12
Toward a Web Search Information Behavior Model

S.A. Knight and A. Spink

Summary Information retrieval (IR) research in the context of the Web involves a number of complex processes. Some are user-related and include cognitive processes, motivational issues, information needs, technology attitude and adoption; and some are system related and include search engine algorithms and interface design. The field currently lacks a comprehensive model of Web interaction in the information behavior context. This chapter first explores a range of information behavior, and information seeking and retrieval model. Research relating to how users seek out and retrieve information in electronic environments will be examined and these models considered for applicability to the information environment of the Web. The exploration begins at the broadest level, examining information seeking models and then interactive IR models, followed by more recent integrated models. The paper then proposes macro model of Web-based information seeking and searching behavior. Further research areas are also discussed.

12.1 Introduction

Information retrieval entails the integration of a number of complex processes within the context of three major factors or entities:

- An information Need (Broder 2002)
- An information Searcher (Kuhlthau 1991)
- An information Environment (Johnson and Meischke 1993)

Not only does each of these entities possess unique characteristics depending on the situation, they also have a considerable influence on each other. This results in a substantial number of variables in regard to the users' information seeking or searching behavior and strategies. Information behavior differs from information seeking behavior[*1] (ISB). ISB represents one component of IB which can also include components such as the nature of the information, its specific context, format, or target audience, and other variables associated with its perceived usefulness or relevancy to the searcher, and searcher characteristics such as his or her cognitive level or efficacy. The term information-seeking behavior is at times

A. Spink and M. Zimmer (eds.), *Web Search, Springer Series in Information Science and Knowledge Management 14.*

mistakenly used in place of "information searching behavior", depending on the author or the system in which the user/searcher is looking for information. For example, within the context of an electronic environment, the action of seeking literally involves "search" strategies, so the seeking behavior is often described as "search behavior". This should not be confused with the term "information searching process" (ISP), which is generally used to specifically describe the cognitive processes involved in searching activities. Heinström (2000) suggests information behavior is best understood in the context of the information needs of the searcher, the inner, or cognitive, processes of the searcher, and the environmental factors relating to the information. These factors have an iterative effect on the searcher's way of responding to the information problem (Heinström, 2000).

From the decades of research into how users find and retrieve information has come a variety of proposed IB, information seeking, and searching behavior models. Wilson's (1981) notion of information need, their personality, and the environment in which they choose to look for the information are core variables that continually influence each other and the overall information seeking process.

Wilson used a framework that modeled information seeking from a "user studies" point of view. This view placed a heavy emphasis on how the user interacted with the information sought and found, rather than how the user interacted with the search system. Human computer interaction (HCI) research has typically concentrated on understanding how users feel about, interact with, and utilise technology, rather than the cognitive processes associated with the task for which they are employing that technology. This deficiency becomes particularly apparent when modeling the human/system interactive process of an activity that is largely cognitive, such as IR. Because of the noted influence of an "information environment to the information behavior of an individual searcher, the major developments in IB modeling will be considered within their historical context. Models will be compared with each other, in order to understand their influence on subsequent models, as well as to gain an understanding of the evolutionary nature of the ISB research discipline. This section will cover some of the major developments, culminating in a discussion relating to the integration of some of the common denominators into a preliminary framework of how searchers interact with Web search engines. The chapter is divided into two model types: information behavior in general and models that emphasis the interactive nature of IR and the role of system feedback in an electronic or online environment.

The historical context of the major IB model developments is closely aligned with two on-line technology revolutions. The first involved the creation of early online IR systems; used by "information professionals" who usually searched on behalf of the person who would ultimately use the found information. The second major development has been the advent of Web search engines, which have made available to any Web-user a practically immeasurable amount of information, with its own unique set of information characteristics. Research into IR, interactive IR and the resulting development of IB models has reflected this dramatic shift in both the end-user/searcher and the information environment.

12.2 Information Seeking Behavior Models

12.2.1 Wilson: Model of Information Behavior

Wilson's complex model (see Fig. 12.1) presented in 1981 and further amended in 1984, was a complicated framework that attempted to capture the information seeking process. The model included the three previously identified entities; namely (1) information user; (2) information need; and (3) information environment (see Fig. 12.1), and the iterative variables of successful (or non-successful) outcomes of specific searches, the possible involvement of other information users, and the ultimate satisfaction (or non-satisfaction) in information results or outcomes on the part of the searcher.

