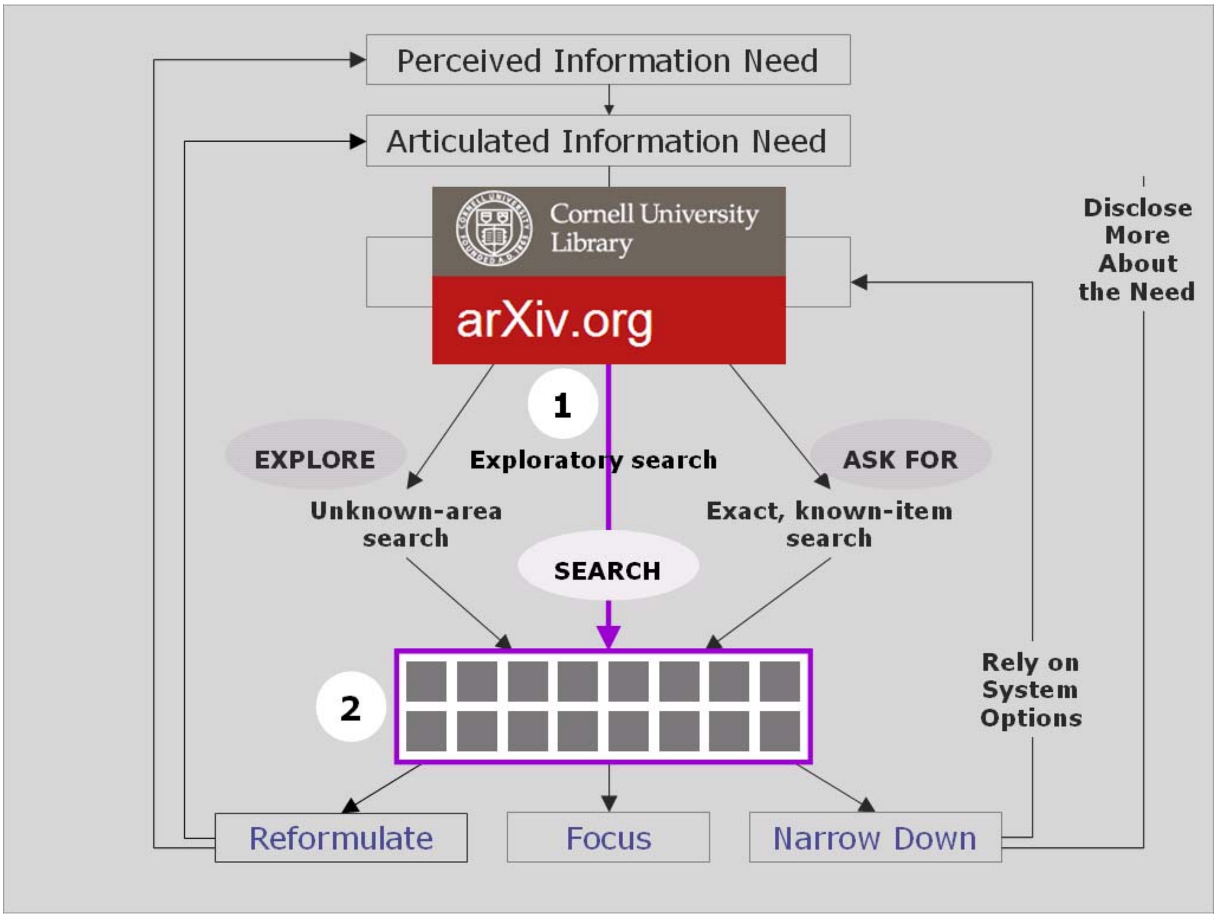


Figure 51: The information-searching model applied to scenario 4, step III

Step IV (Figure 52): Kevin thinks further and decides to go back to a theory that was rather revolutionary at the end of the 1980s and that may shed some light on the topic that he is pursuing. He knows that articles from that period will not be in arXiv, so he turns to SPIRES (1) and searches for materials on the theory (2). SPIRES displays too many results for him to look at (3), although many seem relevant. Kevin decides to narrow down the list and uses the system's option to limit the results to those that were cited the most (4). In this way, he hopes to get to the article that first presented the theory.

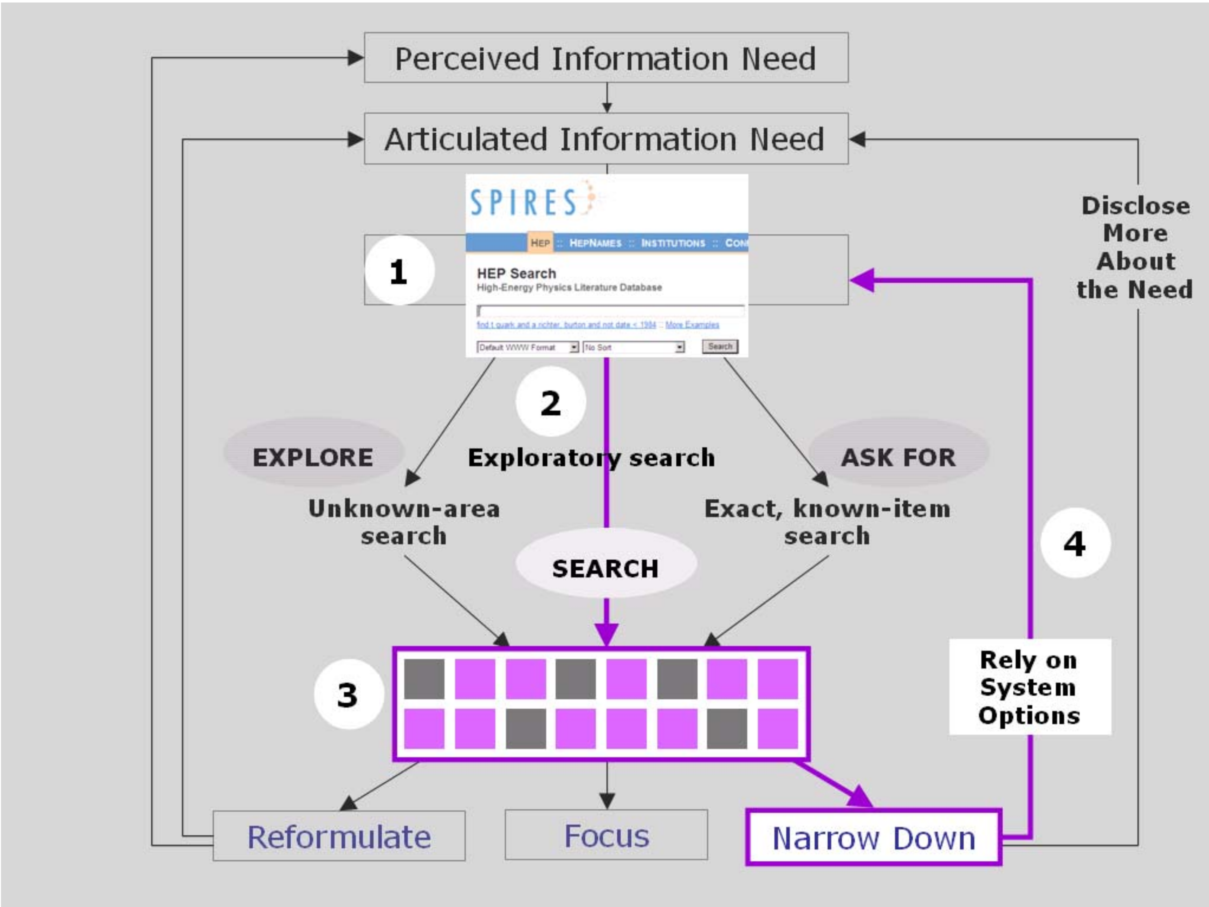


Figure 52: The information-searching model applied to scenario 4, step IV

Step V (Figure 53): The new SPIRES search (1) displays a new result list (2). Indeed, the article that Kevin is looking for is one of the first on the list. Kevin reads the article (3).

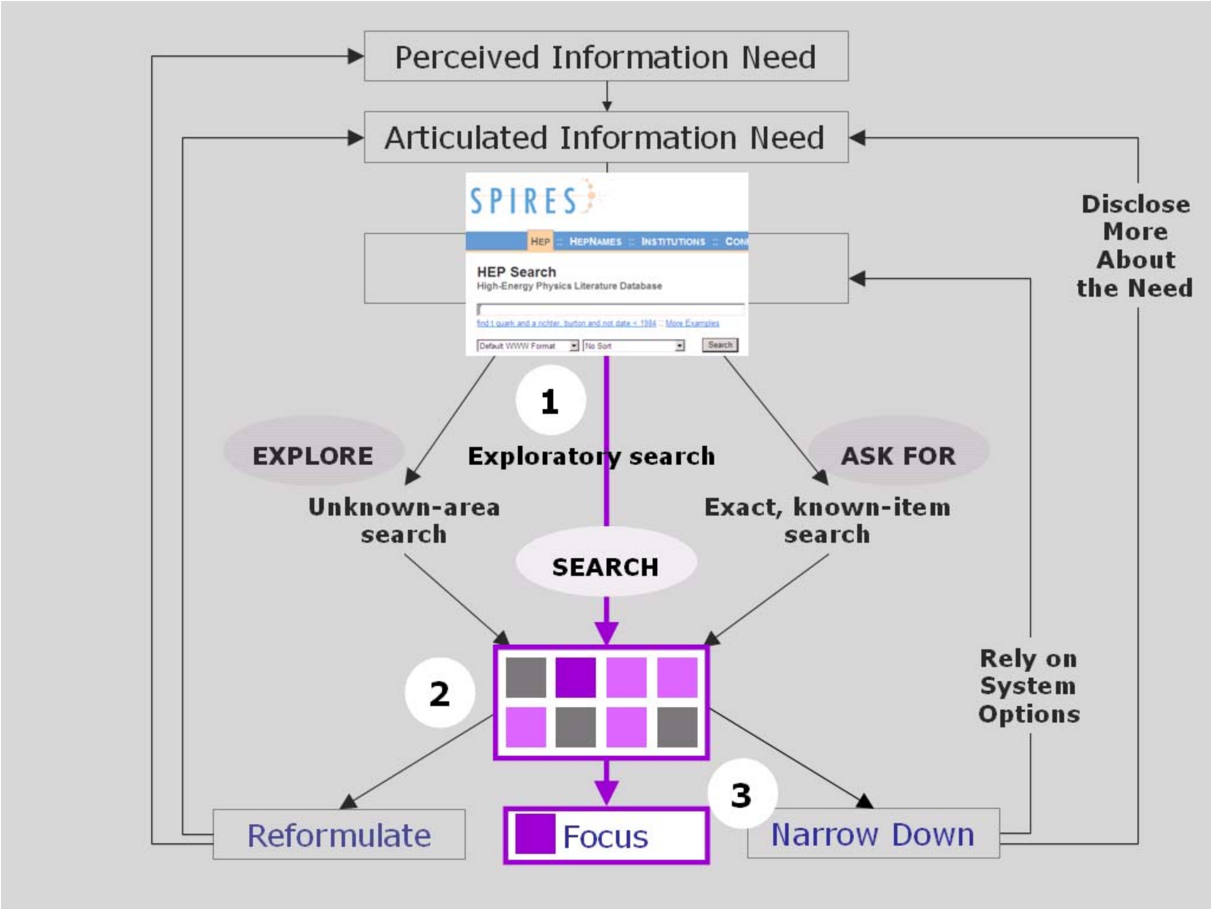


Figure 53: The information-searching model applied to scenario 4, step V

Step VI (Figure 54): After reading the article (1), Kevin decides to see how the idea expressed in the article influenced later research and navigates (2) to one of the articles citing this one (3). Next, Kevin moves on to an even later article (1, 2, 3) and then feels that he has enough background to evaluate the new idea that he was asked about.

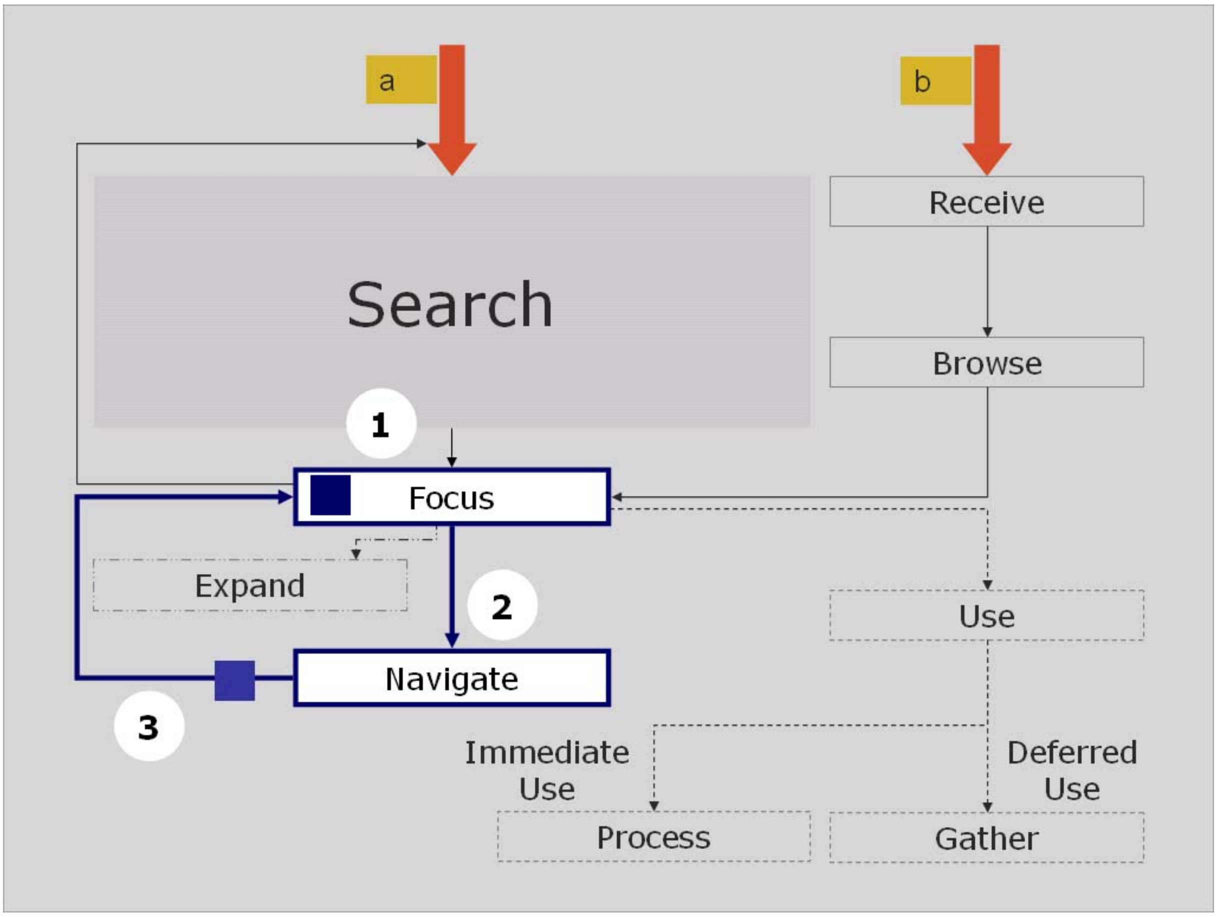


Figure 54: The information-seeking model applied to scenario 4, step VI

5.6.5 Scenario 5

Guy has been asked to give a brief talk on a topic that is not in his direct area of research, so he tries to locate some recent information about the topic. Guy launches a search in Google, entering seven words that relate to the topic (1) (Figure 55). Sure enough, although Google returns many results (2), Guy can spot four items on the first result page that are useful—one presentation, two articles, and a video clip that will help him prepare his talk (3). He does not need more than that.

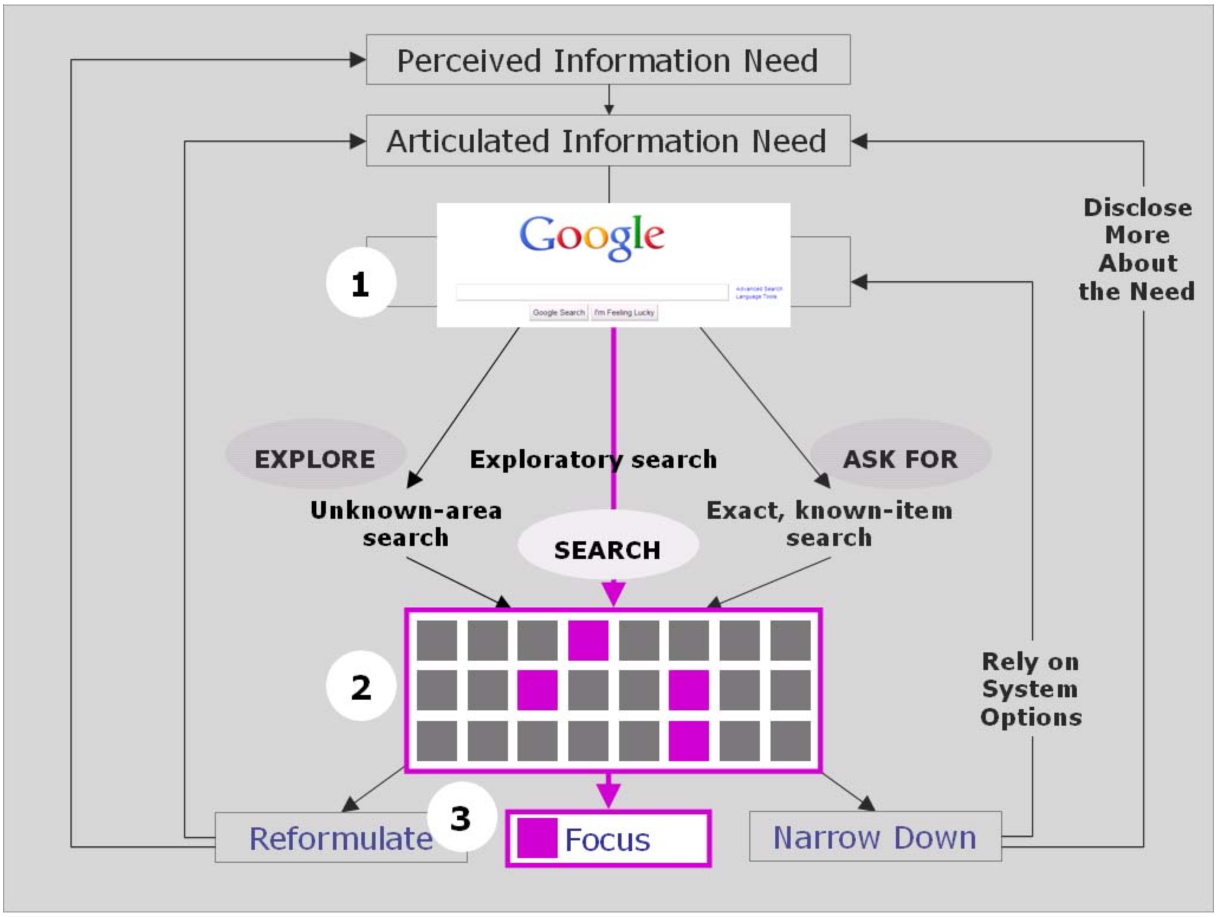


Figure 55: The information-searching model applied to scenario 5

5.6.6 Scenario 6

Step I (Figure 56): Hiro is reading an article that was sent to him by e-mail (1); this article is extremely relevant to a topic that is he exploring at the moment. However, the article's authors do not explain some of the computations that they are relying on. These computations are similar to some that Hiro is dealing with, so he wants to learn more about the ones in the article. Hiro notices that the article refers to a dataset, which he decides to look at (2). The dataset is interesting but does not provide Hiro with enough tools to advance his understanding of the computations.

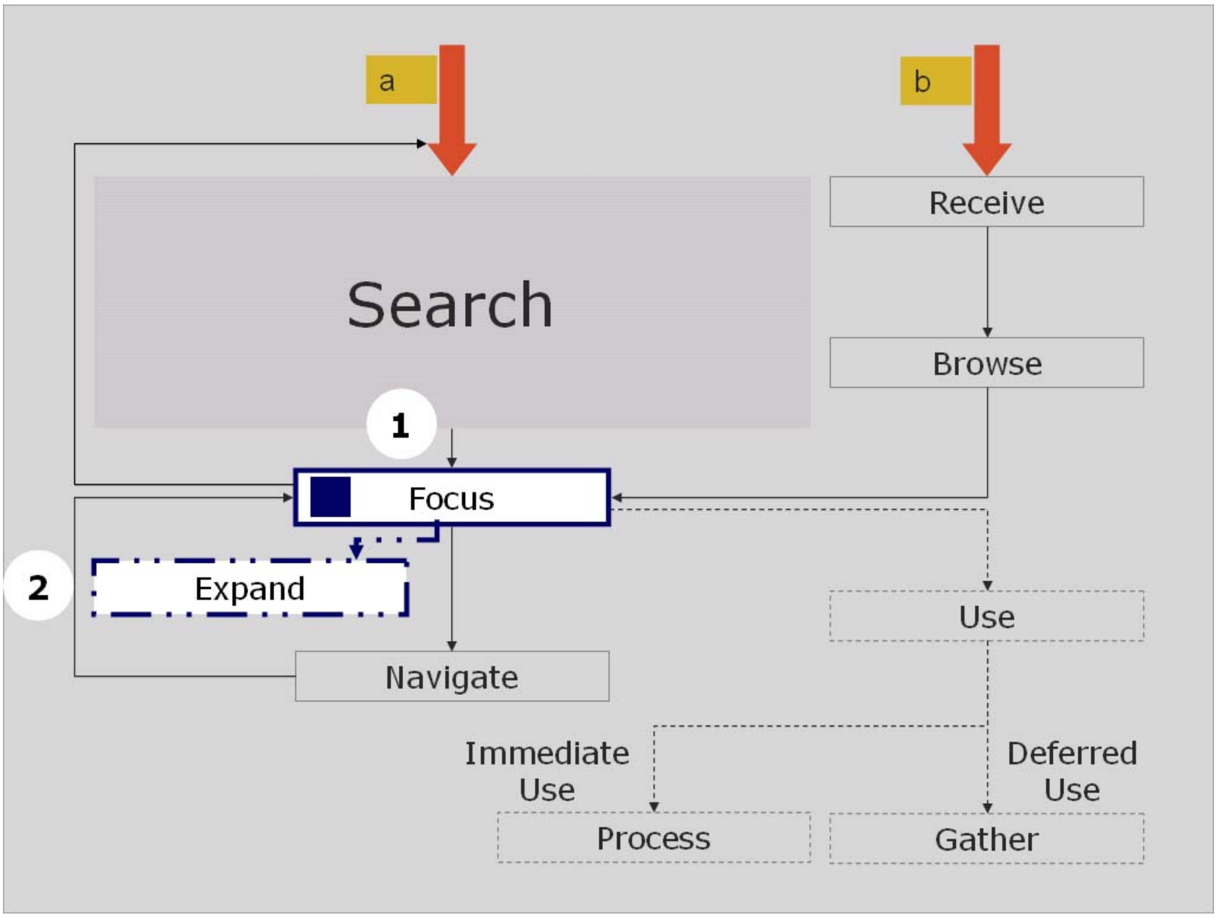


Figure 56: The information-seeking model applied to scenario 6, step I

Step II (Figure 57): Hiro looks in Google for more information about the specific computations mentioned in the article that he read (1). As he expects, he needs to make several attempts (1, 2, 3, 4) in phrasing the query.