Central to Wilson's (1981) model was the information need – which was said to be framed by the users':

1. environment;
2. role; and
3. physiological, affective and cognitive needs. (see Fig. 12.2)

The information need was then said to influence a user's information seeking behavior, although not before it was tempered by any personal, interpersonal, and environmental barriers that the user might encounter.

Wilson's (1981) model lacked a clear description of how people interacted with an IR system in order to find and retrieve the data they sought. What Wilson labeled simply as "information seeking behavior" needed to be defined and explored

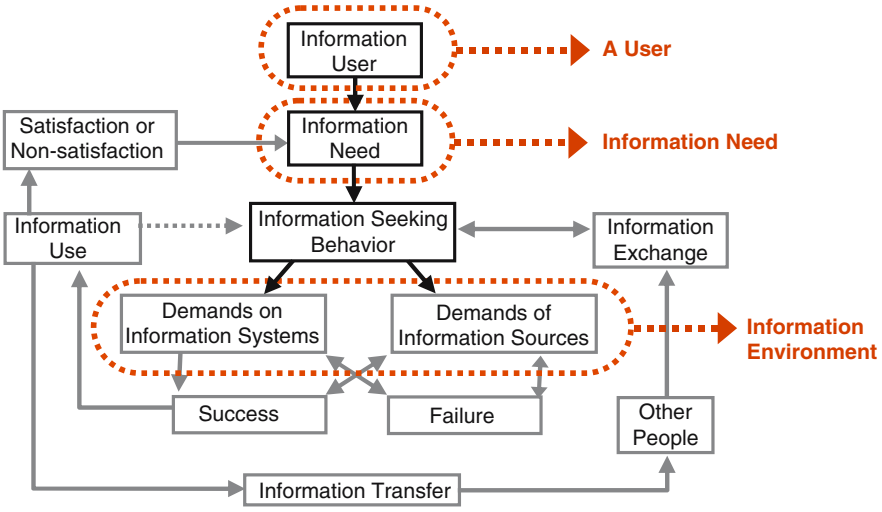

Fig. 12.1 Wilson's (1981) model of Information Behavior (emphasis, Knight 2006)

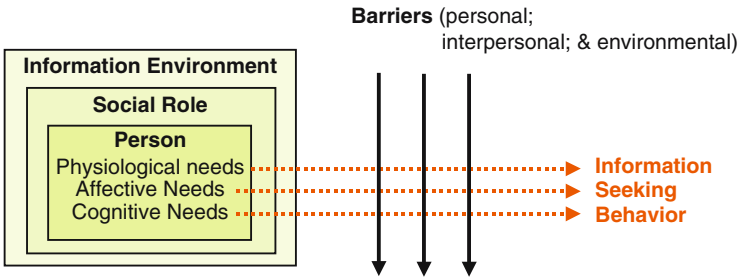

Fig. 12.2 Wilson's (1981) model of Information Seeking Behavior

further. Furthermore, a more extensive understanding of the "information systems" and "information sources" needed to be addressed in future models in order to better appreciate how the information environment – already acknowledged as a major influencing factor – impacted information seeking behavior.

12.2.2 Ellis: Behavioral Model for Information System Design

Ellis' (1989a; 1989b) research into information behavior produced a model describing six information seeking actions/strategies. The framework is illustrated and briefly described in Fig. 12.3, b.

The model was further refined with an additional two actions, verifying and ending (Ellis et al. 1993), and people's actions were described by Ellis as "features" rather than stages; indicating that the behaviors did not necessarily take place in a linear sequence, although clearly some behaviors were part of a sequence of behaviors (Fig. 12.4).

Ellis' framework was built on the observable behaviors and strategies employed by various sets of people (see Table 12.1). The extent of the description of the user's cognitive process related directly to the observable behavior being displayed by the user group in question. Although Ellis used a Grounded Theory methodological approach (Ellis 1989a) when building the model, subsequent testing of the framework using different user groups has produced similar results. It is worth noting that although the model evolves from time to time (see Table 12.1) its structure has remained largely unchanged.

12.2.3 Kuhlthau: Information-Seeking Model

Kuhlthau's (1991) approach was to model people's information seeking behavior in the context of assumed rather than observed cognitive processes. The resulting observable behaviors are not dissimilar in the two models, however Kuhlthau's presuppositions meant a framework could be developed that suggested there was a

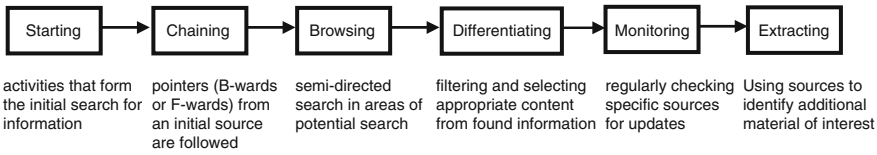

Fig. 12.3 Ellis' (1989a) Behavioral Model of Information System Design

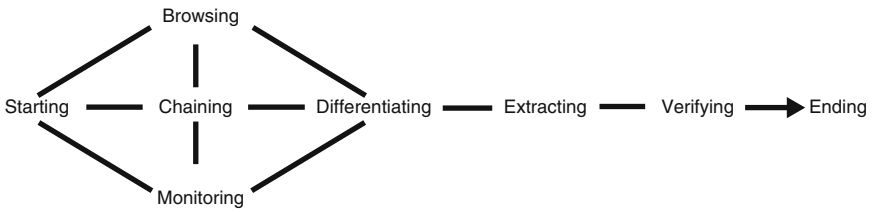

Fig. 12.4 Ellis' (1993) Behavioral Model for Information System Design

Table 12.1 Comparison of Ellis' Information Seeking Behavior Model (1989–1997)

Modelling Information Seeking Behavior (Ellis 1989a)	Info. Seeking patterns of Academic Researchers (Ellis et al. 1993)	Patterns of Engineers & Research Scientists in an Industrial Environment (Ellis and Haugan; 1997)
Starting	Starting	Surveying
Chaining	Chaining	Chaining
Browsing	Browsing	Browsing
Differentiating	Monitoring	Monitoring
Monitoring	Differentiating	Distinguishing
		Filtering
Extracting	Extracting	Extracting
	Verifying	Ending
	Ending	

logical sequence to all information seeking behavior. Each new experience is judged according to these self-made constructs, resulting in the continual reinforcement and/or development of those constructs. Kuhlthau describes an information search process (ISP) or information seeking process as a constructive activity in which the user attempts to find meaning from information (Kuhlthau 1991). The stages of Kuhlthau's model; the information seeker's feelings, thoughts, and actions; and the associated tasks are illustrated in Table 12.2.

Despite the different approaches to modeling user information seeking by Ellis (1989a) and Kuhlthau (1991), the similarities in their observed behaviors are quite remarkable (see Table 12.3), giving credence to Kuhlthau's hypothesis that there seems to be at least some information seeking strategies inbuilt into the human condition.

Table 12.2 Kuhlthau's (1991) Model

Stages		Initiation	Selection	Exploration	Formulation	Collection	Presentation
Human Experience Associated with stages	Affective (feelings)	Uncertainty	Optimism	Confusion/ Doubt	Clarity	Direction Confidence	Satisfaction or Disappointment
	Cognitive (thoughts)	General ~ Vague			Narrowed Clearer	Increased Interest	
	Physical (actions)	Seeking Background Information			Seeking Relevant Info		Focused Info
Tasks		Recognise	Identify	Investigate	Formulate	Gather	Complete

Table 12.3 Observed Information Seeking Stages/Behaviors in Ellis and Kuhlthau's Models

(Ellis (1989a)	Kuhlthau (1991)
Starting	Initiation
Chaining	Selection
Browsing	Exploration
Differentiating	Formulation
Monitoring	
Extracting	Collection
Verifying	Presentation
Ending	Ending

The weakness of both models remains their almost one-dimensional approach to the concept of the contextual variables of the observed information seeking behaviors. Ellis placed a heavy emphasis on the systems (electronic) environment context of the information being sought, while Kuhlthau concentrated on the user's cognitive predispositions towards information and learning. In contrast, Johnson suggests that a fundamental necessity of social action is that it must occur within a context (Johnson 2003), and then suggests that information seeking is a social action. Moreover, without a better understanding of the context of an information search, the information models produced lacked the flexibility to identify key components of the information environment that could trigger changes in an individual's information seeking.