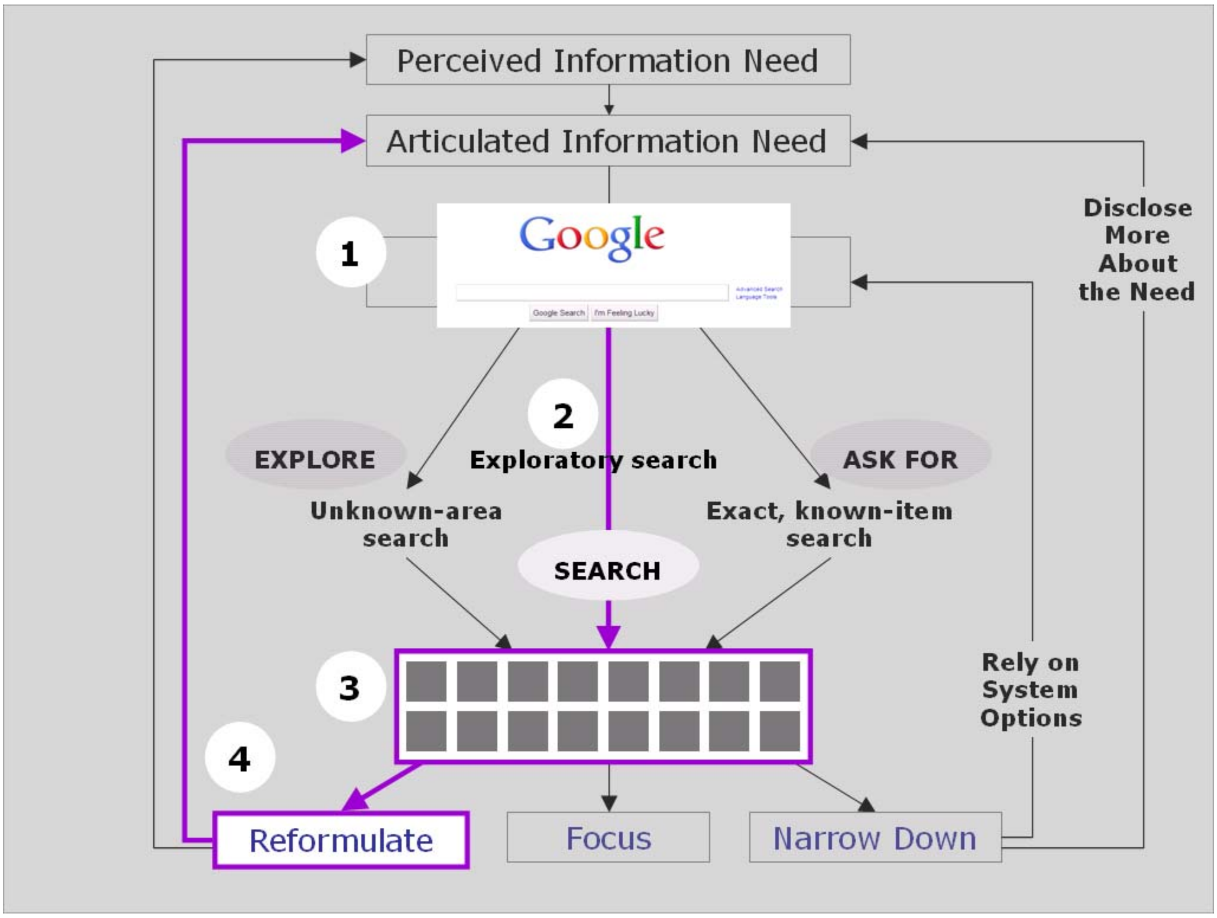


Figure 57: The information-searching model applied to scenario 6, step II

Step III (Figure 58): Hiro finally enters the query that brings up the exact information that he needs (1); the computations are in a thesis available on an institutional Web site (2).

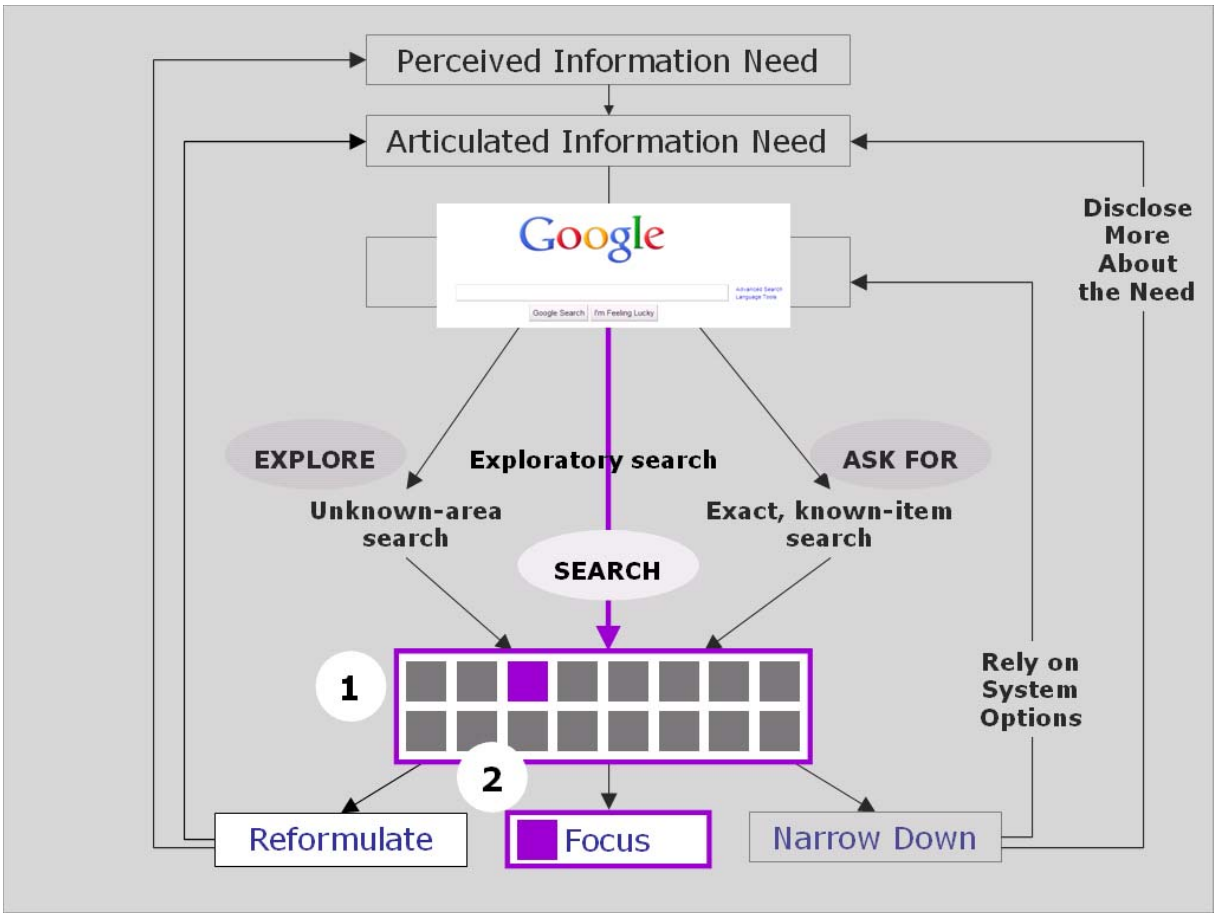


Figure 58: The information-searching model applied to scenario 6, step III

5.6.7 Scenario 7

Step I (Figure 59): George is looking for a solution to a technical problem that is bothering him. He has examined several instruments that he was considering using for a project, but they are not suitable. George decides to use Google to search on the Internet for the ‘ideal’ instrument (1). He phrases his query by listing the attributes of the desired instrument (2), but the results are disappointing—all are irrelevant (3). After realizing that there is no such instrument, he thinks of another approach and looks for something else (4).

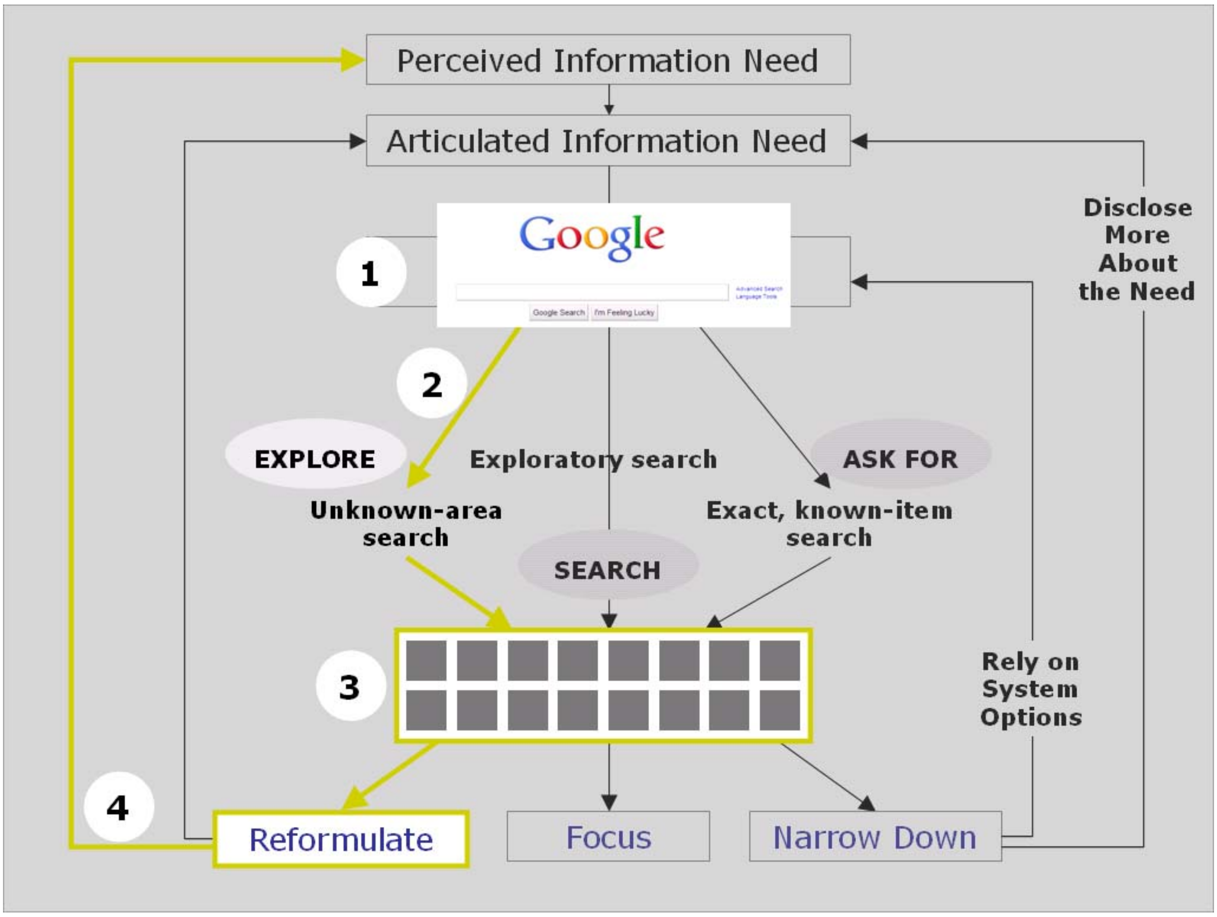


Figure 59: The information-searching model applies to scenario 7, step I

Step II (Figure 60): George uses the Google search box again (1) and looks for a different set of attributes (2). The new results are still not relevant (3), but one of the results hints at a possible direction. Re-evaluating his need once again, George submits a new query (4).

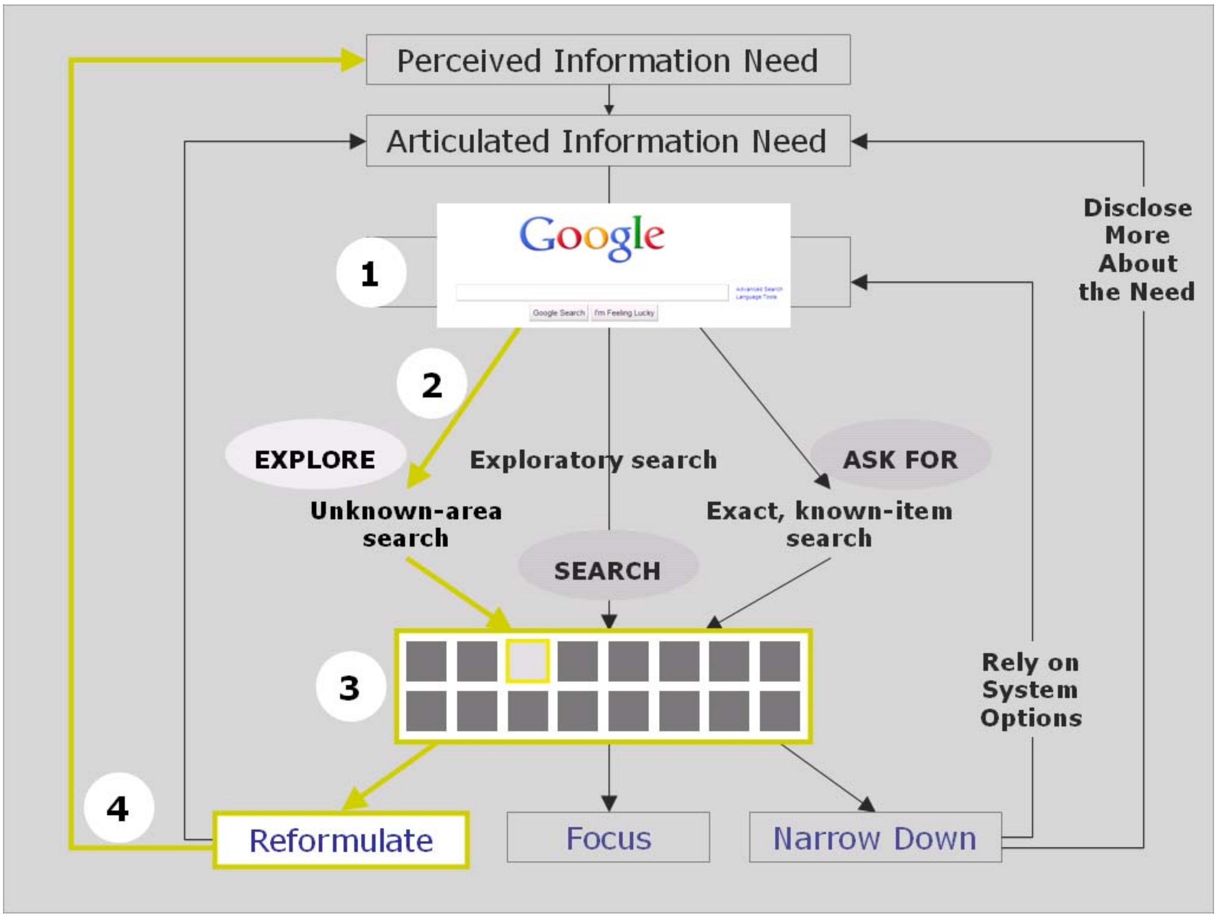


Figure 60: The information-searching model applied to scenario 7, step II

Step III (Figure 61): Google displays many results, including information about two instruments that may be relevant to George's project (1).

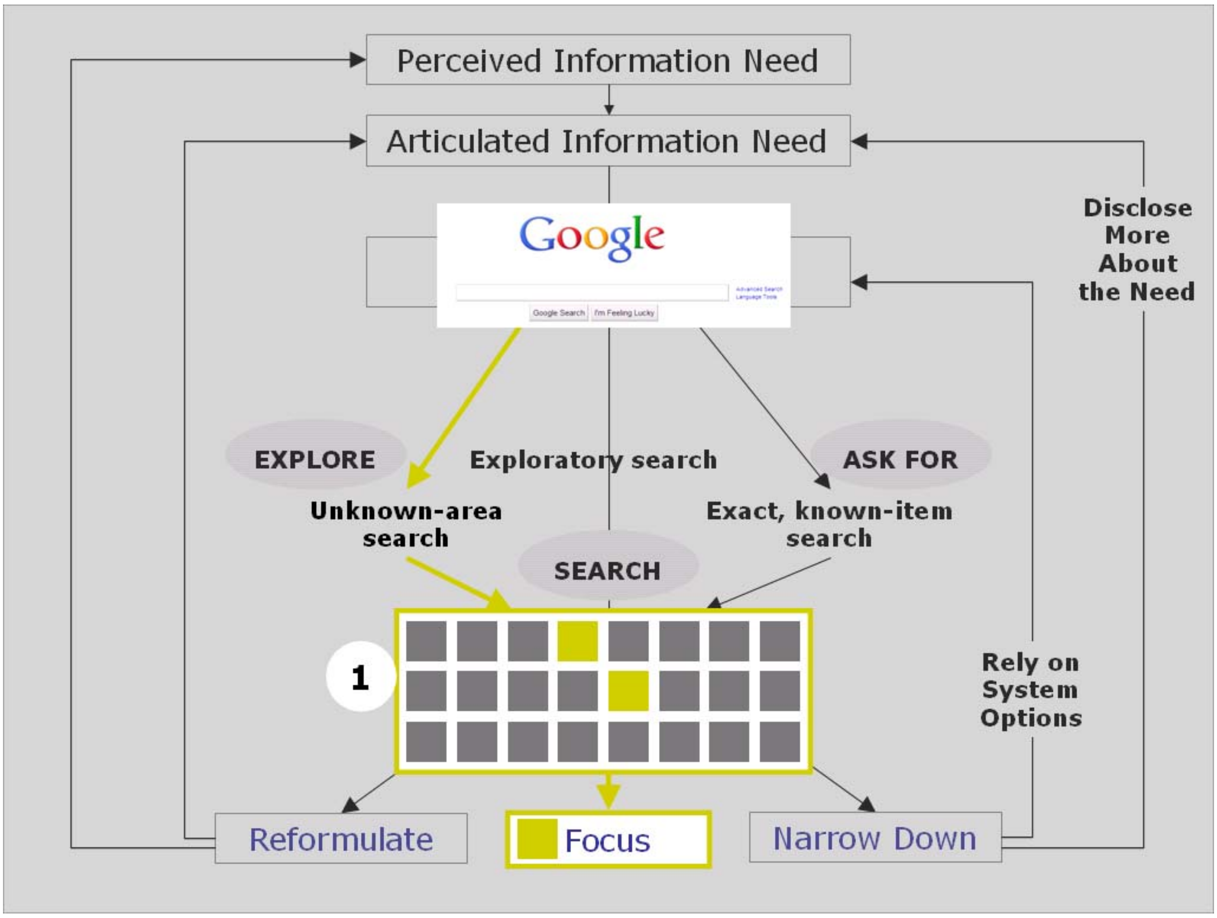


Figure 61: The information-searching model applied to scenario 7, step III

Step IV (Figure 62): George searches in SPIRES for information about the two instruments (1) and finds that these instruments were indeed used in some previous experiments (2). Now George needs to see how he can adapt one of them to make it suitable for his project.