12.2.4 Johnson and Meischke: Comprehensive Model of Information-Seeking

Johnson and Meischke (1991) recognised the influence of context in their research into how women diagnosed with breast cancer went about learning about their condition. They noted that an individual's seeking behavior varied depending on whether she was looking for information about breast cancer prevention, detection, treatment, or for information about dealing with the emotional issues involved with a diagnosis. They noted too that an individual's choice of information source (information-carrier factors) varied depending on the type of information required. By studying information seeking behavior within the context of that behavior, Johnson and Meischke (1993) were able to identify and validate:

1. The relationship between specific motivating factors and an individual's personal information need;
2. How the information need influenced choices relating to information environment ; and
3. The relationship(s) between information environment and individual information seeking behaviours

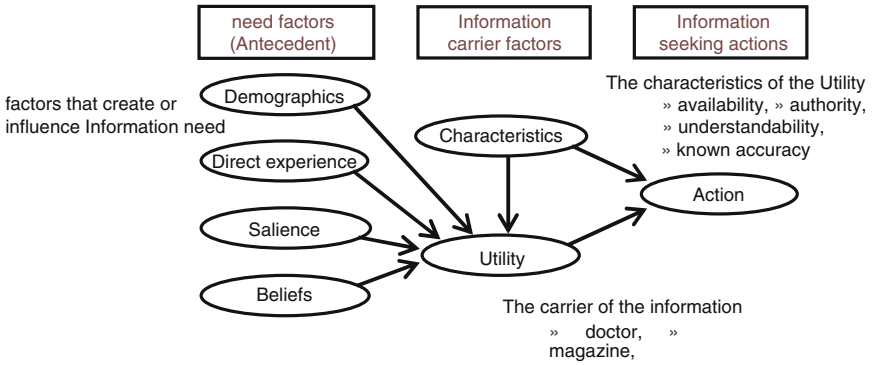

Fig. 12.5 Johnson and Meischke (1993) Comprehensive model of Information-Seeking

In the case of the initial CISM model (see Fig. 12.5) Johnson and Meischke (1993) that the information need (in this case, health-related factors relating to individual beliefs and experience of breast cancer) provided the motive for information seeking actions, which were shaped by information carrier factors. In reality however, the authors found that depending on the actual health-related factors; for example if an individual was not diagnosed with cancer, or they had never been exposed to issues relating to cancer, then the information carriers also played a motivating role in an individual's information seeking. Observations such as this can provide a significant insight regarding the impact of Web push and pull technologies, or how search engines can engage their user-base with "recommended links" or specific page relevancy algorithms.

12.3 Interactive Information Seeking Retrieval Models

The following set of models has been grouped together because of their emphasis on the dynamic interaction between the information need, searcher, and information environment. While interaction was probably always implied in previous models, its iterative affect on user search strategies, processes and outcomes was not always clearly defined.

12.3.1 Marchionini: Information Seeking in Electronic Environments Model

Like Kuhlthau, Marchionini's model (1995) is embedded in social cognitive and personal construct theories. Unlike Kuhlthau, whose primary focus was the affective and cognitive processes being experienced by individual information seekers, Marchionini took a more contextual approach, where the cognitive processes of the

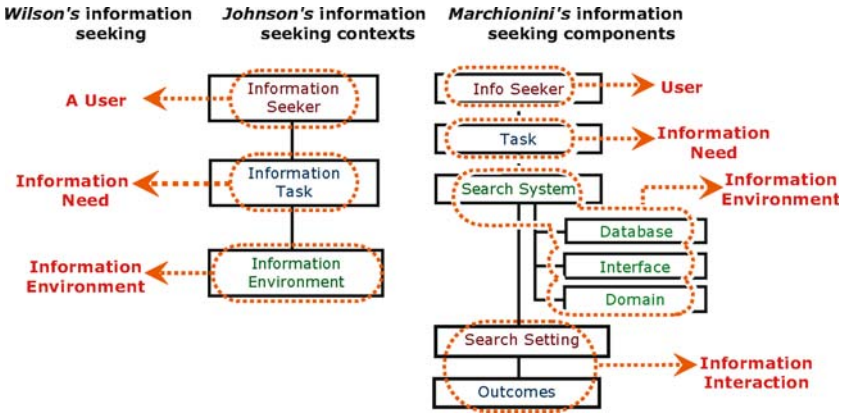

Fig. 12.6 The prominent role of the user, information need, & information environment paradigm in Wilson's (1981), Johnson and Meischke's (1993), and Brodei's (1995) information seeking models

searcher and the increasingly complex electronic information environment were considered within the scaffolding of their interactive relationship to each other. Central to Marchionini's model is the paradigm that information seeking is a natural and necessary mechanism of human existence (Marchionini 1995). It follows then, that in the context of a social science concept of human existence–seen as a series of interactions with the environment – that Marchionini defines information seeking fundamentally as an interactive process within an information environment. Understanding the information environment then, is as important as understanding the searchers cognitive processes, as it is the interaction between the two that establishes and reveals the actual information seeking strategies of the user.

Marchionini identifies eight information seeking components, which can be described as falling into four information entities (or contexts). These contexts are summarised and compared to previous information seeking model contexts in Fig. 12.6. The key difference between Marchionini's information seeking context and the previous information seeking contexts is that he adds a fourth context, namely, the interaction between the three previously considered key entities involved in information searching:

1. An information Need; (Bates 1989; Broder 2002)
2. An information Searcher; (Ellis 1989a; Kuhlthau 1991)
3. An information Environment (Johnson and Meischke 1993)
4. The various interactions between the entities of the searcher, the information need and environment (Marchionini 1995)

Marchionini's information seeking model – built on the contextual understanding developed from the information seeking contexts – is represented in Fig. 12.7.

The key supposition of Marchionini's model is that information seeking is a relatively linear process. Even with iteration taking place at the 'Reflect, iterate,

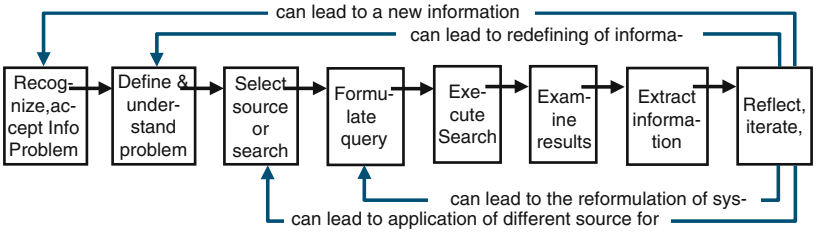

Fig. 12.7 Marchionini (1995) Information Seeking in Electronic Environments

stop' phase of the ISEE model, the implication is that the seeker is still looking and evaluating one information need at a time. The evaluation either leads to the identification of a whole new information need, or reveals possible problems in the search process, resulting in the searcher re-defining the information need, employing another electronic source, or simply formulating a new query. In reality though, information seeking and retrieval is often far more ambiguous than this. Browsing, and more specifically the concept of berry-picking (Bates 1989), is not discussed in Marchionini's model. In the early ′90′s the Web was still in its infancy, and virtually all participants used in prior research into IR and information search behavior still fell into the "information professional" category. These 'end-users' were, in fact, only end-users in the sense that they used the retrieval system. They were not the end-user of the information found. Moreover, they were end-users who had been specifically trained to use the systems, and so possessed a learned bias towards set strategies of searching online database systems. A second reason why Bates' model may not have been universally embraced by the early ′90's ISB research status quo was that it lacked the same degree of empirical testing as other models of its day.

12.3.2 Bates: Berrypicking Model

Bates' theoretical berry-picking model, first suggested as early as 1989 but never empirical validated, is that as an end-user searches, both the information sought and the user's choices regarding what is a relevant result evolves and changes (Bates 1989). Bates argued that the berry-picking model more closely represents the actual behavior of information searchers than previous traditional linear models in that it usually begins with one feature, topic or reference; and moves through a variety of sources, with new information encountered giving new ideas and directions to the original query. The berry-picking, evolving search model of IR is shown in Fig. 12.8. The model illustrates Bates' argument that the result(s) of each query provoke a cognitive response on the part of the searcher, which can either reinforce a search query, lead to expansion or variation of a query, cause a complete overhaul, or even abandonment of a query.

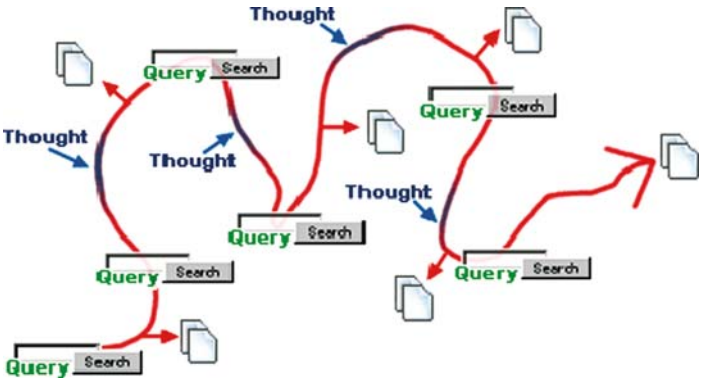

Fig. 12.8 A Berry-picking, evolving search (Bates 1989)

The four major differences noted by Bates between traditional information seeking models and the berry-picking model include, (1) The nature of the query; (2) The nature of the overall search process; (3) The range of search techniques used; and (4) The information domain (the specific data-driven environment) where the search is conducted. The fifth major difference between this model and previous models is that, implicit to the process of ISR is who will use the information. This type of evolving search can only really take place if the information searcher is also the information user, as the progression of the information sought and used is subject to the user making continual judgments regarding its relevancy and interoperability. The interactive nature of self-searchers' (end-users who were the information users) information seeking behavior became a primary focus of information behavior and IR models developed in the mid-1990's. These would become the foundation for models that would be applied to the Web.