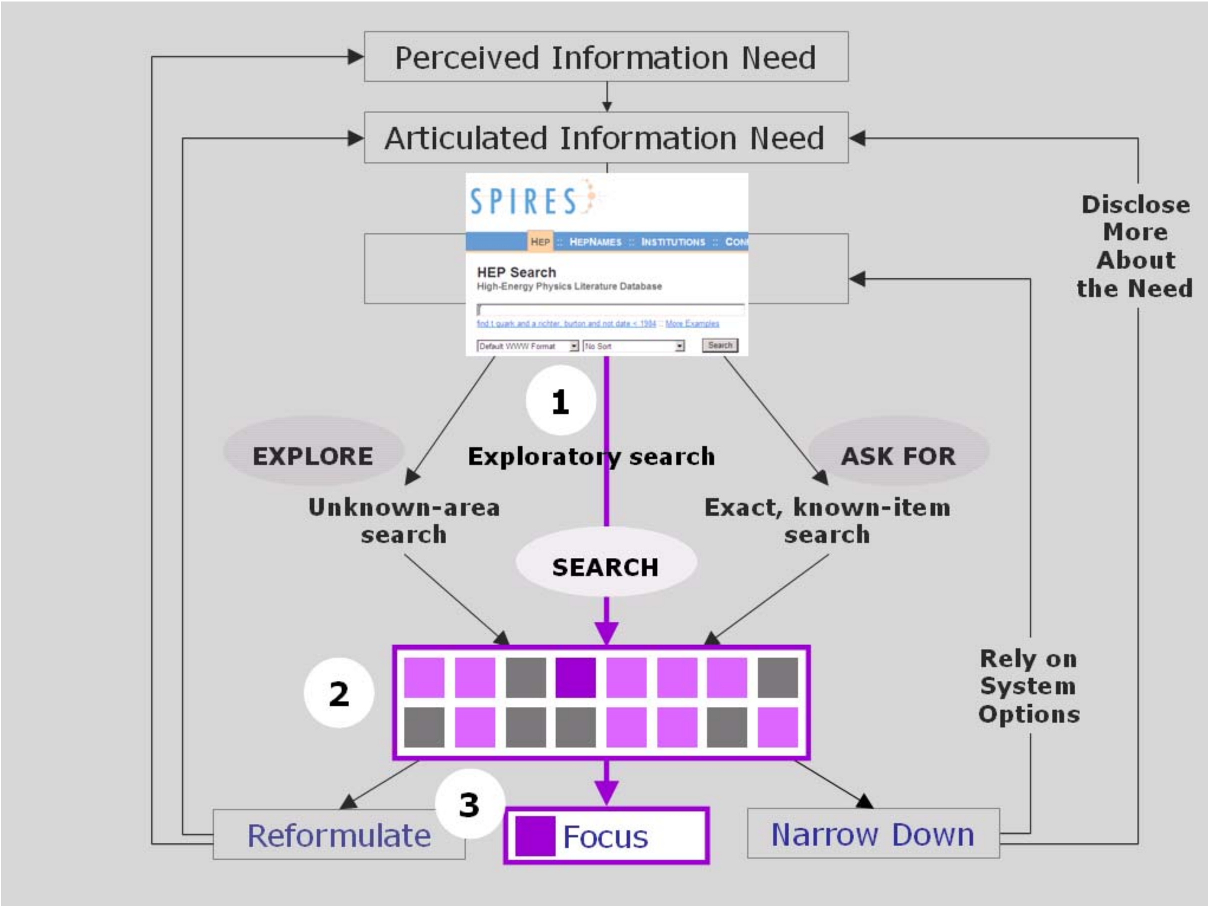


Figure 62: The information-searching model applied to scenario 7, step IV

5.7 Conclusions

Many information-seeking models have been proposed in recent decades; however, either they are too general to provide meaningful insights about the HEP community or they apply to only part of the information-seeking behaviour of that community. The information-seeking model and its subcomponent, an information-searching model, presented in this chapter were defined based on qualitative and quantitative research of the HEP information-seeking behaviour and in light of existing models. This new information-seeking model, tested in seven user scenarios, covers the full spectrum of active information seeking of the HEP community. Nevertheless, the model is general enough to be applied to other scientific communities as well. Further discussion of the model’s applicability to other communities appears in Chapter 7.

Chapter 6 User-Interface Design Derived from an Information-Seeking Model

6.1 Introduction

The model described in Chapter 5 of this thesis represents the information-seeking behaviour of high-energy physics (HEP) community members. To carry out their information-seeking practices, these researchers are currently using multiple information systems, as evidenced in the qualitative and quantitative research described in Chapter 3.

The suite of SPIRES databases, which has been serving the HEP community for over three decades (see Chapter 3), is being migrated to a new technological platform, INSPIRE, under development since 2008. The migration to the INSPIRE infrastructure provides an opportunity to also offer a new user interface. This chapter describes a high-level conceptual user-interface design, derived from the information-seeking model described in Chapter 5, that is likely to serve the team of HEP information specialists who are developing INSPIRE.

Following the description of the suggested user interface, this chapter presents the way in which the user-interface design coheres with the information-seeking model described in Chapter 5. The personas described in Chapter 4 are used to corroborate the applicability of the user-interface components to HEP researchers.

6.2 The Need for a New User-Interface Design

The results of the HEP survey launched in the summer of 2007 and described in Chapter 3 indicate that SPIRES is highly appreciated by the HEP community. However, whereas the quality of the SPIRES content is not debatable, the user interface of SPIRES, in particular its search interface, has aroused much discussion. About 30% of the respondents referred to the search interface in their open-ended answers, half of them praising it and the other half criticizing it.

One of the major objectives of the INSPIRE team was to determine how to better serve the community members who are not happy with the SPIRES user interface but at the same time to avoid disappointing the avid researchers who have relied on SPIRES for many years and would not like to see it changed. In addition, the team needed to consider to what degree contemporary, non-scholarly community tools should play a role in INSPIRE. Finding the balance between the desire to innovate and the need to keep the model that has been so successful for more than three decades is challenging but necessary: the survey results show a drift toward simpler tools—Google and Google Scholar—among young researchers. The new INSPIRE interface needs to address the expectations and the search habits of all HEP members, including the younger researchers, so that it continues to serve as the backbone of HEP research.

This chapter is based on a proposal submitted to the HEP information specialists in June 2009, a proposal that aimed to provide guidelines and concrete ideas that would assist the team in building the user interface of INSPIRE. Some of these ideas have already been implemented by the INSPIRE team while others are under consideration for future versions of INSPIRE.

6.3 Overall Approach to an Actual User-Interface Design

Although the proposed user-interface design was derived from the information-seeking model presented in Chapter 5, the design is an actual proposal; hence, it is offered as a concrete framework relating to technical infrastructure, development capabilities, and the short-term and long-term goals defined by the INSPIRE team.

The proposed user-interface design aims at improving several aspects of the information-seeking practices of the HEP community members. These aspects emerged during the interviews described in Chapter 3 (and summarized in Appendix A) and were further highlighted by the survey results, described in Chapter 3.

The goals of the new design are as follows:

- Provide a comprehensive environment that encompasses the various information-seeking practices of HEP researchers

- Leverage the structure of scholarly publications to streamline information-seeking processes
- Expand the search scope of INSPIRE to cover additional content and new types of materials without overloading the search results with irrelevant materials
- Reduce the average time of information-seeking sessions; in particular, help researchers who are conducting exploratory searches focus on the most relevant documents
- Lessen the difficulty of new researchers in learning information-seeking practices
- Create an interactive environment where researchers add information actively, not just through writing research publications
- Provide a modern approach to HEP information seeking and maintain the innovation that characterizes the scholarly research of the HEP community

6.4 General Recommendations

To achieve the goals of the new interface design, the INSPIRE team should consider following recommendations that are related to the INSPIRE search scope and the handling of researchers' profiles by the system. Although not part of a user-interface design, these two issues are crucial for building a comprehensive scholarly environment and offering personalized services.

6.4.1 Opening Up the Search Scope

As evidenced in the survey results, the focus of SPIRES on HEP materials is both an advantage and a disadvantage (see Chapter 3). On the one hand, such a focus ensures that all items received as search results are relevant and trustworthy. On the other hand, relevant information can be missing when a researcher relies solely on SPIRES, particularly when the researcher is looking for interdisciplinary materials, materials in tangential fields, or materials that are not research articles (reports, lecture notes, presentations, and more).

According to the survey, HEP researchers use several mutually complementary information systems to satisfy their information needs—primarily SPIRES, arXiv, and Google services (Google and Google Scholar). Although researchers manage to set up their environment in a way that enables them to use each of these information systems for different tasks, a tighter integration between the systems is recommended. INSPIRE can achieve such integration by mashing up[20] data and services originating from arXiv, Google, Google Scholar, and perhaps other information systems. For example, arXiv's service of listing recent submissions to arXiv is of high relevance to all HEP researchers; materials that are not articles and that can be found by Google and Google Scholar are of interest to many HEP researchers; and the option to expand searches to scholarly databases in fields that are tangential to HEP—such as the Astrophysics Data System (ADS) or even to arXiv (which is not limited to HEP materials)—will make INSPIRE all the more relevant to researchers. Still other systems, such as specific subject gateways, may be of value in the future.

To mash up data and services, INSPIRE should use technological programming interfaces that would provide access to information systems from INSPIRE in a way that is transparent to researchers (although each such information system is independent and researchers can access it directly).

6.4.2 Setting Researcher Profiles

SPIRES is a free service and does not require researchers to log on so that the system can identify them. However, in order to enable users to personalize settings and to allow authors to update their personal pages and assert authorship of their articles (when name ambiguity occurs[21]), INSPIRE must include a logon mechanism. Supporting authors who want to carry out the procedure of asserting authorship of an item will ease the load of the INSPIRE team members, who work hard to keep the information in the database accurate.

---

[20] A mashup is a method by which a Web application combines data or services from two or more sources and offers the combination as a new service.

[21] Author names appear in publications in the form required by the specific publishers. Thus, variations such as 'Smith, John', 'Smith, John W.', 'Smith, J.', 'Smith, John Wilfred', and 'Smith, JW' are treated by SPIRES as the names of different authors, whereas in reality, these variations all refer to the same author. Furthermore, researchers may have difficulty identifying authors who have common names; for example, 'Smith, J.' may apply to 'Smith, John', 'Smith, Jeffery', 'Smith, Janet', or others. Typically, when searching by author, researchers make the distinction based on further information known to them, such as the institution with which the author is affiliated, the area of research, and other authors of the same article. One of the tasks that the INSPIRE team is carrying out consists of grouping all the variations of an author's name and associating all the author's papers with a single author profile.

Furthermore, authentication is recommended to prevent malicious use of a system and to facilitate trustworthy contributions of information such as ratings, reviews, and tagging. In the future, an authentication mechanism will also enable INSPIRE to include personalized services, such as relevance ranking tailored to an individual's needs, configured by the individual or automatically created through the tracking and analysis of the individual's search and selection patterns.

Because most researchers tend to use their own computer or mobile device, there are other techniques that can be used to set up a personalized environment without requiring an explicit logon—such as the use of 'cookies'[22] stored on the researcher's computer. Such an approach is likely to be more popular among HEP researchers but may be less accurate—for example, when a researcher uses several devices—and will still not accommodate services such as asserting ownership of an item. Another approach could involve an initial logon, but the system would retain the user credentials from one session to another through the use of 'cookies' (similar to the Google approach), eliminating the need to log on every day.

6.5 Requirements for HEP User Interface

The proposal for the INSPIRE user interface takes into account the following requirements, assembled by the author on the basis of the research described in Chapter 3 and correspondence with the HEP information specialists:

- Serving as the major information system of the HEP community, the user interface must be simple, straightforward, and clear. It should not require training.

- The user interface should be extensible and flexible enough to allow the INSPIRE team to develop it over time and to modify features, add new features, and remove features that turn out to be not as useful as anticipated. Such changes should be easy to carry out on an ongoing basis

[22] A cookie is textual information stored by a Web browser on the user's computer. By using cookies, an application—such as INSPIRE—can store information that it needs for authenticating the user and setting the user's preferences.

and should not cause confusion or frustration on the part of the researchers.

- Community tools such as those that enable researchers to tag and review articles should be added in such a way that enables the INSPIRE team to carefully monitor and evaluate the use of these tools. The implementation of the tools should reflect the scientific nature of the community.

- Because many HEP scientists show tremendous loyalty to SPIRES and do not see a need for a change, the system should offer a 'classic' interface option, which will be very similar to the traditional SPIRES interface from the researcher's point of view, in addition to the new, more revolutionary interface. Researchers should be able to set the preferred interface as the default option.

- The system should be easy to use and enable researchers to customize the interface to their needs. The system's default settings should reflect the requirements of the majority of researchers and ease their transition to the new interface.

- The system should provide guidance and suggestions at the point of need and refrain from relying on the researcher's memory, particularly in regard to authors' names and journal titles.

- As described later in this chapter, various design elements and services may be of interest to HEP researchers. However, in order not to clutter the screen with options that are of no interest to a specific researcher or in a specific context, researchers must have control over the elements that make up the screen. That is, researchers should be able to choose and position the screen elements that they prefer, thus arranging their computer 'desktop' to accommodate their exact needs, their way of thinking, and their expertise. Such setup options should be platform dependent: for example, when accessing INSPIRE from a mobile device, a researcher may wish to see only the basic screen elements, whereas on the computer, the researcher might prefer more rather than fewer screen elements. Furthermore, the default interface should be applicable to most researchers so that they do not have to modify it in order to start working with the system.

- Regarding search syntax, the system should support both Google-like searching—allowing researchers to enter a few terms and, optionally, a Boolean operator, as they do when they search in Google—and the SPIRES-formatted searching that exists in SPIRES today (for example, *find author gross, e and title higgs and date > 1997* for finding articles that were authored by E. Gross, have 'higgs' in the title, and were published after 1997).

- The system needs to support a traditional search process whereby the researcher enters a query and browses through a list of results. However, the system should also offer tools that help the researcher better grasp the content of long result lists and use various criteria to drill down to subsets of such lists. The system should also offer navigation tools that enable researchers to find materials that are not displayed in the result list. For example, if a researcher is focusing on a given article, the system should be able to offer links to articles that were published in proceedings from the same conference as the article in focus, that were written by other members of the department with which the article's author is affiliated, or that were viewed during the same search session by researchers who looked at the given article. Furthermore, the system should be able to present items that are not articles but are related to the article in focus, such as datasets, reports, and lecture notes.

6.6 Computing Requirements

The assumption underlying the design presented in this chapter is that HEP researchers are equipped with up-to-date computers and Internet browsers and will increasingly access the system using mobile devices as well. Most interactions with the system will be through personal or laptop computers rather than through public terminals.

6.7 Accessibility

The system should adhere to the international accessibility guidelines published by the World Wide Web Consortium's international accessibility guidelines (Web

Content Accessibility Guidelines 2.0; http://www.w3.org/TR/WCAG20/)—and to the United States government's accessibility guidelines (Section 508 of the US Rehabilitation Act; http://www.section508.gov/).

The suggested design relies on JavaScript for enhanced usability. However, if JavaScript is disabled in the user's computer, an alternative display can be offered. To support both types of display, some programming effort is required. For example, screen elements that can be expanded and collapsed will always be expanded when JavaScript is disabled. Furthermore, because it is recommended that the current design of SPIRES be retained as the INSPIRE 'classic' style—which conforms to the accessibility guidelines—adherence to accessibility guidelines will still be achieved.

6.8 General Layout

An INSPIRE Web page should be built of *tiles*—rectangles that contain data elements and related services—that interact with each to provide a complete user experience. An example of a tile-based screen is the iGoogle window (Figure 63). In the context of INSPIRE, faceted browsing (see 6.14.3), for example, is a functionality that is offered through a single tile that can be placed to the left or right of a result list.

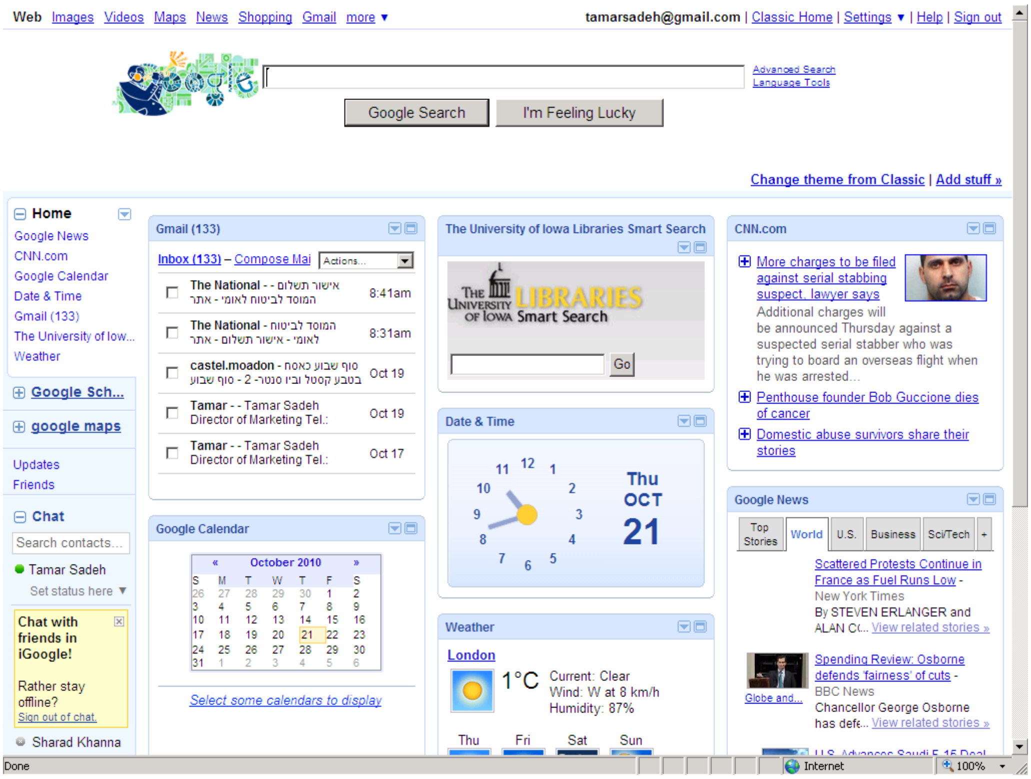


Figure 63: An iGoogle tile-based screen

A researcher can add or remove a tile, expand and collapse tiles, and arrange them to create a layout that suits the researcher's needs (for example, the researcher might put a tile displaying the recent submissions to arXiv at the top of the screen).

The major sections of the proposed screen are as follows (Figure 64):

- **Banner**: the INSPIRE branding area
- **Personalized services**: the logon area, saved records, options for setting a personal profile, and so on
- **Search settings**: the type of search (e.g., search for articles, authors, or conferences) and the search scope
- **Search box**: the query entry box, a link to the advanced search function, and a reset button for starting the query over
- **List header**: the number of items on the result list
- **Sorting order**: the active sorting order, such as by date in descending order; researchers can change the order
- **Selected facets** (if facets were selected for drilling down in a list): the facet values that the researcher selected; these values serve as a navigational 'breadcrumb' trail

- **General services**: recent postings to arXiv, news, popular queries, suggested new searches, and so on
- **Faceted browsing** area: facet categorization with expandable lists of values
- **Result list**: brief information about each item, accompanied by a limited number of services (e.g., an option to download the citation) and a link to the full text. Each item can be expanded to include more details, offer more services, and provide navigation to related items. Some navigation options are displayed on separate tiles.
- **Navigational services that apply to the expanded item** (the result that is in focus): a list of references, list of citations, recommended reading, and so on

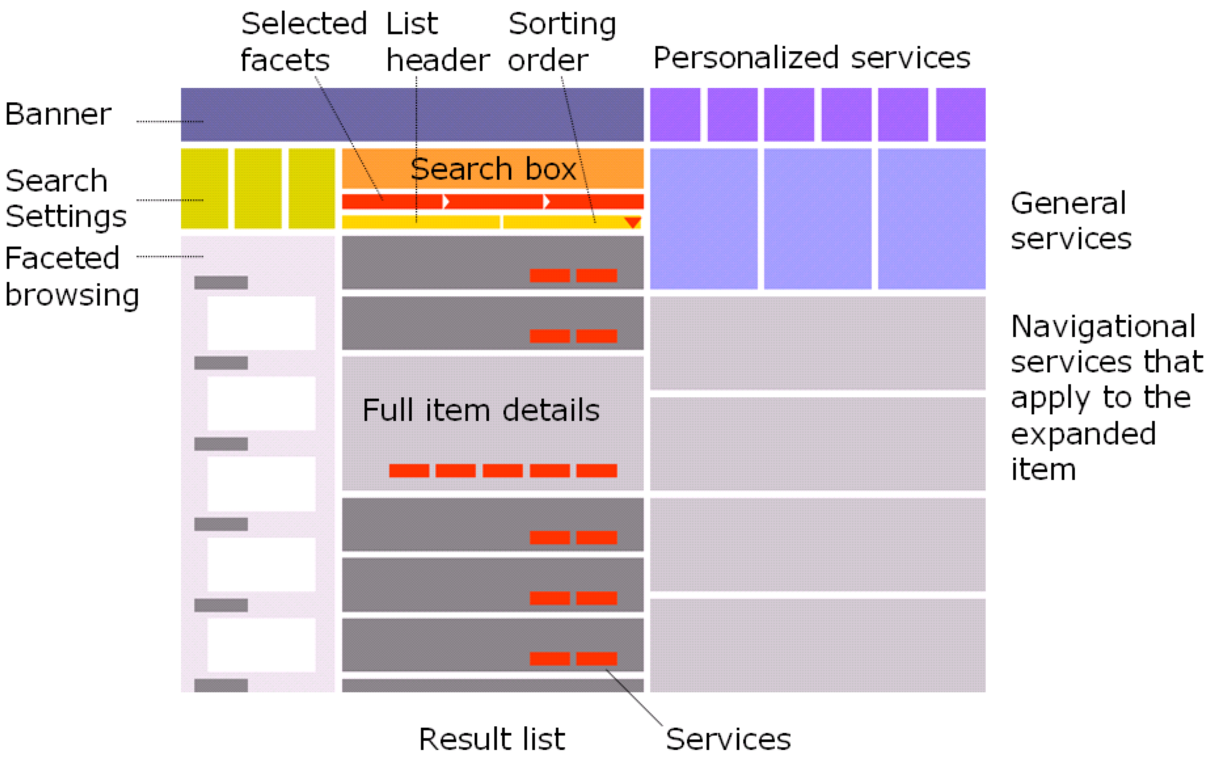


Figure 64: The general layout of the proposed INSPIRE screen

More thorough descriptions of the screen sections are provided later in this chapter. Detailed screen elements, such as the selection of multiple results and the navigation to the next and previous pages, need to be added as part of a detailed design.

Tiles can be multiplied by the researcher as necessary. For example, when an article is in focus, the system displays one or more tiles that include links to related materials (each type of link has a dedicated tile). The predefined set of tiles can be modified by the individual users. A researcher who does not wish to see the links to recommended articles, for example, can remove the tile.

6.8.1 Entities to Be Searched

In addition to scholarly materials, SPIRES keeps track of other information entities such as researchers, institutions, and conferences. Each such information entity can be the ultimate target of a search or can be only a means for a researcher to navigate from one article to another. For example, a researcher may want to look for a particular conference to find out about its content or to use the conference item as a starting point for navigation—for moving to papers presented at the conference. The INSPIRE system should enable the researcher to select the type of material that is the target of the specific search: scholarly publications (the default option), people, conferences, or institutions.

In addition, navigational links that the system provides for each item should enable researchers to navigate from one type of information entity to another entity of the same type or a different type.

6.8.1.1 Scholarly Materials

Articles are the obvious search targets, so the interface needs to be optimized for searching for them. Other types of scholarly materials include books, conference proceedings, reviews, lecture notes, reports, and video clips. Scholarly materials, as a whole, should be set as the default entities to be searched. After the search, the researcher can use the *material type* facet to refine a result list to show specific types of materials. Researchers can also specify the type of material as part of the query, as they already do today.

6.8.1.2 People

Finding an author is a common path to finding relevant materials. The system should enable researchers to find people easily. For historical reasons, ambiguity of author names is common in SPIRES searches; hence, automated suggestions are needed to make researchers aware of all the name variations of an author. Such suggestions also enable researchers to overcome inaccurate spellings of names. When looking for papers by authors, researchers must have a way to look for papers written by individuals (as opposed to papers resulting from a collaboration[23]).

The result of an author search should display an author page listing the publications written by the author and other details, such as the affiliation, the professional history, and the author's students (this type of information is already offered in SPIRES). Researchers should be able to navigate from an author page to articles written by the author.

Authenticated community members should be able to edit their own pages and to 'claim' articles that they have written, thus helping solve the ambiguity of author names. Community members should be able to inform the INSPIRE team of ownership inaccuracies related to papers by other authors but should not be authorized to fix such inaccuracies themselves. An author page should include text added by the author and links to related sites, such as the personal page of the author on the institution's Web site. There should be an option for adding a picture.

6.8.1.3 Conferences

Conferences are of interest to some researchers. The researchers should be able to navigate to materials presented at the conference, including presentations, if available in INSPIRE, and to related conferences.

6.8.1.4 Institutions

Institutions can serve as a hub for finding people and publications.

6.8.2 Search Interface

Although the traditional process of submitting a query, skimming the result list, and selecting one or more items from the list is still relevant for all researchers, they tend to invest less time in articulating a query and spend more time processing the results. Furthermore, researchers strive to find an anchor—one item that is relevant to their needs—and then accumulate other items of interest

[23] Many projects in the HEP domain are carried out by collaborations of thousands of researchers. Papers that are the outcome of such projects are submitted by all collaboration members and hence have thousands of author names.

through navigation from the anchor item to other items, using a web of links that is provided by the system.

As discussed in Chapter 5, a search can be for a known item (*Ask For* in the information-searching model), in which case the researcher is likely to type or paste an article title, an arXiv ID, an author name, or a combination of any of these in the search box and receive a relatively small number of results; it can be an exploratory search (*Search* in the information-searching model), in which case the system may display a large number of results and must provide researchers with tools to sort them out; or it can be a search related to an unknown area (*Explore* in the information-searching model), in which case the system should provide suggestions for new queries in addition to the tools that can help researchers refine the result list. In all these cases, the system should facilitate the modification of an existing query and the creation of a new query.

When searching, researchers convert their perceived information need into an articulated information need, that is, a combination of search terms (a query) and pre-search selections that characterize the requested results and thus narrow down the search scope. Such pre-search selections may include the type of materials (such as article, review, book, or lecture notes), the language in which the materials were written, and, in a search environment that covers multiple information repositories, the repository in which the system should search. Researchers can use default values for such settings or change the settings as appropriate.

Two pre-search selections are proposed for INSPIRE: the entities to be searched (see 6.8.1) and the search scope.

6.8.3 Query

The default interface should allow for both Google-like queries (e.g., one or more search terms, quotation marks for exact phrases, some Boolean operators, and commands such as *define:*) and queries that adhere to the syntax that is offered today by SPIRES—for example, a SPIRES query for an article that has ‘quark’ as part of the title field, was written by Burton Richter, and was published in 1984 or later would look like this: *f t quark and a richter, burton and not date < 1984*. The system needs to be able to parse both types of queries and interpret them.

In addition, an advanced search form should be offered as an option. The system's default setting should be the basic search, but researchers should be able to set the advanced search as their default option.

The advanced search form should display field values (based on the actual data in the system—author names, publication titles, institution names, journal titles, and so on) as the researcher is entering the query. For example, the researcher starts typing an author name, and the system immediately displays actual author names that correspond to the typed letters, narrowing down the list as more letters are typed. In addition, a list of values (for example, journal titles and material types) is displayed when relevant and upon request, but the researcher is always able to type a value rather than choose it from the list.

The query box should be available at the top and bottom of each page. In both locations, the query box should include a new-query option as well as the query at hand, because the latter provides the context for the result list and because people often add query words to narrow down searches. If the page displays an item to which the researcher navigated (not a search result), the query box is empty.

The researcher can explicitly expand the search scope to cover other scholarly databases, Amazon, Google, and Google Scholar. The system can offer this functionality either by including results obtained from other information systems in the search results or by offering additional results that are not blended with the main results, primarily when results are obtained from systems that are of a different nature, such as Google.

To enable researchers to search in additional databases that are similar in nature to the SPIRES databases, the search interface needs to include a list of such databases; the system should enable the researcher to select one or more resources from the list. If the number of resources on that list is seven or less, the system should display all resources; otherwise the display should be organized in expandable groups, by topic or by proximity to the HEP discipline, and allow multiple selections. Researchers can set their own default scope and, if they adhere to the system's default scope, they can collapse this option altogether (but can expand it at a later stage).

It is recommended that INSPIRE simultaneously search in information systems that are loosely related to HEP, such as Google, and place the results on designated screen tiles, one tile per information system. The researcher can link to each result or, when relevant, switch to the information system's interface to see the search results there.

If the researcher selects one or more facet values to narrow down a search (see 6.14.3), the selected values should be displayed as a breadcrumb trail below the query and the system should allow the researcher to deselect any of the values, regardless of the order in which the researcher selected them. For example, the researcher narrows down an initial search by material type, year range, and journal title, yielding the following breadcrumb trail: *Reviews > 2005-2010 > International Journal of Modern Physics A (IJMPA)*. Now the researcher decides to remove the date range, so the system displays just the material type and journal title: *Reviews > International Journal of Modern Physics A (IJMPA)*.

6.8.4 Result List and Full-Record Display

The result list proposed for INSPIRE has the following characteristics:

- If there are no results or the number is very small, the system offers alternative terms (Did you mean…?). The system automatically expands the search to include related terms, predefined in the system, such as British and American spellings of the same word, singular and plural forms of nouns, and general or discipline-specific acronyms along with their spelled-out version (e.g., *SUSY* and *supersymmetry*).
- There is more than one sorting order, and researchers can set their preferred sorting order as a personal default setting. The sorting order can be by date (descending and ascending), relevance to the query (without additional considerations such as popularity or citation rate), and number of citations. An optional sorting order could be popularity (number of people who have demonstrated interest in the item) and a combination of all of the previous factors (relevance to the query, date, number of citations, and popularity). The system can determine the popularity sorting order by quantifying the kind of interest that people showed in an item (whether they looked at the full-record display, downloaded the full text, or saved the citation, for example).
- The heading of the result list specifies the number of results and the sorting order, which the researcher can modify.

- The system suggests new searches based on an analysis of the query and the result list. For example, the system could suggest searching for topics (keywords) that characterize the list or for terms that are related to those used in the query, helping researchers avoid dead ends in their searching process.
- Faceted browsing is offered to enable a researcher to narrow down large result lists, by the type of material, date range, topic, journal title, publisher name, collaboration, institution, online availability, or other factors. Researchers can remove facets that are not helpful for them; for example, a researcher who is not familiar with publishers may decide not to have the publisher facet displayed on the screen. A discussion of faceted browsing in greater detail is provided later in this chapter (6.14.3).
- The number of results per page can be customized. The default number is 25, and researchers can set any other number as their preference.
- Services are offered for a whole page or for specific items. The researcher can select some or all items displayed on the page and request a service such as printing, downloading citations in a specific format, 'pushing' citations to another tool, and sending items by e-mail.
- The system initially displays only basic information for each item, but each record can be expanded—on the same screen—to show the details of the full record. The basic information includes the title, authors and their institution (or collaboration), the publication date, citation information, and the number of citations. Online availability is displayed prominently, including a link to the full text. In addition to offering a link to the full text in arXiv—if the full text is available from there—the system displays a link to the published version, if it exists. For researchers affiliated with institutions that implemented an OpenURL link resolver, an OpenURL is added for each item. The use of the OpenURL enables researchers to access the copy of the full text for which their institution has a subscription.
- If an abstract is available, the system offers it as part of the full record display and highlights query words that appear in the abstract.
- If full text is used for the search and the item was selected because of words that appear in the full text, the system embeds—in the full record display—a snippet from the text that includes the query words. Query words are highlighted wherever they appear.

- The full-record display includes only the information that is helpful to the researcher. Information that is not likely to differ from one record to another, such as the list of arXiv mirror sites, is collapsed by default but expandable upon request. The researcher can set one site as a personal preference and change the selection when needed.
- The full-record display may include links to information elements that relate to the article, such as datasets, spreadsheets, images, and video clips.

6.9 Navigation from a Selected Item

Every selected item should be considered a starting point from which the researcher can navigate to other items. Navigational links should represent the associations discussed in Chapter 5 and should be derived from the metadata fields of the item and from information gathered by the system, such as usage data. Once the researcher navigates to a new item, that item becomes the main item on the screen and can offer navigational links to further items. The researcher can thus follow a web of links that characterize scholarly publications.

The INSPIRE system should offer navigation through various methods, which can include metadata fields, references, citations, tags, and suggestions.

6.9.1 Metadata Fields

Some metadata fields can serve for navigation to items that share the fields' values with the item in focus. A researcher who clicks a field such as conference, for example, expects to see works that were presented at the same conference as the item in focus. The system performs a new search—with the clicked term as the search term—and displays a new result list.

6.9.2 References

Every reference should include one or more hypertext links to the full text of the item itself on arXiv and on the publisher's site when the full text is available in either of these locations, and also to the metadata item in INSPIRE, if it exists in INSPIRE. If the full text is not available to the researcher, the system should display only the INSPIRE record, if it exists. If the full text is not available online, other institutional services such as information about print holdings and document delivery should be displayed (these can be generated by an institutional link resolver).

6.9.3 Citations

The information about articles (indexed in SPIRES and INSPIRE) that cite the article in focus is gathered by SPIRES (and INSPIRE). Links to such articles should follow the same guidelines as for references (full text, INSPIRE record, and institutional services, when applicable; see 6.9.2).

6.9.4 Tags

Tags are expected in today's systems, but in the HEP context, the implementation of tags can be considered of low priority.

When a researcher clicks a tag, the system displays other documents that share that tag. Tags should be displayed in two groups: a personal group, where the researcher sees other documents to which he or she assigned the same tag; and a community group, where the researcher sees documents to which the tag was assigned by all community members.[24]

6.9.5 Suggestions

The INSPIRE system should provide a list of suggested articles, similar to Amazon.com's "Customers who viewed this also viewed…" suggestions. The list is based on an analysis of the information-seeking behaviour of researchers as recorded by the system over time.

6.10 Navigation Options Not Related to Searches

General navigation options should appear on the screen regardless of the query. Each such option is displayed as a separate tile, and researchers can remove any of these tiles. The following options are proposed:

- **Recent Papers in My Area:** recently published articles, as per a set of preferences defined by the researcher. This arXiv service, which provides such lists of recent publications, can be incorporated in this tile. The researcher can click the tile to navigate to the full list in arXiv. The

[24] Tags gain relevance only when there is enough activity around them—that is, enough researchers define tags that are of a general nature and assign them to records.

researcher can define several such tiles and set preferences for each, thus monitoring several research areas.

- **My Recent Queries**: recent queries within current session. The researcher can click a query to re-launch it.
- **Recent Queries**: a list of the community's recent queries. The researcher can click a query to launch it.
- **Popular Queries in the Last Week**: a list of the community's popular queries in the last week. The researcher can click a query to launch it.
- **Recommended articles**: a list of articles that community members have recommended explicitly. The list should be sorted by date.
- **My Items**: records marked during the current session. If the researcher has logged on, records can be saved from session to session.

6.11 Services

Services should be offered for a specific record, for multiple selected records on a single page, or for a page of results. The system should enable the researcher to select and deselect parts of the list on each page.

Some services, such as rating and reviewing items and sending corrections to the INSPIRE team, apply only to a specific record and should appear in the context of a full-record display. Other services, such as printing, downloading, saving, tagging, and sending by e-mail, can apply to a selected list.

6.12 Community Tools

Although tools that enable social networking are becoming popular in other domains, the inclusion of such tools in the context of INSPIRE is questionable: for scientists, writing a review is considered an important work-related task (typically they would publish a formal review), and hence most researchers would not write brief reviews or rate items in INSPIRE. Furthermore, social networking typically requires that the researcher be identified by the system; HEP researchers are not accustomed to logging on to SPIRES and arXiv and hence will be even less likely to use social networking tools.

However, some researchers are expected to review items as part of their service to the community.

6.12.1 Reviews and Ratings

Researchers can assign a rating to a record. To make the rating more meaningful, assigning it should be in conjunction with writing a short review, although the researcher can refrain from doing either.

6.12.2 Tagging

Researchers can assign tags and keep them private or designate them for public use. As a researcher types a tag letter by letter, the system should display a choice of existing relevant tags that start with the same letters.

6.12.3 Recommendations

Unlike the automated suggestions derived from the system's monitoring of the community's information-seeking behaviour ('people who looked at this item also looked at…'; see 6.9.5), recommendations in the community area are provided explicitly by researchers who are logged on to the system. Once there is a great enough volume of recommendations, the list of recommended articles can be displayed as part of the general services, and recommendations can be shown along with the items that they describe.

6.13 Researcher Preferences

Researchers who are logged on can set their preferences, such as the number of results per page, the search scope, the sorting order, the tiles they want to see on their desktop, and the arrangement of the tiles on the screen. These settings can also be saved as 'cookies', in which case the system will always use the most recent settings on the specific computer or mobile device.

6.14 The Researcher's Workflow

The suggested workflow for the new Web interface is illustrated in the figures in this section. The schemes presented in these figures do not attempt to cover the full spectrum of options; rather, they illustrate the main components and the interaction between them. (A more detailed description of the suggested main screen is provided in Figure 64)

The workflow is described under the assumption that all basic components are present on the screen.[25]

6.14.1 Entering the System

There are two entry points to the Web interface of INSPIRE: the researcher can enter the system for the purpose of looking for information, either by active searching or by monitoring information that is 'pushed' to the researcher by the system (Figure 65); or the researcher can navigate to a specific item in the system by following a link offered by another system such as a reference management tool or a Web search engine. In the latter case, the entry point is the one described in 6.14.5.

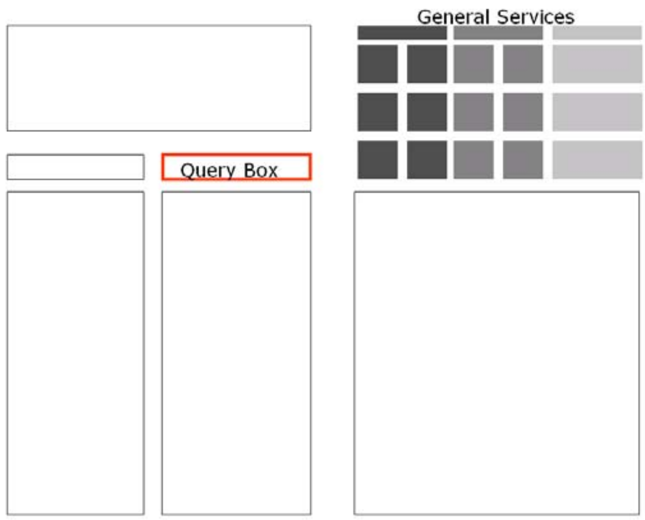


Figure 65: Main entry point to the INSPIRE Web interface.
The information-seeking process begins at this point. The query box (outlined in red) serves as the initial step for directed searches.

---

[25] Researchers may decide to remove screen components, in which case the workflow may differ from the one illustrated here. If fewer components are present, less functionality is offered. For example, if the faceted browsing tile is removed, the only way the researcher can narrow down a result list is by adding a query term.

Upon entering the system, the researcher can take advantage of the general services (see 6.10), which are already displayed as per the researcher's preferences, for an undirected information-seeking process. These services may include the list of recent submission from arXiv and news from other sources. The researcher can also start an active, directed information-seeking process by entering a query.

6.14.2 Querying

Once a researcher enters a query, a result list is displayed (Figure 66).

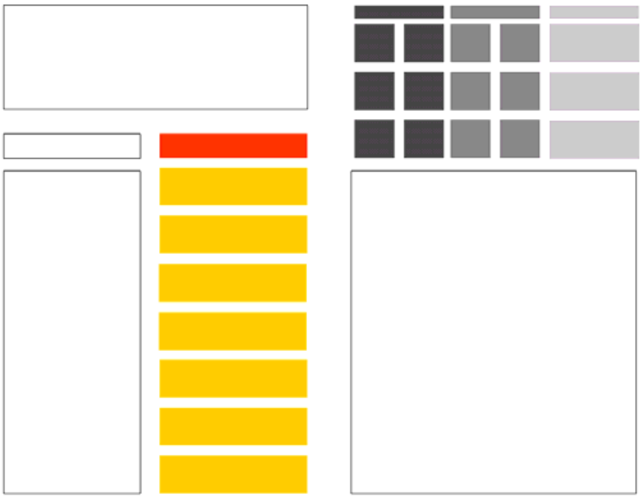


Figure 66: Result list (items in yellow) displayed as a response to a query

Depending on the search options set by the researcher, the result list may include scholarly publications, authors, conferences, or institutions. The suggested workflow is the same for all entities, but the options that are available at each stage differ depending on the retrieved entity's type.

The researcher now has several options:

- If the researcher decides that the results are not relevant to her information need, she can modify the query or enter a new query, obtaining a new result list.
- If the result list is too long, the researcher can refine the list by adding query terms and submitting the revised query or by using the faceted browsing options.
- If an item of interest is displayed on the screen, the researcher can use the item immediately (typically, the researcher would obtain the full text

of the item). The researcher can also focus on the item by choosing it, in which case the system displays more details about the item and all the items that are related to the item in focus.

6.14.3 Faceted Browsing

The system can exploit the result list by analysing it and presenting it as a multidimensional structure in which items are grouped in a way that corresponds to common methods of classifying information; researchers can then examine the list from various angles, gaining a better understanding of its content and its relevance to their query.

The system can group records in a result list by using system-predefined metadata fields that are available for each item on the list and can be regarded as attributes of the item. Such attributes include subject, author, type of material, journal title, and year of publication. Some attributes can have only one value, such as the year of publication, while other attributes—for example, author and subject—can have multiple values. The values for the subject attribute can be derived from an authoritative list such as the Physics and Astronomy Classification Scheme (PACS).

The system can display multiple dimensions of a result list by presenting the list of attributes alongside the result list and, for each such attribute, a list of the values that are shared by the records in that group (for example, one of the attributes could be the type of item, with the values *book*, *journal*, *article*, and so on). Each time the researcher selects a value, the result list is redisplayed to show all the items that share that value. In this manner, the system enables the researcher to look at the results from different angles, each time focusing on a specific characteristic of the result list. Furthermore, the list of attributes with their values provides a summary of the results: at a glance, the researcher can see which topics characterize the result list, which types of items the list contains, in which journals articles are published, and so on. As Hearst (2006) explains, such interface 'shows previews of where to go next, and how to return to previous states in the exploration, while seamlessly integrating free text search within the category structure. The approach reduces mental work by promoting recognition over recall and suggesting logical but perhaps unexpected alternatives at every turn, while at the same time avoiding empty results sets. This organizing structure for results and for subsequent queries can act as scaffolding for exploration and discovery.' (Hearst 2006, 61).

The system has two ways of identifying the values of the attributes that will constitute the multiple dimensions of a result list. In one method, typically referred to as *faceted categorization* or *faceted browsing*, the system groups the results on the basis of one attribute at a time—extracted from the structured, library-defined metadata—such as date of publication or subject. In the second method, typically referred to as *clustering*, the system scans numerous metadata fields of each record and extracts phrases that are repeated in multiple records. For instance, the system locates the phrase 'Introduction to Seiberg-Witten Theory and its Stringy Origin' in the title field in record 1, 'We analyze the action of Toric (Seiberg) duality on the combined mesonic and baryonic moduli space of quiver gauge theories obtained from D3 branes at Calabi-Yau singularities' in the abstract field in record 2, and 'Nathan Seiberg' in the author fields in record 3. In this example, the system identifies the word 'Seiberg' as the common thread among the three records, puts them in the same group, and labels it 'Seiberg', thus indicating to the user that all records under this label are related to Seiberg. (See Hearst 2006 and Sadeh 2008 for a discussion about clustering and faceted browsing.)

The way in which the information system displays subsets of the result list and provides access to them is the same for the two underlying technologies that are used to create the subsets—faceted categorization and clustering. Moreover, both technologies can be applied to the same result list. For example, the system might use clustering to group records by topic but faceted categorization to group records by type of material, journal title, publication date, and publisher. As Wilson et al. (2009) explain in their evaluation of advanced search interfaces, 'these modes [clustering and faceted browsing] can capitalise on users' ability to filter and navigate through information using recognition rather than recall' (Wilson et al. 2009, 1408). Kaki (2005) points out the benefit of faceted categorization primarily for exploratory search.

Figure 67 illustrates the way faceted categorization or clustering is offered to the researcher in the suggested INSPIRE interface. In this simplified scheme, results can be grouped by two attributes: the first—for example, material type—is represented by circles, where the circle colour indicates a value (for example, *article*, *book*, *review*, *conference proceeding*, or *lecture notes*), and the second attribute—for example, subject—is represented by triangles, in the same manner. In this example, each record on the result list may have one value for the first attribute (the circle) and one or more values for the second attribute (the triangle).