12.3.3 Ingwersen: Cognitive IR Interaction Model

Ingwersen proposed that IR was a set of dynamic interactive processes, which occurred at multiple levels within the "cognitive space" of the user and the "information space" of the IR system. By using this poly-representation (1992; 1996) for information behavior, Ingwersen was able to at least begin to model an interactive process, said to occur not only between a user and the IR system, but also between the user and the information objects within the system with a more focused understanding of the actual information system being used, and the interactive cognitive processes that occur between the user and the system in order for information to be retrieved and ultimately used (Fig. 12.9).

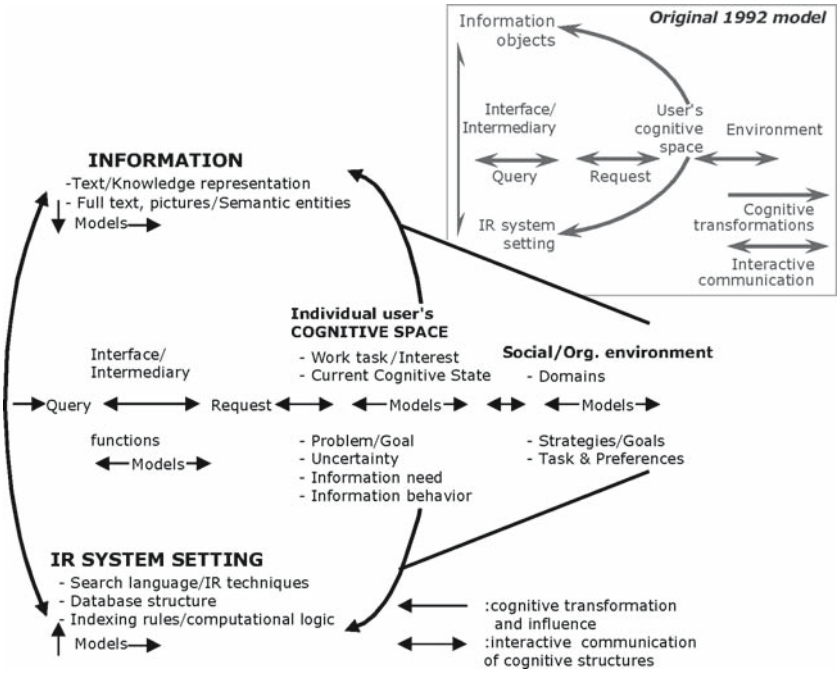

Fig. 12.9 Ingwersen's Cognitive Model of IR interaction (1992; 1996)

12.3.4 Saracevic: Stratified Interactive IR Model

The stratified interactive model (Saracevic 1996) of IR was based on an acquisition-cognition-application (A–C–A) type model of interaction. The model borrowed heavily (conceptually) from human computer interaction (HCI). The model is based on the assumption that users interact with IR systems in order to use information; that is, apply the information acquired through a cognitive process. Including "information use" as a part of the model was – like interaction – somewhat implied in previous models, but had not yet been explicitly positioned into the information seeking behavior models, perhaps because it can be safely assumed that a user would not take the time to specifically seek out information unless they were going to use it for something. Saracevic however, suggested that understanding the reason why a user sought out information was an important part of discerning the influencing factors on the interaction between the user, the IR system, and the information objects through the system.

In his stratified model, Saracevic (1996) proposed three levels, or strata, of IR interaction (Fig. 12.10):

1. A surface level of interaction – a sequence of events (interactions) between the user and the interface of the IR system.