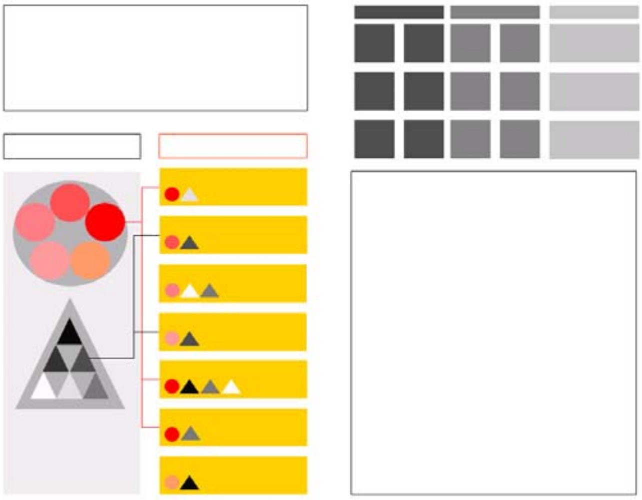


Figure 67: A result list displayed with facets for faceted browsing.
The circles represent one attribute, and the triangles, another attribute.

Once the values are presented on the screen, a researcher can select any of them and thus narrow down the list to show only the records that bear this value (Figure 68). The selection process is repetitive; that is, every time the researcher selects an option, a new result list is displayed, and on the basis of this list, the system updates the groupings and provides new options for selection. The selected value or values are displayed at the top of the list as an indication to the researcher that the list has been narrowed down by the specific value. The researcher can also remove the value, in which case the system refreshes the display to show the result list without narrowing it down by the specific value.

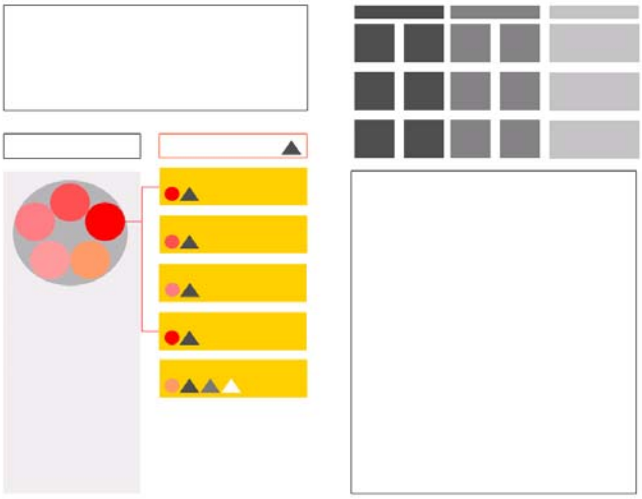


Figure 68: Result list narrowed down by the selection of a facet value.
The circles represent one attribute, and the triangles, another attribute.

The researcher can continue selecting options for any of the attributes until a concise, manageable result list is attained, enabling the researcher to focus on an item of interest (Figure 69).

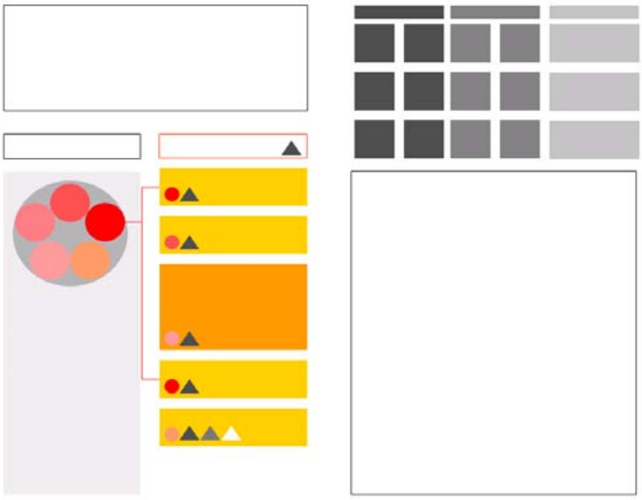


Figure 69: The researcher selecting an item to focus on (the orange rectangle).
The circles represent one attribute, and the triangles, another attribute.

If the initial search was for another information entity offered by INSPIRE—an author, an institution, or a conference—the system uses the corresponding attributes for constructing the facets. For example, authors can be grouped by research topic, institution, and academic degree.

6.14.4 Focusing on an Item

At any point in the process, the researcher can obtain the bibliographic record or the full text (if available online) and use it—for example, save it or send it by e-mail. Furthermore, once the researcher selects an item of interest, the system displays related items (Figure 70). Each colour indicates a different kind of relationship to the item in focus, such as items cited by the one in focus, items referencing it, items written by the same author, and items that other researchers looked at while looking at it.

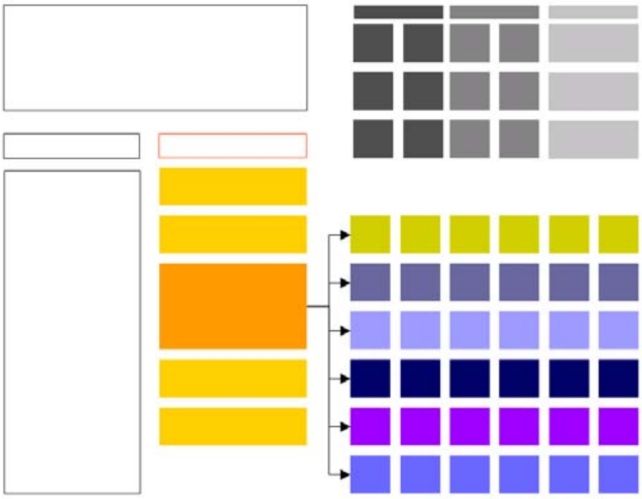


Figure 70: Display of items related to the item in focus

6.14.5 Navigating

The researcher can select any of the records displayed on the screen, thus making the selected record the item in focus (the blue box in Figure 71). In this case, the item has become the one in focus because of a link that the researcher followed, not because the item was the result of a query. Such a link can also be offered by another system, such as a reference management tool or a Web search engine, which directs the user to the context of a specific item in INSPIRE.

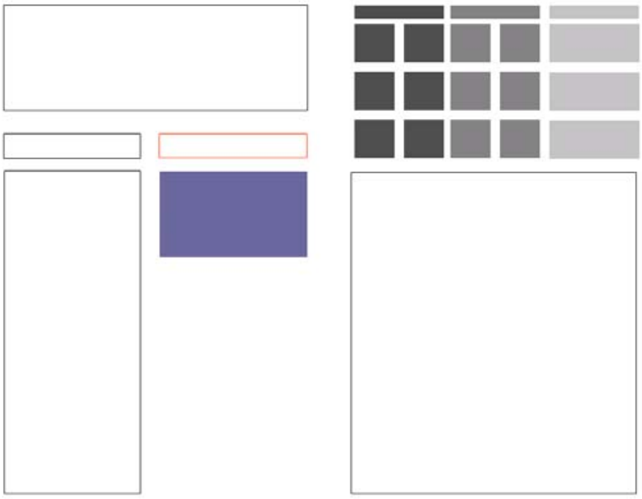


Figure 71: Item in focus.
The blue box represents an item that has become the one in focus as a result of the researcher's navigation.

When a new item becomes the one in focus, the system refreshes the screen and displays all the items that are related to this new item (the orange box in Figure 72).

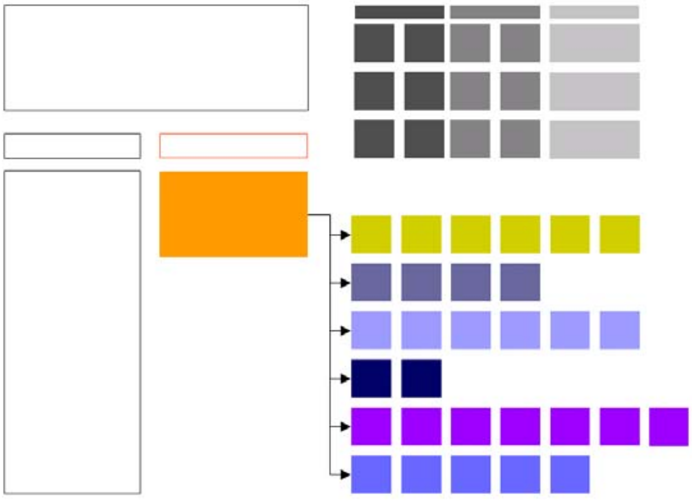


Figure 72: A refreshed screen.
The screen displays the web of links for an item that just became the one in focus.

The item in focus is not necessarily a scholarly output; it could be an author, an institution, or a conference that the researcher either searched for explicitly or navigated to. For example, the researcher can look for a specific institution or can navigate to the institution by following a link from an author's record. The list of related items depends on the type of item in focus: for example, if an author is in focus, the system may display the list of students of that author and the institutions to which the author is (or was) affiliated, in addition to the list of publications and the list of papers citing the author's publications. The researcher can thus navigate from one information entity to another.

6.15 How the Proposed User Interface Concretizes the Information-Seeking Model

The model detailed in Chapter 5 describes abstract information-seeking behaviour that is currently supported by several information systems. User-interface design, by nature, is restricted to a concrete information system and hence can only aspire to cover the full spectrum of the HEP information-seeking behaviour represented by the model. The illustrations in this section provide a high-level description of the way in which the various user-interface components correspond to the abstract actions and workflows portrayed as part of the information-seeking and information-searching models described in Chapter 5.

The information-seeking process can start with an explicit search or with the monitoring of lists of items that are generated automatically on the basis of the researcher's preferences (such as a list of new submissions to arXiv or INSPIRE) (Figure 73). Searching and monitoring correspond to entry points *a* and *b*, respectively, in the information-seeking model (see 5.4).

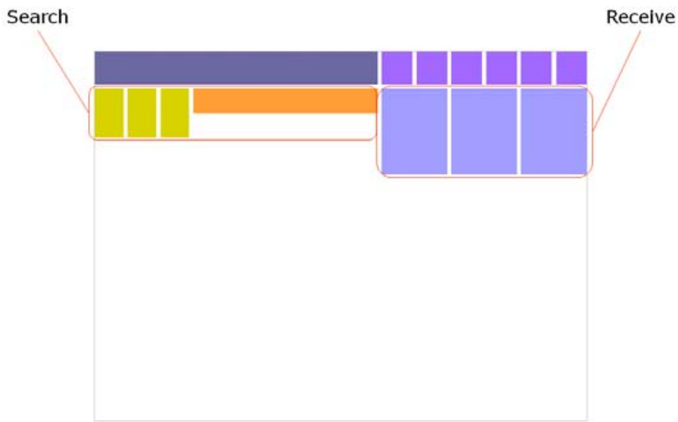


Figure 73: Entry points to the information-seeking process.
One entry point is for explicit searching, and the other is for the monitoring of automated updates.

Researchers can browse through the automated lists (Figure 74) and focus on items of interest (Figure 75); these actions correspond to the Browse and Focus actions in the model (see 5.4).

Figure 74: Browsing through lists of automated updates

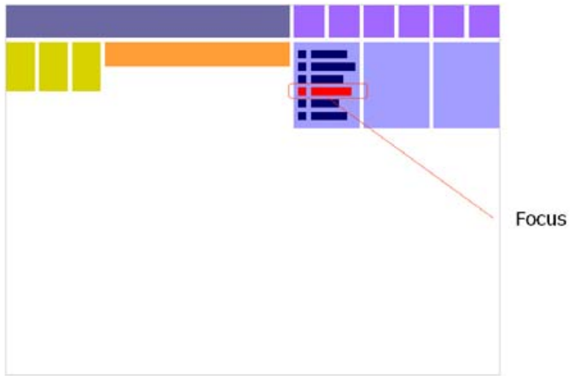


Figure 75: Focusing on an item of interest

As noted earlier, a user-interface design is constrained by the capabilities of the system. Although presented on one screen, automated lists can originate from various systems. If, for example, a list represents new materials in INSPIRE, the system will display the item in focus (see page 202). However, if the display is based on a service leveraged from another information system—for example, a list of new submissions generated by arXiv—a researcher who clicks an item is taken to the environment of that other system. The model describes the focus action as identical in these two cases, but in reality, the INSPIRE system alters its behaviour according to the origin of the information it displays.

The researcher who starts the information-seeking process with an explicit search typically enters one or more terms and, optionally, sets pre-search parameters—for example, to change the default search scope (Figure 76). The researcher may also decide to use an advanced search form. These activities correspond to the Query action of the information-seeking model described in 5.5.

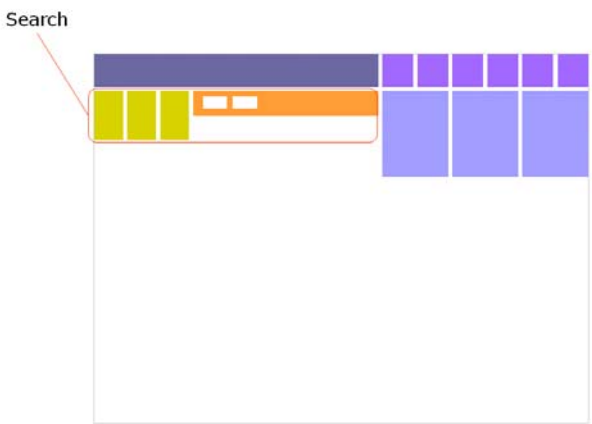


Figure 76: Searching by entering search terms and setting search preferences

The user interface looks the same for queries of the various types described in Chapter 5 (Ask for, Search, Explore; see 5.5). However, the type of search dictates the kind of services that are available or that the researcher is likely to find useful. For example, a known-item search (Ask for) will probably yield only very few results, and hence the system options for refining the search either will not be displayed or will be of no use to the researcher.

In all cases, the system displays a result list of zero or more items following the submission of a query. The researcher examines the list (the model's Scan action) and makes decisions about the next steps. If the query is of the Explore or Search type, the result list might be empty (if the query—the Articulated Information Need in the model—does not match materials stored in the system) or the results might seem irrelevant to the researcher. In such cases, the researcher may need to reformulate the query (Figure 77) or to select from the system's suggestions and submit a new query (the model's Reformulate action). If the result list is too long to be scanned thoroughly, the researcher is likely to narrow it down (the model's Narrow Down action).

Figure 77: Reformulating the query

Narrowing down can be done by disclosing more about the information need—adding more search terms to describe the need more specifically (Figure 78)—or by using the system's options (Figure 79).

Figure 78: Narrowing down the list by adding query terms

Figure 79: Narrowing down the list by using the system's options

The system provides faceted browsing as means to inform the researcher about the contents of the list and to enable the researcher to drill down to subsets of the list, thus narrowing it down. Narrowing down a result list is an iterative process during which the researcher can choose various attributes that describe the required item, such as the type of item, the topic, and the year of publication (Figure 80 A, B, and C).

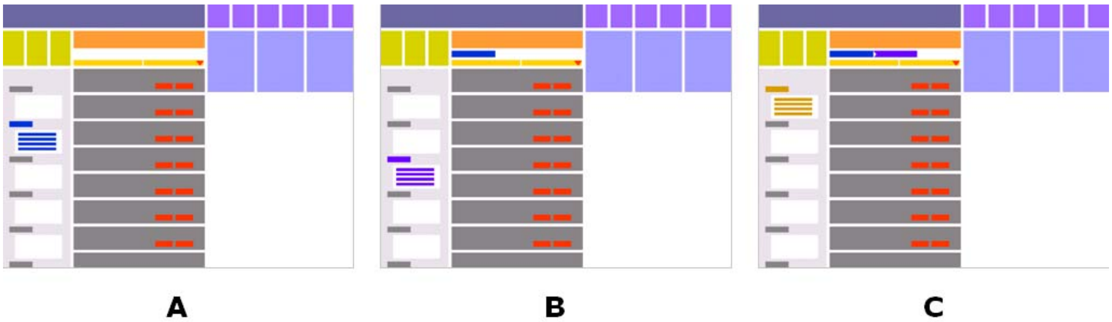


Figure 80: Faceted browsing

The narrowing down process ends when the researcher sees the required item on the screen or when the size of the result list is such that the researcher can examine it thoroughly (Figure 81). The researcher can expand the list at any point by deselecting any of the attributes.

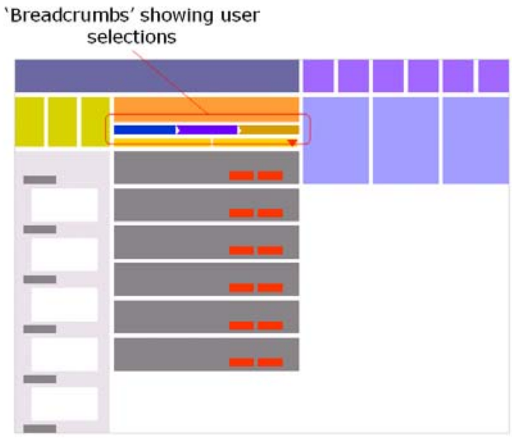


Figure 81: Result list that is narrowed down

At any time, the researcher can focus on an item of interest (Figure 82). The researcher can obtain a variety of services that relate to the item, such as an option to download the full text or the citation.

Figure 82: Focusing on an item

Once an item is in focus, the system displays lists of related items (Figure 83). Each list represents a specific relationship, such as items that cite the item in focus, items that are cited by the item in focus, and items that were looked at by other researchers along with the item in focus.

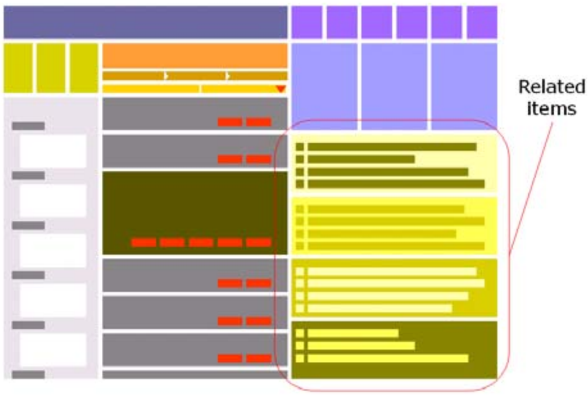


Figure 83: Lists of related items

The researcher can navigate to any related item (Figure 84). Such navigation corresponds to the Navigate action of the information-seeking model.

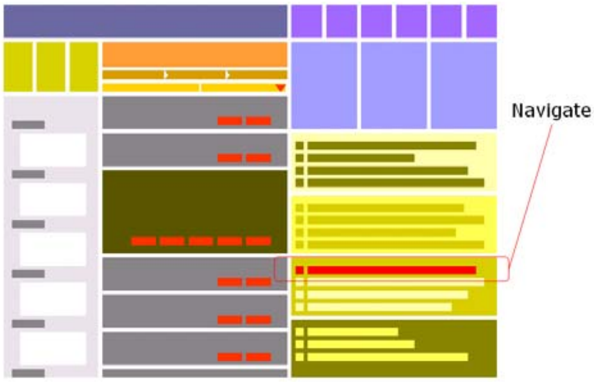


Figure 84: Navigating to related items

When the user navigates to a new item, that item becomes the item in focus and the system displays items related to this new one (Figure 85). The user can then continue navigating or can invoke a new query.

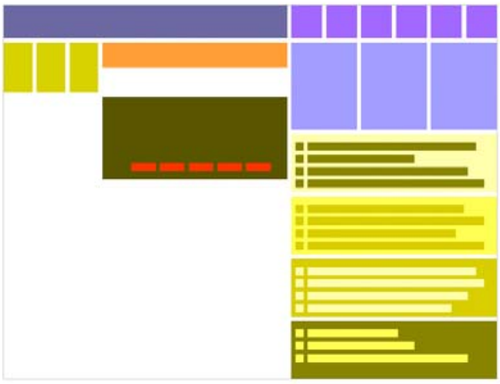


Figure 85: New item becoming the item in focus

As can be seen in these illustrations, all the actions and workflows that are described in the information-seeking and information-searching models are supported by the proposed user-interface design. The gap between the models and the user interface lies in the limited capability of INSPIRE to support actions (other than browsing) that relate to items that originate from other systems.

6.16 Evaluation of the Suggested User Interface

To assess the degree to which the suggested user interface would serve the HEP community, the author has defined a matrix whereby the personas described in Chapter 4 are used as hypothetical evaluators (Figure 86, Figure 87, and Figure 88). A list of system features was compiled, and the author used her best judgment[26] to rank the features on a scale of 1 to 5 according to the likelihood of a persona's using them (5, definitely; 4, very likely; 3, sometimes; 2, rarely; 1, never). The results are calculated for the more significant[27] subgroup of four personas (Ed, Laura, Kevin, and Guy) and for the whole group of six personas. The higher the score is, the more important it is for the developers of INSPIRE to offer the functionality or option.

As can be seen in Figure 86, Figure 87, and Figure 88, there are slight differences between the results obtained from the subgroup and those obtained from the whole group. The addition of the last two personas—which both represent younger people and experimentalists—skews the group evaluation toward a greater willingness to experience new options, work with user-interface elements that are familiar to users of popular Web services, and expand the search scope to other databases and other material types. On the other hand, the last two personas are more focused on immediate fulfilment of their exact information needs (typically newer articles and documents of a variety of other material types) and less inclined to broaden their knowledge horizons or to contribute to the community.

---

[26] The author relied on her deep familiarity with HEP researchers (following a long period of meetings, conversations, and e-mail exchanges with community members) and her ability to identify with community members through the use of the persona's characteristics.

[27] As explained in Chapter 4, user-interface design guidelines recommend the use of three to four personas. Therefore, a subgroup of four personas was defined, representing researchers who are more likely to use the system extensively and whose needs and expectations are of foremost importance in the development of the system. However, because comprehensive coverage of the types of HEP researchers can be accomplished only by considering all six personas together, referencing the subgroup of four and the entire group of six achieves both adherence to the guidelines and, as a means of further reassurance, attentiveness to a larger population's needs.