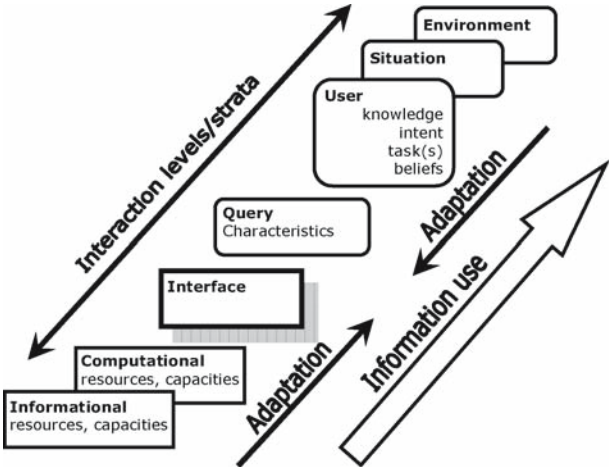

Fig. 12.10 Saracevic's Stratified Model of IR Interaction

2. A cognitive level of interaction – which identifies both the user's thinking and system's information objects as cognitive entities. At this level of interaction, the user is making judgments regarding the results (or feedback) given by the system.
3. A situational level of interaction – a context driven interaction, influenced by the original information need and how the user and/or system might categorize, or even iteratively change, the need.

The user's own pre-existing knowledge of the information, or the system, can influence the each of the levels of interaction, as well as any changes in strategies and categorizations of the information made, as the user chases the information being sought.

Saracevic acknowledged that elements within the three levels of interaction can, and in fact do, change as the process of IR is occurring. What, and how, those changes occur however, was not fully established in his model, as much of the research was still at the hypothesis stage. Empirical data was required, and needed to be analysed to establish the significant factors that influenced the interactive processes, so that that model could be tested. From the point of view of IR systems design, the strength of Saracevic's model is that it shifted the focus on IR from that of a static process to an interactive, and therefore highly dynamic one (Saracevic 1996), challenging system designers to re-consider the effectiveness of automated retrieval systems (Spink et al. 1997).

12.3.5 Spink: Search Process Model

As the importance of interaction became established in the research literature relating to ISB within a systems environment, authors began to question how the interactive

process actually took place. Until Spink's research in the mid-to-late 1990s, relatively little empirical research had been done that observed IR from an interactive point of view. Spink's search process model (1997) was developed from the hypothesis that a variety of feedback mechanisms were the major influencing factors in the interactive IR process, which involved such things as the user's "evaluation of the IR system output, user's judgments, and query modification" (Spink 1997). The empirical research undertaken by Spink set out to map the types and frequency of interactive feedback during mediated IR (Spink 1997). The goal was to identify user judgments, user search strategies and the interactive feedback loops within the search process. A major focus of the research was to understand the role of feedback in the interaction. Previous models had acknowledged feedback existed; mainly in relation to (1) user relevance judgments and (2) number of result (magnitude), however this research generally considered feedback to be somewhat linear, rather than an ongoing loop process.

Spink's research confirmed that these feedback mechanisms did in fact exist within the interactive IR search process, and proposed that a further three feedback mechanisms existed. The five different types of interactive feedback identified included;

1. *Content Relevance Feedback* (CRF) consisted of a query, followed by one or more relevance judgments, resulting in a modified or reformulated query.
2. *Term Relevance Feedback* (TRF) consisted of a user utilizing a term within the retrieved objects to modify any search strategies. Spink noted that this type of interaction occurred in 60% of observed online searches.
3. *Magnitude Feedback* (MF) consisted of user using the number of results to either broaden or refine the search for information. This type of interaction occurred in 45% of the observed online searches.
4. *Tactical Review Feedback* (TCF) consisted of users choosing to use strategy-related commands, such as the display sets (DS) command, to make judgments relating to the system's output, such as viewing a search history. Tactical review feedback only occurred in 7% of observed online searches, however it would have been interesting to note whether intermediary type searchers (information professionals) represented a higher proportion of this type of feedback, as it implies a familiarity with both the IR system and specific IR system strategies.
5. *Terminology Review Feedback* (TMR) is like the tactical review feedback, in that this strategy-related interaction involved the user requesting the display of terms in the inverted file. It occurred in only 1% of observed searches.

Importantly, the feedback mechanisms listed above did not occur as an either/or manifestation. As Fig. 12.11 illustrates, each search strategy could consist of more than one cycle of user-queries, that is ~ a user session/interaction with the system could consist of multiple feedback transactions, leading to additional inputs, or queries, which could in turn lead to different feedback and new inputs.