Predicted Relevance of New Services to Personas

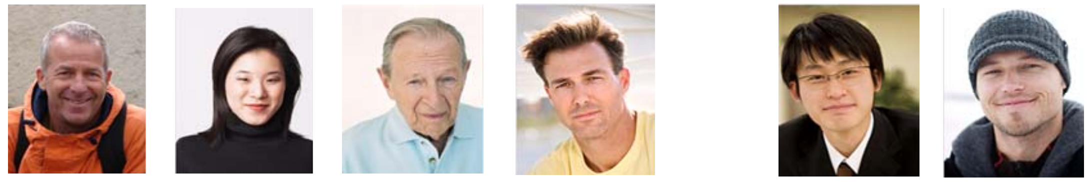





	Ed						Laura						Kevin						Guy							Hiro						George						
5:Definitely; 4: Very likely; 3: Sometimes; 2: Rarely; 1: Never.	5	4	3	2	1		5	4	3	2	1		5	4	3	2	1		5	4	3	2	1		Average for the first four	5	4	3	2	1		5	4	3	2	1		Total average (all six)
Use new INSPIRE interface		X				4	X					5				X		2	X					5	4.00	X					5	X					5	4.33
Set classic entry point as a default					X	1					X	1	X					5					X	1	2.00					X	1					X	1	1.67
Set personal preferences		X				4	X					5					X	1	X					5	3.75	X					5	X					5	4.17
Look for scholarly materials	X					5	X					5	X					5	X					5	5.00	X					5	X					5	5.00
Look for people		X				4	X					5		X				4		X				4	4.25	X					5			X			3	4.17
Look for institutions					X	1				X		2				X		2			X			3	2.00		X				4		X				4	2.67
Look for conferences					X	1		X				4					X	1				X		2	2.00		X				4				X		2	2.33
Look explicitly at the lists of values (e.g., journal titles)			X			3		X				4	X					5		X				4	4.00	X					5	X					5	4.33
Expand the search to use other databases		X				4			X			3				X		2		X				4	3.25	X					5	X					5	3.83
Search in parallel in other information systems such as Google, Google Scholar, and amazon.com		X				4			X			3					X	1	X					5	3.25		X				4	X					5	3.67
Use the facet value breadcrumbs		X				4	X					5				X		2		X				4	3.75	X					5	X					5	4.17
Use Did You Mean? feature	X					5	X					5	X					5	X					5	5.00	X					5	X					5	5.00
Sort by relevance			X			3			X			3					X	1	X					5	3.00	X					5	X					5	3.67
Sort by date (descending)		X				4	X					5	X					5	X					5	4.75	X					5	X					5	4.83
Sort by date (ascending)				X		2	X					5				X		2					X	1	2.50					X	1					X	1	2.00
Sort by popularity			X			3		X				4					X	1		X				4	3.00			X			3				X		2	2.83
Sort by number of citations	X					5	X					5	X					5			X			3	4.50	X					5			X			3	4.33

Figure 86: Matrix showing the likelihood of a persona to use a proposed feature, part 1

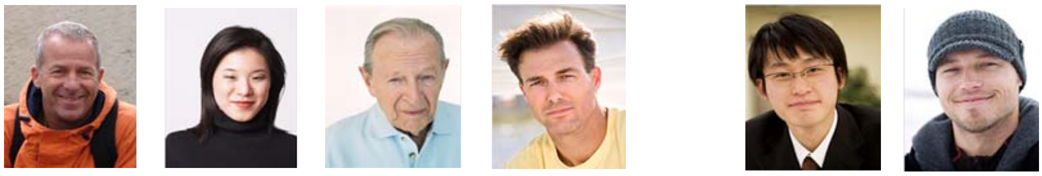
Predicted Relevance of New Services to Personas



	Ed						Laura						Kevin						Guy							Hiro						George						
5:Definitely; 4: Very likely; 3: Sometimes; 2: Rarely; 1: Never.	5	4	3	2	1		5	4	3	2	1		5	4	3	2	1		5	4	3	2	1		Average for the first four	5	4	3	2	1		5	4	3	2	1		Total average (all six)
Use faceted browsing			X			3	X					5			X			3	X					5	4.00	X					5	X					5	4.33
Narrow down to materials available online				X		2	X					5				X		2		X				4	3.25	X					5	X					5	3.83
Narrow down by date				X		2		X				4					X	1			X			3	2.50			X			3			X			3	2.67
Narrow down by material type		X				4	X					5				X		2		X				4	3.75	X					5	X					5	4.17
Narrow down by publisher					X	1					X	1			X			3					X	1	1.50					X	1					X	1	1.33
Narrow down by journal				X		2		X				4		X				4				X		2	3.00				X		2					X	1	2.50
Narrow down by institution				X		2			X			3			X			3					X	1	2.25				X		2					X	1	2.00
Narrow down by collaboration			X			3	X					5				X		2	X					5	3.75		X				4		X				4	3.83
Narrow down by topic				X		2	X					5				X		2		X				4	3.25		X				4	X					5	3.67
Change the default number of results per page					X	1				X		2					X	1	X					5	2.25			X			3		X				4	2.67
Select multiple citations on a page (or all citations on a page) for a specific service			X			3	X					5				X		2	X					5	3.75	X					5	X					5	4.17
Use suggested new searches		X				4		X				4			X			3	X					5	4.00	X					5		X				4	4.17
Read the abstract		X				4	X					5	X					5		X				4	4.50	X					5	X					5	4.67
Look for the query words in the displayed snippet or the abstract	X					5	X					5	X					5	X					5	5.00	X					5	X					5	5.00
Set one arXiv mirror site as a default	X					5	X					5	X					5	X					5	5.00	X					5	X					5	5.00
Change from the default arXiv mirror site when deafult is not responding	X					5	X					5		X				4	X					5	4.75	X					5	X					5	4.83

Figure 87: Matrix showing the likelihood of a persona to use a proposed feature, part 2

Predicted Relevance of New Services to Personas



	Ed						Laura						Kevin						Guy							Hiro						George						
5:Definitely; 4: Very likely; 3: Sometimes; 2: Rarely; 1: Never.	5	4	3	2	1		5	4	3	2	1		5	4	3	2	1		5	4	3	2	1		Average for the first four	5	4	3	2	1		5	4	3	2	1		Total average (all six)
Follow links to related information such as datasets, spreadsheets, images, video clips				X		2		X				4					X	1		X				4	2.75		X				4		X				4	3.17
Navigate through author name		X				4	X					5	X					5	X					5	4.75	X					5	X					5	4.83
Navigate through institution				X		2				X		2				X		2					X	1	1.75				X		2					X	1	1.67
Navigate through conference				X		2				X		2					X	1					X	1	1.50				X		2				X		2	1.67
Navigate through topic				X		2	X					5			X			3			X			3	3.25		X				4	X					5	3.67
Navigate through collaboration				X		2	X					5				X		2		X				4	3.25		X				4	X					5	3.67
Navigate through references				X		2	X					5			X			3				X		2	3.00	X					5			X			3	3.33
Navigate through citations	X					5	X					5	X					5	X					5	5.00	X					5	X					5	5.00
Navigate through tags				X		2		X				4					X	1		X				4	2.75		X				4			X			3	3.00
Navigate through 'people who looked at this item also looked at'			X			3		X				4				X		2	X					5	3.50			X			3	X					5	3.67
Look at explicit recommendations		X				4		X				4		X				4		X				4	4.00		X				4			X			3	3.83
Look at 'recent papers in my area'		X				4	X					5		X				4			X			3	4.00	X					5			X			3	4.00
Use more than one setting for recent papers		X				4			X			3				X		2				X		2	2.75		X				4			X			3	3.00
Look at my recent queries		X				4	X					5				X		2				X		2	3.25			X			3		X				4	3.33
Look at 'recent queries'				X		2					X	1				X		2				X		2	1.75					X	1					X	1	1.50
Look at 'most popular queries in the last week'			X			3				X		2				X		2				X		2	2.25				X		2				X		2	2.17
Rate an item					X	1					X	1					X	1				X		2	1.25					X	1					X	1	1.17
Review an item			X			3					X	1					X	1				X		2	1.75					X	1					X	1	1.50
Tag an item			X			3			X			3					X	1		X				4	2.75		X				4					X	1	2.67
Send a correction to INSPIRE team	X					5	X					5				X		2			X			3	3.75					X	1					X	1	2.83

Figure 88: Matrix showing the likelihood of a persona to use a proposed feature, part 3

As can be seen in Figure 86, Figure 87, and Figure 88, some of the features are not likely to be very relevant to some or all of the personas. These features are listed nevertheless to demonstrate that they were considered during the design phase.

By sorting the features by the average ranking score, the developers of INSPIRE can optimize their efforts and prioritize the components that they want to add to the system.

The features of the new user interface that are considered most likely to be useful for the researchers are those whose matrix score is at least 4.00 for either the subgroup of four personas or the entire group of six. The following features (some of which are already implemented in SPIRES) pass that threshold:

- A new user interface, if it is expected to be more user friendly than the current one
- Lists of possible values for fields such as material types and journal titles
- Did You Mean? suggestions
- Sorting by date
- Sorting by number of citations
- Faceted browsing
- Suggested new searches
- The display of the abstract
- Highlighting of query words
- Multiple selection of items for obtaining a service
- Setting one arXiv mirror site as a default and changing it only when necessary
- Navigation through author names
- Navigation through citations
- Explicit recommendations
- Lists of new articles

The following features, with a score of 3.00 to 3.99, are considered likely to be used:

- Setting personal preferences
- Searching seamlessly in other databases (if so defined by the researcher)

- Searching simultaneously in other information systems (with the results displayed separately)
- Facet value breadcrumbs
- Sorting by relevance
- Sorting by popularity
- Narrowing down results to materials available online
- Narrowing down results by material type
- Narrowing down results by journal
- Narrowing down results by collaboration
- Narrowing down results by topic
- Navigating through topics
- Navigating through collaborations
- Navigating through references
- Navigating through ‘people who looked at this item also looked at…’
- ‘My queries’
- Sending corrections to the INSPIRE team

6.17 Long-Term Monitoring of the New User Interface

The behaviour of the users of a new system is not always predictable because such behaviour is often influenced by factors that are beyond the scope of the individual system. For example, the appearance of Web search engines, particularly Google, had a considerable impact on the way users formulate queries in all information systems, including those used for academic research. The appearance of new mobile devices and the improved worldwide accessibility to the Internet is likely to further change the overall information behaviour and hence have an impact on information seeking in a scholarly environment.

Today, Web-based systems are regarded as ongoing projects, and the developers of such systems typically monitor the user behaviour and further develop and modify the systems as appropriate. This section discusses suggestions that will help the INSPIRE team gather information and analyse the researchers’ behaviour once they start using the system. Such analyses will serve as a basis for projected changes and further development.

It is important to note that the information gathering suggested in this section applies to INSPIRE alone and hence does not reflect the overall information-seeking behaviour of all researchers. Information gathered from all the information systems used by HEP scientists would have been more valuable for an analysis of the full spectrum of information seeking but gathering such information is not technically feasible at this stage; researchers use a variety of systems that do not enable the INSPIRE team to trace information behaviour. However, if, indeed, INSPIRE is developed to offer more content and services originating from other information systems, as discussed earlier in this chapter, the information that the INSPIRE team can gather will be more comprehensive.

6.17.1 Log Files

Designing log files from the outset provides an efficient, cost-effective, and comprehensive method of information gathering that, in turn, facilitates automated analyses of user behaviour.

The INSPIRE team should be able to extract from the INSPIRE log files data about sessions, queries, facets, search results, services, tiles, full-text availability, and user settings. In consideration of privacy issues, the information gathered should not be saved in a way that would associate it with individual researchers.

6.17.1.1 Sessions

Information about sessions enables the developers of a system to monitor how often people come to the system and how quickly they satisfy their information needs. The data that should be available for each session includes the following:

- Duration
- Number of pages viewed
- Number of queries submitted
- Number of items viewed (that is, how many times a user paid special attention to an item by requesting the full details of the item or requesting a service for it, for example)
- Number of items downloaded, including citations and the full text of articles
- Number of items saved or sent by e-mail

In addition, each session should be assigned an identification number (ID), to be used for logging other information (see sections 6.17.1.2 through 6.17.1.8). Having a session ID as part of the data gathered makes possible further analysis of the actions that a user performs during a session and supports an understanding of user workflows without having to monitor individuals over time.

6.17.1.2 Queries

The following information should be available about each query submitted to the INSPIRE search engine.[28]

- Query terms, as entered
- The type of interface used (basic or advanced). If the former is used, the data should indicate whether the query was Google like (e.g., *burton richter quark*) or SPIRES like (e.g., *find a richter, burton and t quark and date > 1984*).
- An indication of the origin of the query (for example, whether the query was entered by the user, invoked when the user clicked a link, generated as a refinement for a query, generated by a robot submitting a query, sent by a reference management tool as a bookmark, sent by Google, or sent by Google Scholar)
- Search scope

6.17.1.3 Facets

Although the information systems of the HEP community have never provided facets, researchers are most likely familiar with using them from commercial Web sites such as eBay and Amazon. Therefore, monitoring the usage of the facets is high priority, particularly to check which facets are more popular and are placed by the researchers at the top of the list, and which facets can be omitted. Information that should be available includes the following:

- Number of times that a facet was selected
- Number of facets used for a single query

[28] Technically, the INSPIRE search engine interprets a link to a record in INSPIRE as a query for a known item. Therefore, from the perspective of the search engine , there is no difference between a query that is formulated by a user, a hypertext link in INSPIRE, a bookmark stored on a personal page, and a hypertext link in a Web search engine, all of which lead to one or more records in INSPIRE.

6.17.1.4 Search Results

Monitoring the results of a query can give an indication of the type of query (Ask For, Search, or Explore; see 5.5) and the degree to which the sorting order corresponds to the user's needs. Together with other data, such as the type of search and the use of facets, the search results portray the effort that was required by researchers to satisfy their information needs. The following information should be available regarding result lists:

- o The number of results per query
- o The number of pages looked at
- o The position in the result list of the items selected by the user (first item on the list, second item on the list, and so on)

6.17.1.5 Services

Services should be easily modified, added to the INSPIRE system, or removed from it without disrupting the user workflow or requiring training for the researchers. Services should be monitored to see how useful they are. The logs should record the requests for each service.

6.17.1.6 Tiles

The use of customizable tiles enables researchers to create a personal information-seeking desktop. However, having too many tiles on a screen may lead to cluttered displays and decrease the usability of the new interface. Because not all researchers are likely to modify the default setup, it is of great importance to understand which tiles are more useful and modify the default display accordingly. To that end, the system should provide information about the number of researchers who selected a specific tile and how often they use that tile.

6.17.1.7 Full-text availability

Although not necessarily related to the user interface, the usage of the materials that INSPIRE makes available for searching should be monitored. Relevant information includes the following:

- o Number of times that the full text of articles was not available
- o Number of requests for full text from arXiv, a publisher's site, or institutional repositories
- o Number of times that full text was available through both arXiv and a publisher's site, and which source the user selected

6.17.1.8 User Settings

For further developing personalized services and for fine-tuning user default settings, the INSPIRE team should monitor the changes that individuals apply to the default settings. Information reflecting user preferences should include the following:

- The number of records per page
- The tiles that were selected to be part of a user's 'desktop' and by how many users each tile was selected
- The type of default search box (basic or advanced)

6.17.2 Survey

The 2007 HEP survey proved to be most instrumental in the understanding of HEP researchers' perceptions regarding their information systems and practices and in setting the design of the new INSPIRE system in motion. The relatively high rate of participation in the survey—about 10% of the HEP community members—demonstrates the high significance of the HEP information systems for the members. It is anticipated that launching another survey a year after the introduction of a new system will evoke similar participation by the community and generate feedback that can be used for improving the new system.

6.17.3 User Feedback

With the HEP community's commitment to improving their information system and content, it is anticipated that valuable feedback will be provided through feedback forms that the new interface will offer. Researchers are likely to write to the INSPIRE team and share their impressions.

6.17.4 Interviews

To acquire additional qualitative information that will help the INSPIRE team further develop and modify the system, revisiting the HEP community members who were interviewed as part of this study will be useful when these researchers start working with the new interface. Recording their feedback at several intervals is recommended to monitor the researchers' first impressions of the new interface, the speed at which they learn to use it, and, later, how rapidly they become accustomed to the new system and start taking advantage of new functionality.

6.18 Conclusions

On the basis of the HEP information-seeking model described in Chapter 5, a user-interface design is proposed and evaluated. The personas described in Chapter 4 are used for further evaluating the design, and methods of long-term quantitative and qualitative monitoring of the design's success are described. It is anticipated that the proposed user interface, which provides an information environment that supports the information-seeking practices of the HEP community in a friendly and efficient manner, will shorten the search process, improve the findability of quality materials, and thus support HEP academic research. Such an interface is expected to be applicable to other disciplines, as well, as discussed in Chapter 7.

Chapter 7 Conclusions and Future Research Directions

The research hypothesis that triggered this study emerged from the realization that scholarly searching is at a crossroads. As in the past, researchers are in constant need of information; however, with the increasingly rapid pace of publishing, immense quantities of information to explore, and many new types of materials to discover, obtain, and evaluate, researchers must use a variety of tools, some of which are outdated and others of which were designed for general information seeking. Information systems available today do not optimally address the information-seeking behaviour of scholars, particularly those who belong to scientific communities—which are the focus of this research; as a result, scholarly discovery is often cumbersome and incomplete.

The hypothesis of this study is that an information-seeking system that is designed to address the nature of scholarly materials and the information-seeking behaviour of scholars, particularly the members of one scientific community, will increase the effectiveness of the scholars' searches and enable them to find and obtain relevant materials with greater ease and precision than current practices do.

To test this hypothesis and fulfil the aim of this study, the author created an original model of researchers' information seeking, and on the basis of the model, suggested a software user interface that can be feasibly implemented by providers of scholarly information.

By carrying out a case study on the high-energy physics (HEP) research community, the author was able to gain an understanding of the information-seeking behaviour of its members. The information-seeking model and the proposed user interface were evaluated in view of the characteristics and needs of this scientific community and were found to satisfy the researchers' needs on a theoretical level, thus fulfilling the aim of this study. More empirical findings are required to ascertain the success of the user-interface design in helping the HEP researchers find and obtain relevant materials with the ease and precision expected from the model. However, such empirical findings will be available only after the suggested user-interface design is implemented. A follow-up study is suggested to examine these empirical findings.

The present study's objectives were addressed through the following activities:

- The information-seeking behaviour and search practices deployed by HEP researchers were examined through a series of interviews and observations (Chapter 3 and Appendix A).

- More than 2,100 responses obtained from a HEP survey that took place in the summer of 2007 were examined; in particular, the open-ended responses of these surveys were analysed (Chapter 3).

- On the basis of qualitative and quantitative research regarding the characteristics of HEP researchers and their information-seeking practices, a set of six personas, representing typical members of the HEP community, was developed (Chapter 4). These personas were used in the design and the evaluation of an information-seeking model, described in Chapter 5, and the user interface of an information system, described in Chapter 6.

- An original model was developed that leverages existing models of information behaviour, information seeking, and information searching and reflects the full spectrum of active information-seeking and information-searching practices of HEP scholars and the nature of the data that these researchers look for (Chapter 5). The model was evaluated through the use of seven scenarios involving the personas developed in Chapter 4.

- On the basis of the new information-seeking model described in Chapter 5, a software user interface was designed as the future interface for the HEP INSPIRE (the new version of SPIRES) information system. The user-interface design was corroborated through the model, and the personas described in Chapter 4 were used for to evaluate the design. Methods were suggested for long-term quantitative and qualitative monitoring of the way in which this design supports HEP researchers (Chapter 6). It is argued that the proposed user interface, which provides an information environment that accommodates the information-seeking practices of the HEP community in a friendly and efficient manner, will support HEP academic research by shortening the search process and improving the findability of quality materials.

This thesis contributes to the body of information-science knowledge in several areas:

- For the first time, to the author's knowledge, an information-seeking model is defined that addresses a specific scientific community that is highly focused on research and has distinct information-seeking patterns.
- The development of personas that represent scholars of the target community as an aid in the definition of an information-seeking model constitutes an innovative approach to modelling; such personas are usually applied in practical rather than theoretical contexts.
- The concretization of an information-seeking model into a new user-interface design that can be feasibly implemented provides a novel means of evaluating the theoretical model in an empirical context and of supporting a specific scientific community with its well-defined information needs.

The high-energy physics (HEP) community is the target research community of the study. However, the information-seeking model and the user-interface design derived from this model may serve other scientific communities as well. The HEP community, described in Chapter 3 (see particularly 3.10), has many characteristics in common with other scientific communities. However, the manner in which the HEP information-seeking practices have developed over time sets that community apart in various ways.

First and foremost, the scale of collaboration among members of the HEP community is unparalleled in any other scientific community. HEP projects carried out by thousands or even tens of thousands of scientists from all over the world, across institutional and national boundaries, have driven the development of methods to make publications written by community members readily available to other community members. The invention of the World Wide Web by Tim

Berners-Lee and the creation of an open repository—arXiv—by Paul Ginsparg, both HEP community members, are fundamental to the success of HEP research. Furthermore, the high publication rate has triggered the creation of tools, such as the new submissions service by arXiv, that enable researchers to keep abreast of developments in their area.

While other communities share some of these characteristics—for example, medical researchers face similar challenges, if not greater ones, in staying up to date—a major differentiator exists in the concentration of the HEP community's projects in a very small number of institutions and these institutions' acknowledgment of the importance of scholarly communication early on. The HEP scholarly communication infrastructure, which was put in place already in the 1960s, is supported entirely by the HEP institutions and therefore can continue to evolve in a way that corresponds to the community's needs. By establishing the communication channels by itself and not through external organizations, such as for-profit publishers, the community also developed a great awareness of the importance of open access to scholarly materials. Furthermore, the community's determination to use arXiv as a preprint repository had a significant influence on the issuing of its materials in a traditional manner by publishers: publishers acknowledge the originality of the materials written by HEP researchers despite the fact that researchers make these materials available through arXiv in parallel with submitting them for publication in a peer-reviewed journal.

The information-seeking model presented in this study is general in that it does not relate to a specific system; rather, it describes behaviour that characterizes scientific research. However, the availability of a means to concretise the behaviour cannot be taken for granted. Medical researchers, an example of another community that is relatively well supported by the members' institutions and by publishers, are likely to search for scholarly materials in PubMed, a database that can be regarded as a community-based initiative because it is created and maintained by the United States National Library of Medicine. These researchers can also define 'alerts', whereby they receive notification from PubMed when new materials that match their needs are published. However, the medical information landscape differs from that of high-energy physics: privacy issues, ethical concerns, and patents that involve the prospect of very large sums of money make the publication of medical-related materials slower and more restricted, leading to possible difficulties in obtaining materials at the point of need.

Other communities, such as that of geochemistry researchers, differ from the HEP community in various respects, such as the rate of new publications. Because other communities generate fewer new publications of interest than HEP scientists, their members do not rely on an alerting system such as those provided by arXiv and PubMed. Furthermore, no information systems have been created specifically for such communities. Whether the lack of better tools is the result of different information-seeking practices or vice versa is a topic that merits further research.

Further research is also needed to examine the information-seeking model proposed in this study in light of its relevance to non-HEP research communities.

Several meetings with the HEP team of information specialists—at the Deutsches Elektronen-Synchrotron (DESY) research centre in Hamburg, in May 2008, and at the European Organization for Nuclear Research (CERN) in Geneva, in February 2009 and September 2009—and e-mail exchanges with team members enabled the author to make a modest contribution to the creation of the INSPIRE information system that is replacing SPIRES. At the Hamburg meeting, the author presented her initial evaluation of the directions in which SPIRES needs to evolve (see https://indico.desy.de/conferenceOtherViews.py?confId=800). After analysing the open-ended questions of the survey conducted by the HEP information specialists, she presented her findings at CERN early in 2009, and later that year, she presented the proposed user-interface design to the same forum. At the time of this writing, INSPIRE is already in the beta-testing phase. However, at this point, the main difference between INSPIRE and SPIRES is that INSPIRE relies on a new technological infrastructure that has better performance (a much faster response time) and includes an improved search engine. Personalized features are expected in upcoming versions (http://www.projecthepinspire.net/). The user-interface design has not yet been changed.

The work that has already been done by the information specialists is likely to be welcomed by the HEP researchers who use the system. Anecdotal comments made to the author by HEP community members show that they like the improved response time and the fact that they do not have to remember the SPIRES-specific query syntax (even though they still use it for more complex queries). However, two major issues have not changed: the coverage—still pure HEP literature—and the conservative user-interface design. The changes that have already been implemented may prove to 'convince the convinced'—make the many avid users of SPIRES become avid users of INSPIRE—but may not attract new users. The team of HEP information specialists will need to conduct another survey to check the effects of the current changes on the HEP community.

However, in addition to the team's work and the suggestions provided in this study, other factors may change the information-seeking practices of HEP researchers in various ways, as described in the rest of this chapter.

The 2007 HEP survey showed that young researchers are more attracted to Google and Google Scholar than researchers who have spent a number of years in HEP research. However, a one-time survey does not provide a means to evaluate trends; this survey cannot help information specialists determine whether these young researchers will keep on relying on general information systems for the discovery of HEP content or will change their information-searching behaviour once they spend more time in the HEP environment and understand the benefits of using the information systems developed and maintained by the community. As discussed in Chapter 3 of this study, the involvement of the young generation of researchers in maintaining INSPIRE's accuracy and relevance to the community is crucial.

The information and communication behaviour of younger researchers—including but not limited to their information-seeking behaviour—is a clue to other types of changes that may occur in their HEP-related practices and that may have a substantial impact on the HEP information systems. The wide adoption of Internet communication and social networking tools such as Skype, Facebook,[29] and

[29] According to the Facebook statistics page (http://www.facebook.com/press/info.php?statistics), there are more than 500 million active users and 50% of them use Facebook daily. Furthermore, 'more than 30 billion pieces of content (web links, news stories, blog posts, notes, photo albums, etc.) [are] shared each month'.

LinkedIn may have an effect on the user-interface design of scholarly information systems, including HEP information systems. While some HEP researchers are expected to prefer keeping the user experience of their work-related environment separate from that of their social media environments (such a view was clearly evident in interviews conducted with researchers during this study), other researchers may want to have a unified user experience across all the systems that they use. The challenge of the HEP information specialists is to address both the approach of those who look for a contemporary user interface and the more conservative approach.

Social networks, primarily Facebook, may also change people's search expectations, in general, and assign people's social context a major role in the search process. One of the main messages of such networks is the context that they define for members: a person serves as a hub that is linked to other people and to the information they can provide; in other words, the social networks build a 'social graph' in which users serve as the graph nodes, but each user can be regarded as the centre of his or her own social graph.

Today, information generated by Facebook users is available only in the context of Facebook (it is not available to Google to index) and represents a more humanistic approach to information providing: in a manner that is similar to people's everyday behaviour—their reliance on the information that they acquire from people whom they know and trust—Facebook plans to offer search capabilities that leverage the context of the user to better satisfy the user's information needs. According to the vision of Mark Zuckerberg, the founder of Facebook, 'users will query this "social graph" to find a doctor, the best camera, or someone to hire—rather than tapping the cold mathematics of a Google search. It is a complete rethinking of how we navigate the online world, one that places Facebook right at the center. In other words, right where Google is now' (Wired 2009).

The Facebook approach to searching may have an impact on scholarly information seeking as well. It is not unrealistic to think of a community graph—primarily in the context of a well-defined group such as the HEP community—that will enable researchers to look for materials written or read by scholars whom they know and trust, starting from the scholars who are the closest to them and then branching out to include other community members. Although today scholars can implement a similar information-seeking approach by searching explicitly for materials written by specific authors, a change of the type that social networks are bringing about would be huge. If this change, indeed, happens, it may help researchers obtain relevant materials; on the other hand, however, it may obscure materials that are of high relevance but that originate from outside the expected domain.

In the context of HEP, a shift that renders researchers the pivot of the information environment may require a redesign of the environment: rather than offering content that is selected strictly on the basis of subject matter, the HEP information system would open up to offer materials written—or used—by members, including materials that are not HEP content or are of various new types. Information-seeking practices would likely change accordingly, and the information system would need to support these new practices and offer new services.

Other changes that may have impact on the user-interface design of scholarly information systems relate to technological developments. Mobile devices, still used primarily for general information seeking and exchange, may become convenient enough for scholarly research. The use of touch screens and methods for zooming in and out, selecting, and scrolling are likely to change the current design concepts.

There is no doubt that scholarly information behaviour is changing along with the general information behaviour of humans. Technological advances enable automated systems to be more context sensitive and, by identifying the location, affiliation, past behaviour, and preferences of the person who accesses the system, can tailor their services to that person. The distinction between human-human interaction and human-machine interaction is becoming blurred. The scholarly domain strives, on the one hand, to rely on assessment measures that are objective, but on the other hand, acknowledges the invaluable contribution of specific scholars to academic research; thus, the scholarly domain must leverage the developments that are likely to govern general information-seeking practices while retaining its academic focus.

Jamali and Nicolas note that ‘as information technologies…develop, information services are improved and as a result information seeking activities of scientists go though changes and adjustments. This is a cycle where research on information behaviour of scholars leads to better information services and improved information services might make the scholars alter their information seeking activities and behaviour, hence the need for continuous study of the information-seeking behaviour of scholars' (2008, 444). Indeed, much research is expected to take place in these areas in the coming years, and one can foresee that improvements will continue to be made in the ways in which information systems help scholars accomplish their tasks.

Appendix A: Interviews with High-Energy Physics (HEP) Researchers

Appendix A contains the summaries of the semi-structured interviews conducted with the HEP researchers in 2008 and the beginning of 2009. All the researchers work in the high-energy physics department of the physics faculty of a particular leading scientific research institution. Every interview lasted between an hour and an hour and a half and took place in the office of each HEP researcher. The author took notes during the interviews; there were no recordings. The interviewees were informed about the author's research and the objectives of the study as outlined in Chapter 1 of this thesis.

The author began all the interviews with the following questions:

- What is your academic status and role?
- What kind of research do you conduct, theoretical or experimental?
- How often do you look for scholarly materials?
- Which information resources do you use? How frequently do you use each of them?
- What do you use SPIRES for? How do you find the interface of SPIRES? What do you like and dislike about SPIRES?
- What do you use arXiv for? How do you find the interface of arXiv? What do you like and dislike about arXiv?
- What do you use Google for? What do you use Google Scholar for? What do you like and dislike about Google and Google Scholar?
- What services do you like in each of the information resources?
- In your opinion, how could SPIRES be improved?

In some cases, the interviewees were also asked to demonstrate the way they search.

Because the aim of the interviews was to understand the overall information-seeking behaviour of scholars rather than obtain specific details, the conversation was not firmly structured. The author let the interviewees take the lead in the discussion and raise other topics that they found relevant to HEP information seeking, such as their views on the use of datasets as part of the scholarly information corpus.

The interviewees were extremely cooperative. All commented that the work of the HEP information specialists is very important for any HEP-related research and thus were happy to assist in improving the HEP information systems.

After these interviews, the interaction with some of the researchers continued on a less formal basis in the form of e-mail correspondence and occasional meetings.

1. Alan

Alan is a senior professor, a theorist.

Alan is very fond of SPIRES. He has a link to SPIRES on his desktop and uses it every day. He also uses arXiv, but less than he uses SPIRES. The use of arXiv is typically related to new publications: every morning he looks at the new publications added to arXiv, and sometimes he looks for articles deposited in recent weeks, mainly after an absence of a week or so. He does not use arXiv for searching.

SPIRES, in his opinion, is a wonderful resource and covers all his needs. He considers himself an expert in searching SPIRES, and if he does not know a specific command (e.g., how to look for an item through the author's affiliation—he never uses this command) he uses the help page, which he finds clear.

When searching for an unknown item, Alan tries to narrow down the initial result list to about 100 hits—he can easily look at all of them. To narrow down a result list, he adds words to the query box, which is always displayed above the result list.

For example, Alan might start with a query in the following form:

`Find t yyyy a xxxx`

where *t* stands for *title* and *a* stands for *author name* (the SPIRES search syntax requires field indications). For example, the query *Find t supersymmetry a nir* would bring up all papers written by Nir that have *supersymmetry* in their title.

If there are too many results, he adds a term:

`Find t yyyy a xxxx t zzzzz`

Alan uses the SPIRES advanced search form (called Easy Search) only when he wants to enter the arXiv ID of an article.[30] He does not use other options of this search form. He often uses *exact author* (*ea*) rather than *author* (*a*) when there are several authors with the same surname—he learnt about this feature from colleagues. His searches typically include AND; sometimes he uses NOT—e.g., **`AND NOT collaboration`**.

Alan typically tries to find one article and then get to other articles by following links, either to articles cited by that article or to articles that cite that one. He says that this is the best way to conduct an exploratory search, in his experience.

Alan thinks that the citation index in SPIRES is extremely important and that the accuracy of this information and the related features offered by SPIRES are outstanding. According to him, some departments have a weekly meeting to keep track of the usage of their articles through the citation index of SPIRES. When he compares the citation index in SPIRES to that offered by Web of Knowledge (WoK), he finds the SPIRES service clearly superior. He uses WoK only when he has to find information about researchers dealing with other fields, and actually he wonders why other disciplines do not use the SPIRES infrastructure for their own purposes—but separately from SPIRES.

Alan does not see any problem with the current coverage of SPIRES. He does not encounter cases where he misses information in his research fields. He uses Google only when he needs to find information about other disciplines for purposes such as evaluating candidates from other areas.

[30] An arXiv ID is part of the metadata available for records stored in SPIRES.

Alan uses the arXiv service of new items, but he does not like to get the information by e-mail; he prefers going to arXiv and looking at the new items, grouped according to the three areas he deals with.

Another issue that Alan mentioned relates to open access repositories: he encourages open access and sends his papers to journals that offer open access rather than to those that are licensed.

Alan is a happy user of SPIRES and cannot think of a reason to change the current interface. He cannot think of anything missing in SPIRES—interface or content—and has no problem with the complexity of the interface. Nevertheless, Alan is interested in the INSPIRE initiative and will be glad to follow up and serve as a 'guinea pig' for testing any new ideas.

2. Benjamin

Benjamin is a senior professor, an experimentalist.

Benjamin says that he rarely looks for papers. When he does, he typically looks for specific ones. This happens once a week or less. He uses Google as his main resource and rarely goes to SPIRES. He is not sure what the difference is between SPIRES and arXiv, and when he looks for an article in Google and links to arXiv, he calls the process 'using arXiv'.

Benjamin thinks that Google is great. He also likes Amazon, especially the fact that Amazon seems to 'know' him and tailors services for him, e.g., recommendations. However, he thinks that these personalized features are more relevant to other areas of interest than to pure research. When asked to list the information systems that he uses in order of their relevance to him, he listed the following:

1. Google
2. Amazon
3. arXiv
4. Google Scholar

Benjamin does not know how to use the interface of SPIRES and finds it outdated, complex, and confusing. He does not use the arXiv interface, either. He does not know exactly how to find his own articles and says that the citation index in SPIRES is more relevant to theorists, who publish many papers, not to experimentalists like him. He finds the Easy Search form of SPIRES even more complex than the basic search.

Benjamin believes that SPIRES was the first system to implement a social network for the community, and he considers it a nice feature, but he does not believe that the information that is offered in SPIRES about researchers and research activities is of much importance. On the other hand, he would like the authors' pages to include pictures: he says that before he meets with someone, he typically looks for the person's picture so that he will not be embarrassed for not remembering the face of a person he has met before.

Benjamin thinks that alerts could be of much relevance to him—getting information based on specific criteria that he can define—but does not know how to set alerts in SPIRES or in arXiv.

When Benjamin searches in Google, he uses a few words from the title. He may add the surname of the author. Typically, he sees the relevant articles on the first result page, and if not, he adds query terms and searches again. He links to arXiv to read the articles and strongly dislikes being linked to a publisher site, mainly licensed ones. He believes all scientific information should be free.

Benjamin, despite the fact that he is not new in the HEP community, does not invest time to learn the interface of SPIRES, which is complex. He expects a Google-like interface. He would benefit from services such as alerts—these may convince him to use the SPIRES database. Otherwise he is likely to abandon it.

3. Colin

Colin is a senior professor, an experimentalist.

Listing the information systems that he uses, Colin gives arXiv first, as the most relevant resource for his needs. He uses SPIRES as well, and sometimes he uses Google.

According to Colin, SPIRES has no rivals when it comes to finding information about conferences, job and research opportunities, and people. He loves the feature that automatically generates the list of publications of a specific person (himself), although he would like it in another format so that he would be able to copy and paste without any further processing. However, when he looks for articles, he typically prefers arXiv, through a mirror site at a nearby University that provides a very fast response time.

Colin finds the search interface of SPIRES unfriendly. When he searches for articles, he looks for words in the title. According to him, the language used in HEP literature is very specific and standard; HEP researchers are forced to use the same terminology and stick to specific phrasing. Furthermore, the titles always reflect the subject of the article, so he does not think that a subject field in necessary. Colin uses SPIRES when looking for old articles: then he typically looks for the name of a person, and through the person he navigates to the list of publications, some of which are not available online, in which case he traces them in other repositories. He sometimes uses arXiv to search for authors.

When asked for his opinion about a possible extension of the search capabilities of SPIRES to cover abstracts and the full text of articles, he said that the system should first search the metadata, then the abstract, and only then search the full text. Typically the metadata and the abstract provide enough information about the article, but sometimes he remembers an idea that was expressed in an article but does not remember which article. In such cases, a full-text search is very useful.

Colin suggested a few enhancements to SPIRES. He likes the fact that people are represented in the database and would add more content to personal pages, including a large, clear photograph. Also, he suggests a synonym list specific for HEP. For example, he says that the terms *collision* and *interaction* are both used to describe the same phenomenon, although they are not regarded as synonyms in English. He expects the system to automatically look for alternative terms also or at least to suggest them. SPIRES should treat acronyms specific to HEP in a similar way—for example, the system should regard *SUSY* and *supersymmetry* as the same term.

Colin suggests that SPIRES add ratings by the community. According to him, the citation index of SPIRES is useless in the kind of research in which he is involved, where hundreds of scientists collaborate and every paper bears the names of them all. The fact that a name appears on an article is therefore meaningless. Another argument that Colin brings in favour of ratings is that they can be disconnected from the actual use of the article: a scholar can recommend or critique an article without having to write another article that cites the original one.

Regarding the coverage of SPIRES, Colin suggests that SPIRES add internal publications in a way that will keep them within the community but provide access to them through the same interface. He is referring to the following types of internal publications:

- 'Technical notes'—raw notes that have not been edited, checked, or reviewed
- 'Physics notes'—notes that have been reviewed internally
- Conference presentations—presentations that were given publicly and hence belong to the public domain
- 'Scientific notes'—notes that are reviewed and submitted for publication in journals

Colin does not think that providing access to datasets make sense. First, it is not 'fair' in the sense that the creator of a dataset has invested a great deal of time, effort, and money to create it, while others are spared all the bother and take advantage of a ready-made work. Second, and more important, he believes that experiments are very complex, and someone who does not know exactly how a dataset was created can jump to wrong conclusions that will require much effort on the part of the dataset creator to correct. Essentially, a dataset is not really usable to someone who does not know it well, and documenting everything about it for other people to use is not realistic. Furthermore, actual experience shows that when datasets were provided to others—such as the large electron–positron collider datasets that were offered in the public domain—hardly anyone looked at them.

Although Colin uses SPIRES, he reserves its use typically for services that he cannot get elsewhere. SPIRES is not at the top of his priority list. He is likely to use it more once SPIRES provides better services.

4. David

David is from another country, a 27-year-old postdoctoral theorist researcher who came to the institution for two years. His PhD was done partly at a university in his country and partly at CERN. The time he is spending as a postdoctoral researcher is devoted to realizing achievements that will pave his way to finding a good job. The competition is tough.

David arrived a few months ago and, in general, likes the people and the atmosphere at the institution. He thinks that the modest size of the department and the individual attention by faculty members is an advantage. On the other hand, it is not easy to be in a foreign country, away from family and friends. There are quite a number of people from abroad, and they are all looking for company; hence friendships are formed even though they cannot replace his 'real' friendships at home.

At this stage, David is looking for interesting topics that will enable him to come up with new findings or new directions. Eventually, he will summarize his research in a paper or a series of papers that he hopes will demonstrate that he is worthy of a position in a leading institution. He reads quite a lot and tries multiple directions, typically discussing his ideas with his advisor and other staff.

David understands very well what the difference is between a known-item search and an exploratory search. Most of his searches are for known items that were suggested to him by other people. He feels that looking at those items is much more effective than searching for materials by himself, because the people who recommend materials to him know the authors or the areas of research better than he does. Although he spends about 20% of his time looking for new materials, only very rarely does he perform an exploratory search.

David uses SPIRES for searching for known items and also for exploratory searches, although for the latter, he may consider Google as well (he does not know of Google Scholar). He greatly prefers the way SPIRES looks—the design is nice and clean, and it is easy to see titles, author names, institutions, and so on. He never had any training for using SPIRES, so he may be lacking knowledge about the functionality provided by SPIRES; he started using it after seeing others use it, because, according to him, it is the default information system for HEP.

David uses arXiv only as a source of new submissions. For searching, he always uses SPIRES. He likes the fact that SPIRES allows him to follow links from one article to another, mainly from an earlier article to a later one that cites the earlier one. The SPIRES citation index should be considered with some reservations, he believes. It is more important to know who cited the paper than the number of citations.

David does not like Google—first, because of the amount of 'noise' it brings up, and second, because of the design, which, according to him, is unfriendly: the information is not well formatted.

Although he has a Facebook profile, David is not a fan of social networks and uses them rarely. In general, he says, he is not very good at using gadgets and new tools.

5. Elena

Elena is a theorist, in her second year of her doctoral studies. She obtained a master's degree at the same institution and her bachelor's degree at another university.

Elena spends about half an hour a day searching for materials. She searches for articles most of the time; rarely does she search for information about conferences, organizations, or people.

Elena uses the following information systems:

- SPIRES, for all types of searches for articles
- arXiv, for monitoring new submissions
- Google, for finding information that is not pure HEP

Elena loves SPIRES and does not feel the need for any change in its interface or content. She estimates that 70% of her searches are for known items; the rest are exploratory searches. When she searches in SPIRES, she uses the Easy Search function because she does not remember the SPIRES syntax. She set Easy Search as her entry point into the system. She typically searches by keyword or author name; she rarely searches for a title or date. She occasionally uses the Top Cite option but says she uses it more out of curiosity than need.

Elena looks for an anchor—one relevant article—and relies on the navigation options in SPIRES or, even more so, in the full text of the 'anchor' article. She uses the SPIRES links to later citations of the papers and the references in the full text to find earlier papers. She likes finding a reference that includes a link or that is written in such a way that she can copy and paste the title into the search box of arXiv or SPIRES. She dislikes references that appear as a journal name, volume, issue, and so on because then she needs to reassemble the reference and put each subpart in another field of the Easy Search form of SPIRES in order to obtain the article.

To obtain a reference to an article in a specific format, she uses the relevant SPIRES service; however, sometimes the process becomes cumbersome: she needs to copy the arXiv ID, go to arXiv, find the article, link to SPIRES, and then get the appropriate format of the reference.

Occasionally, Elena looks for people. She often knows only the author name; by looking at the author's publications one by one she manages to find the right article. She is aware of the ambiguity of author names, but she typically looks at the longer list of publications, and only when she does not find what she is looking for does she try a variation of the name.

To find a conference, she uses a dedicated Web site that lists all conferences.

Elena does not like links to publishers' sites for obtaining the full text of articles; she always prefers the arXiv version. She uses arXiv every morning, scanning the daily list of new articles. If she is away for a while, she looks at recent articles as soon as she comes back. She also uses arXiv if she has the arXiv ID of a specific article; other than that, she never searches in arXiv.

6. Fred

Fred is 28 years old and in his first year of doctoral studies as a theorist. He obtained his master's degree in another university and is currently struggling to acquire a deeper familiarity with the research topics, the tools used, and the people at the faculty.

Fred uses SPIRES for searching and thinks that it is the best information system available in his field. Lacking familiarity with specific articles and with the terminology, he finds an item of interest in SPIRES and relies on the references and citations to obtain materials that are relevant to his research. He may use arXiv for searching, typically when he has the arXiv ID, but in most cases he uses SPIRES. Because he does not know the query syntax of either SPIRES or arXiv—he does not have the patience to learn it—he uses the Easy Search form in SPIRES and the advanced search form in arXiv, which do not require any knowledge of the query syntax.

Fred uses SPIRES also to find materials that are not found in arXiv, mainly theses: when looking for these, he adds the word *thesis* as a query word, and this typically 'does the trick', as he says.

Fred likes SPIRES very much because of the information it holds. However, because his field is closer to astrophysics, he sometimes discovers that the materials that he is looking for are not found in SPIRES, in which case he needs to use other information systems, including Google.

7. Gordon

Gordon is 31 years old and in his first year of doctoral studies as a theorist. He obtained his master's degree at the same institution and is familiar with the research topics, the tools, and the faculty members of the department.

Gordon searches mostly for known items and uses SPIRES for his searches. He likes SPIRES and thinks it is a great resource. Besides the content that he values, he likes the citation links, which he uses to follow ideas through the web of articles. He thinks that being able to see scientific work before it is published in a peer-reviewed journal is extremely valuable. Of course, he knows that the full text of preprints resides in arXiv, but because he can access the preprints by searching in SPIRES, he considers SPIRES the source for everything. Only when he cannot find something in SPIRES does he try another information system—typically Google or Google Scholar. This usually happens when he needs older materials. Another option for obtaining older materials is to go to the physical library.

Gordon uses arXiv for looking at new postings, but because he has so much to read, he looks at such postings on a weekly or even monthly basis—certainly not a daily basis. He does not go directly to arXiv but receives the list of new submissions by e-mail.

Gordon uses SPIRES for conducting exploratory searches. He tries keywords from the title along with author names and follows citations and references until he reaches the relevant materials. If he does not know the exact spelling of an author's name, he tries looking in Google for the name he has in mind, and then uses Google's suggestions to extract the correct name and search for it in SPIRES. He does not like Google, in general—Google displays too much 'noise', he thinks. SPIRES, from his perspective, is inclusive for his field.

Appendix B: List of Publications

Sadeh, T. (2010) Open products, open interfaces, and Ex Libris open-platform strategy, *Library Review* 59 (9), pp. 677-689. doi: 10.1108/00242531011087006.

Sadeh, T. (2008) Information-seeking behavior in the high-energy physics community. Paper presented at the HCI and Information Services Conference, Charles University, Prague, November 10, 2008. Available at http://uisk.ff.cuni.cz/detail.do?articleId=6662.

Sadeh, T. (2008) Multiple dimensions of search results. Paper presented at the Analogous Spaces Interdisciplinary Conference, Ghent University, Belgium, May 15-17, 2008.

Sadeh, T. (2008). User experience in the library: a case study, *New Library World*, 109 (1/2), pp. 7-24. doi: 10.1108/03074800810845976. Available at http://www.emeraldinsight.com/journals.htm?articleid=1642000.

Sadeh, T. (2007) Time for a change: new approaches for a new generation of library users, *New Library World* 108 (7/8), pp. 307-316. doi: 10.1108/03074800710763608. http://www.emeraldinsight.com/10.1108/03074800710763608. (The article received the Emerald Literati Club Award for Excellence 2008.)

Sadeh, T. (2007). Transforming the metasearch concept into a friendly user experience, *Internet Reference Services Quarterly* 12 (1-2), pp. 1-25.

Sadeh, T. (2007) User-centric solutions for scholarly research in the library, *LIBER Quarterly* 17 (3/4). Available at http://liber.library.uu.nl/publish/articles/000215/article.pdf.

Sadeh, T. (2007). User experience for library users: time for a change, *Password* 01/2007, pp. 20-23 (printed version).

Sadeh, T. (2006) Google Scholar versus metasearch systems, *High Energy Physics Libraries Webzine* 12 (February 2006). Available at http://library.web.cern.ch/library/Webzine/12/papers/1/. (The Informed Librarian Online selected the article as an Editor's Pick in May 2006.)

Sadeh, T., and Ellingsen, M. (2005) Electronic resource management systems: the need and the realization, *New Library World* 106 (5/6), pp. 208-218. doi: 10.1108/03074800510595823. Available at http://www.emeraldinsight.com/journals.htm?articleid=1501637&show=html. (The article received the Emerald Literati Club Award for Excellence 2006.)

Sadeh, T. (2004) Developing an electronic resource management system: Verde from Ex Libris. *LIBER Quarterly* 14 (3/4): 322-334.

Sadeh, T. (2004) The challenge of metasearching, *New Library World* 105 (1198/1199), pp. 104-112. doi: 10.1108/03074800410526721. Available at http://www.emeraldinsight.com/journals.htm?articleid=1509164.

Sadeh, T. (2004). To Google or not to Google: metasearch design in the quest for the ideal user experience. Paper presented at European Library Automation Group (ELAG) 2004 Conference: Interoperability: New Challenges and Solutions. Trondheim, Norway, June 9-11, 2004. Available at http://www.exlibrisgroup.com/resources/metalib/To_Google_or_not_to_Google.pdf.

Sadeh, T. (2003) Integrated access to hybrid information resources, *LIBER Quarterly* 13 (1-4), pp. 299-311.

Sadeh, T., and Walker, J. (2003) Library portals: toward the semantic Web, *New Library World* 104 (1/2), pp. 11-19. doi: 10.1108/03074800310458241. Available at http://www.emeraldinsight.com/journals.htm?articleid=860135. (The article received the Emerald Literati Club Award for Excellence 2004.)

Sadeh, T. (2001) MetaLib and SFX: managing heterogeneous resources in the scholarly environment. Paper presented at CASLIN 2001, Library of Academy of Sciences of Czech Republic and National Library of the Czech Republic. Beroun, Czech Republic, May 27-31, 2001. Available at http://www.caslin.cz:7777/caslin01/sbornik/metalib.html.

Sadeh, T. (2001) SFX and the OpenURL framework for the hybrid library. Paper presented at European Library Automation Group (ELAG) Conference 2001: Integrating heterogeneous resources. Prague, Czech Republic, June 2001. Available at http://old.stk.cz/elag2001/Papers/TamarSadeh/TamarSadeh.pdf.

Sadeh, T. (2001) Technologies and strategies for the interconnectivity of the hybrid library. Paper presented at the 10th Panhellenic Academic Libraries' Conference.

Thessaloniki, Greece, October 15-17, 2001. Available at http://www.lib.uom.gr/palc10/english/sadeh.doc.

References

Addis, L. (2002) Brief and biased history of preprint and database activities at the SLAC library, 1962-1994. Available at http://www.slac.stanford.edu/spires/papers/history.html [last accessed: 3 December 2010].

Aymar, R. (2008) Scholarly communication in high-energy physics: past, present and future innovations [Online] *CERN-OPEN-2008-015*. Available at http://cdsweb.cern.ch/record/1115073/files/CERN-OPEN-2008-015.pdf [last accessed: 3 December 2010]. (Published in 2009 in *European Review*, 17, pp. 33-51. doi: 10.1017/S1062798709000556.)

Bates, M.J. (1989) The design of browsing and berrypicking techniques for the online search interface, *Online Review*, 13(5), pp. 407-424. Version used for this study: http://gseis.ucla.edu/faculty/bates/berrypicking.html [last accessed: 3 December 2010].

Bates, M.J. (2002) Toward an integrated model of information seeking and searching. Paper presented at the 4th International Conference on Information Needs, Seeking and Use in Different Contexts, Lisbon, Portugal, September 11, 2002. *New Review of Information Behaviour Research*, 3, pp. 1-15. Version used for this study: http://gseis.ucla.edu/faculty/bates/articles/info_seeksearch-i-030329.html [last accessed: 3 December 2010].

Bawden, D. (2006) Users, user studies and human information behaviour: A three-decade perspective on Tom Wilson's 'On user studies and information needs'. *Journal of Documentation*, 62(6), pp. 671-679.

Bawden, D., and Robinson, L. (forthcoming) Individual differences in information-related behaviour: what do we know about information styles?, in Spink, A., and Heinström, J. (eds), *New Directions in Information Behaviour*, Bingley: Emerald.

Belkin, N.J., Cool, C., Stein, A., and Thiel, U. (1995) Cases, scripts, and information-seeking strategies: On the design of interactive information retrieval systems, *Expert Systems with Applications*, 9(3), pp. 379-395. Version used for this study: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.48.490 [last accessed: 3 December 2010].

Blomkvist, S. (2002) The user as a personality: Using personas as a tool for design. Position paper for the workshop 'Theoretical perspectives in Human-Computer Interaction' at the Interaction and Presentation Laboratory of the Royal Institute of Technology, Sweden, September 3, 2002. Available at http://www.nada.kth.se/~tessy/Blomkvist.pdf [last accessed: 3 December 2010].

Brown, C.M. (1999) Information seeking behavior of scientists in the electronic information age: Astronomers, chemists, mathematicians, and physicists. *Journal of the American Society for Information Science*, 50(10), pp. 929–943. Version used for this study: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.118.1322&rep=rep1&type=pdf [last accessed: 3 December 2010].

Centre for Information Behaviour and the Evaluation of Research (CIBER) (2008) Information behaviour of the researcher of the future: executive summary. A British Library and JISC study. Available at http://www.ucl.ac.uk/slais/research/ciber/downloads/ggexecutive.pdf [last accessed: 3 December 2010].

CERN (2008) *First beam in the LHC—accelerating science*. Available at http://press.web.cern.ch/press/PressReleases/Releases2008/PR08.08E.html [last accessed: 3 December 2010].

CERN (2010) *Particle physics INSPIREs information retrieval*. Available at http://cerncourier.com/cws/article/cern/42094 [last accessed: 3 December 2010].

Connaway, L.S., and Dickey, T.J. (2010) *The digital information seeker: Report of findings from selected OCLC, RIN and JISC user behaviour projects*. Available at http://www.jisc.ac.uk/media/documents/publications/reports/2010/digitalinformationseekerreport.pdf [last accessed: 3 December 2010].

Cooper, A. (1999) *The Inmates Are Running the Asylum*. Indianapolis, IA: SAMS/Macmillan.

Cornell University Library (2008) *Online scientific repository hits milestone*. Available at http://news.library.cornell.edu/content/online-scientific-repository-hits-milestone [last accessed: 3 December 2010].

Dantin, U. (2005) Application of personas in user interface design for educational software, in Young, A., and Tolhurst, D. (eds) *Proc. Seventh Australasian Computing Education Conference (ACE2005)*, Newcastle, Australia. CRPIT 42, ACS, pp. 239-247. Available at http://crpit.com/confpapers/CRPITV42Dantin.pdf [last accessed: 3 December 2010].

Dervin, B., and Nilan, M. (1986) Information needs and uses, *Annual Review of Information Science and Technology*, 21, pp. 3-33.

DESY (2008) High-energy physics labs join to build a new scientific information system. *Interactions.org* [Online] May 28, 2008. Available at http://www.interactions.org/cms/?pid=1026243 [last accessed: 3 December 2010].

Ellis, D. (1993) Modelling the information-seeking patterns of academic researchers: a grounded theory approach, *Library Quarterly,* 63(1), pp. 469-486.

Gentil-Beccot, A., Mele, S., Holtkamp, A., O'Connell, H.B., and Brooks, T.C. (2008) Information resources in high-energy physics: Surveying the present landscape and charting the future course, *arXiv:0804.2701* [Online]. Available at http://arxiv.org/abs/0804.2701 [last accessed: 3 December 2010]. (Published in 2009 in *Journal of the American Society for Information Science and Technology,* 60(1), pp. 150-160.)

Ginsparg, P. (2008) The global-village pioneers, *physicsworld.com* [Online]. Available at http://physicsworld.com/cws/article/print/35983 [last accessed: 3 December 2010].

Goldschmidt-Clermont, L. (2002) Communication patterns in high-energy physics. *High Energy Physics Libraries Webzine* [Online], March 2002, 6 (text of preprint written in 1965). Available at http://library.cern.ch/HEPLW/6/papers/1/ [last accessed: 3 December 2010].

Hagedorn, K., Chapman, S., and Newman, D. (2007) Enhancing search and browse using automated clustering of subject metadata, *D-Lib Magazine* [Online], 13(7/8). Available at http://www.dlib.org/dlib/july07/hagedorn/07hagedorn.html [last accessed: 3 December 2010].

Haines, L.L., Light, J., O'Malley, D., and Delwiche, F.A. (2010) Information-seeking behavior of basic science researchers: Implications for library services, *Journal of the Medical Library Association*, 98(1), pp. 73–81. Available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2801986/. [last accessed: 3 December 2010].

Hammond, S. (2006) Using psychometric tests, in Breakwell, G.M., Hammond, S., Fife-Schaw, C., and Smith, J.A. (*eds*) *Research Methods in Psychology*. London, UK: Sage Publications Ltd.

Hemminger, B.M., Lu, D., Vaughan, K.T.L., and Adams, S.J. (2007) Information seeking behavior of academic scientists, *Journal of the American Society for Information Science and Technology,* 58(14), pp. 2205–2225. Available at http://onlinelibrary.wiley.com/doi/10.1002/asi.20686/abstract [last accessed: 3 December 2010].

Hearst, M.A. (2006) Clustering versus faceted categories for information exploration, *Communications of the ACM*, 49(4), pp. 59-61. Available at http://people.ischool.berkeley.edu/~hearst/papers/cacm06.pdf. [last accessed: 3 December 2010].

Heuer, R.D., Holtkamp, A., and Mele, S. (2008) Innovation in scholarly communication: Vision and projects from high-energy physics, *Information Services and Use*, 28(2), pp. 83-96. doi: 10.3233/ISU-2008-0570. Available at http://arxiv.org/abs/0805.2739. [last accessed: 3 December 2010].

Hylegårde, J. (2009) Personality traits and group-based information behaviour: an exploratory study, *Information Research*, 14(2), paper 402. Available at http://InformationR.net/ir/14-2/paper402.html [last accessed: 16 March 2011].

Ingwersen, P., and Järvelin, K. (2005*) The Turn: Integration of Information Seeking and Retrieval in Context*. Dordrecht, the Netherlands: Springer.

Jamali, H. R., and Nicholas, D. (2008) Information-seeking behaviour of physicists and astronomers, *Aslib Proceedings*, 60(5), pp. 444-462. doi: 10.1108/00012530810908184. Version used for this study: http://eprints.rclis.org/16127/1/JAMALIi-FINAL-preprint.pdf. [last accessed: 3 December 2010].

Käki, M. (2005) Findex: Search result categories help users when document ranking fails, *CHI 2005 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* doi:10.1145/1054972.1054991. Version used for this study: http://courses.ischool.berkeley.edu/i247/f05/readings/Kaki_Findex_CHI05.pdf. [last accessed: 3 December 2010].

Kreitz, P.A., and Brooks, T.C. (2003) Subject access through community partnerships: A case study, *Science & Technology Libraries* [online preprint], 24(1-2), pp. 153-172. Available at http://arxiv.org/pdf/physics/0309027. [last accessed: 3 December 2010].

Kuhlthau, C.C. (1991) Inside the search process: Information seeking from the user's perspective, *Journal of the American Society for Information Science*, 42(5), pp. 361-371. Available at http://comminfo.rutgers.edu/~kuhlthau/docs/10.1.1.119.2997.pdf. [last accessed: 3 December 2010].

Marchionini, G. (1995) *Information seeking in electronic environments*. Cambridge, UK: Cambridge University Press.

Markey, K. (2007) Twenty-five years of end-user searching, part 1: Research findings, *Journal of the American Society for Information Science and Technology,* 58(8), pp. 1071–1081. doi: 10.1002/asi.20462. Available at http://deepblue.lib.umich.edu/bitstream/2027.42/56093/1/20462_ftp.pdf. [last accessed: 3 December 2010].

Markey, K. (2007) Twenty-five years of end-user searching, part 2: Future research directions, *Journal of the American Society for Information Science and Technology,* 58 (2007), pp. 1123-30. doi: 10.1002/asi.20601. Available at http://deepblue.lib.umich.edu/bitstream/2027.42/56094/1/20601_ftp.pdf. [last accessed: 3 December 2010].

Morville, P. (2005) *Ambient Findability*. Sebastopol, CA: O'Reilly.

Mulder, S. (2006) *The User Is Always Right: A Practical Guide to Creating and Using Personas for the Web*. With Ziv Yaar. Berkeley, CA: New Riders Press.

Murphy, J. (2003) Information-seeking habit of environmental scientists: A study of interdisciplinary scientists at the Environmental Protection Agency in Research Triangle Park, North Carolina, *Issues in Science & Technology Librarianship*, 38. Available at http://www.istl.org/03-summer/refereed.html. [last accessed: 3 December 2010].

OCLC Online Computer Library Center (2005) *Perceptions of Libraries and Information Resources: A Report to the OCLC Membership*. Dublin, Ohio. Available at http://www.oclc.org/reports/2005perceptions.htm. [last accessed: 3 December 2010].

OCLC Online Computer Library Center (2006). *College Students' Perceptions of Libraries and Information Resources: A Report to the OCLC Membership*. Dublin, Ohio. Available at http://www.oclc.org/reports/perceptionscollege.htm. [last accessed: 3 December 2010].

OCLC Online Computer Library Center (2007). *Sharing, Privacy and Trust in Our Networked World: A Report to the OCLC Membership*. Dublin, Ohio. Available at http://www.oclc.org/reports/sharing/default.htm. [last accessed: 3 December 2010].

Perfetti, C. (2002) Personas: Matching the design to the user's goals, *Intercom online*, Nov. 2002. Available at http://www.stc.org/intercom/PDFs/2002/200211_35.pdf [last accessed: 3 December 2010].

Research Information Network (RIN) (2006) *Researchers and discovery services: Behaviour, perceptions and needs.* Available at http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/researchers-and-discovery-services-behaviour-perc. [last accessed: 3 December 2010].

Robbins, L. (2007) SPIRES-HEP database: The mainstay of high-energy physics, *Issues in Science and Technology Librarianship* [online], 49. Available at http://www.istl.org/07-winter/electronic2.html. [last accessed: 3 December 2010].

Robertson, S. (2010) On queries and other messages. Paper presented at the 2nd International Workshop on Contextual Information Access, Seeking and Retrieval Evaluation, Milton Keynes, UK, March 28, 2010.

Randolph, G. (2004) Use-cases and personas: A case study in light-weight user interaction design for small development projects, *Information Science* 7, pp. 105-116. Available at http://www.inform.nu/Articles/Vol7/v7p105-116-175.pdf. [last accessed: 3 December 2010].

Sadeh, T. (2007a) Time for a change: New approaches for a new generation of library users, *New Library World* 108 (7/8), pp. 307-316. doi: 10.1108/03074800710763608. http://www.emeraldinsight.com/10.1108/03074800710763608. [last accessed: 3 December 2010].

Sadeh, T. (2007b) User-centric solutions for scholarly research in the library, *LIBER Quarterly* 17 (3/4). http://liber.library.uu.nl/publish/articles/000215/article.pdf. [last accessed: 3 December 2010].

Sadeh, T. (2007c) User experience for library users: Time for a change, *Password* 01/2007, pp. 20-23 (printed version).

Sadeh, T. (2008a) Information-seeking behavior in the high-energy physics community. Paper presented at the HCI and Information Services Conference, Charles University, Prague, November 10, 2008. http://uisk.ff.cuni.cz/detail.do?articleId=6662. [last accessed: 3 December 2010].

Sadeh, T. (2008b) Multiple dimensions of search results. Paper presented at the Analogous Spaces Interdisciplinary Conference, Ghent University, Belgium, May 15-17, 2008.

Sadeh, T. (2008c) User experience in the library: a case study. *New Library World* 109 (1/2), pp. 7-24. doi: 10.1108/03074800810845976. http://www.emeraldinsight.com/journals.htm?articleid=1642000. [last accessed: 3 December 2010].

Van de Sompel, H., and Beit-Arie, O. (2001) Open linking in the scholarly information environment using the OpenURL framework, *D-Lib Magazine.* 7(3). Available at http://www.dlib.org/dlib/march01/vandesompel/03vandesompel.html. [last accessed: 3 December 2010].

Vogelstein, F. (2007) Great Wall of Facebook: The social network's plan to dominate the Internet—and keep Google out, *Wired Magazine*. Available at http://www.wired.com/techbiz/it/magazine/17-07/ff_facebookwall. [last accessed: 3 December 2010].

Wilson, M., Schraefel, M.C., and White, R.W. (2009) Evaluating advanced search interfaces using established information-seeking models, *Journal of the American Society for Information Science and Technology* 60(7), pp. 1407-1422. Version used for this study: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.136.4180&rep=rep1&type=pdf. [last accessed: 3 December 2010].

Wilson, T.D. (1981) On user studies and information needs, *Journal of Librarianship*, 37(1), pp. 3-15. Version used for this study: http://informationr.net/tdw/publ/papers/1981infoneeds.html. [last accessed: 3 December 2010].

Wilson, T.D. (1999) Models in information behaviour research, *Journal of Documentation*, 55(3), pp. 249-270. Version used for this study: http://informationr.net/tdw/publ/papers/1999JDoc.html. [last accessed: 3 December 2010].

Wilson, T.D. (2000) Human information behaviour, *Informing Science*, 3(2), pp. 49-55. Special Issue on Information Science Research. Available at http://www.inform.nu/Articles/Vol3/v3n2p49-56.pdf/. [last accessed: 3 December 2010].