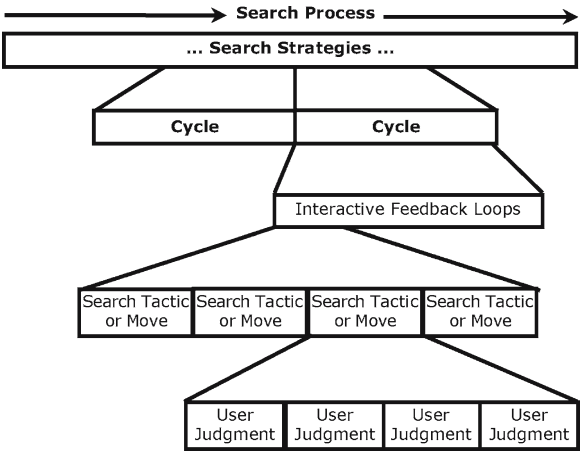

Fig. 12.11 Elements of the Interactive Search Process (Spink 1997)

12.4 Building a Web Interaction Model

The berry-picking (Bates 1989), cognitive (Ingwersen 1996), stratified (Saracevic 1997) and feedback process (Spink 1997) models provided a backdrop for the emerging "user" and "information environment " of the online IR systems of the early and mid 1990's. However, like the more linear models before them, they required a rethink and extensive testing before they could be applied to the emerging 'information environment' of the Web. Understanding the contextual makeup of IR on the Web is essential if researchers are to even begin to understand how users search and find information on/in the Web. The practical application of such research would include the design of appropriate Web search engine algorithms and interfaces, that better reflect (1) the cognitive processes of the typical Web information seeker (Spink and Jansen 2004). A big-picture focus also brings researchers back to the original supposition of information behavior models, that IR occurs in the context of an information need (or problem); an information searcher; and an information environment (Spink and Jansen 2004); and should always consider how these three contexts interact together (Marchionini 1995) in order to appreciate the extreme diversity of IR interactions.

Before the advent of the Web, the users of IR systems were largely "information professionals". These were made up of two types of individuals, those who were "intermediaries" – generally librarians who used online systems to search and retrieve information on behalf of a client who was ultimately the user of the information, and "educated professionals" – end-users who sought information directly connected with their work or profession (Ojala 1986). The enormous growth of the Web has provided an environment for a whole new user group with

a vast computational capacity to search for information. This new "end-user" is different from the previous online environment end-user in a number of ways:

1. They are not necessarily the "information professionals" of the previous generation of online searchers.
2. They are unlikely to have any formal training in developing appropriate search queries or retrieval strategies. In fact, the Web has introduced an entirely new generation of people – who have never even seen an IR system–to online IR (Brooks 2003).
3. They are likely to use a wider variety of search strategies, with more inconsistent results.
4. They are usually cognitively and physically on their own – unable to directly ask intermediaries or other users how to refine a query or improve a search result (Rieh 2004).
5. They are likely to be searching for a wider variety of information type and format.
6. They are more likely to be the "information-user" of the information they are seeking.

This change in end-user profile means that new dynamic variables of different user interactions have to be considered (Spink and Saracevic 1997): user cognitive ability, personality, information task, search outcomes, and PC capabilities. These all become important variables that can influence information search behavior (Hsieh-Yee 2001). The change in the "user" has been accompanied by a dramatic change in the on-line information environment. Web search engine environments differ from traditional online library information systems in a number of key areas:

1. Open architecture – resulting in no enforceable quality standards regarding the accuracy or quality of content.
2. Open classification and meta-tagging system – resulting in Web pages failing to be indexed appropriately by search engines (Doctorow 2001).
3. Highly dynamic use of the hypertext – favouring browsing over query - making in many instances.
4. Dynamic/fluid content structure – resulting in pages being "moved" within directories of a given Website, and frequent 404 errors (where pages no longer exist as formerly known URL's).
5. Partial representation – at any one time a Search Engine can literally only provide a "snap-shot" of the Internet at one given time in history. Servers that are offline or networks that have temporarily been interrupted cannot be "indexed" by a crawling search engine (Sullivan 2002).
6. Sheer volume – the sheer size of the Internet means that the snap-shot a search engine takes of the Internet at any one time is likely to represent less than 30% of the known Web.

Understanding how these users interact with this "utility" is the key to developing sound information behavior models and ultimately to building effective Web based IR systems. Initially, applying what had been learned from the years of research into

information seeking behavior in online environments seemed the logical step to understanding how users would retrieved desired information on the Web. However, early ISB studies that focused on traditional, managed, IR systems were unable to provide a rich picture of the interactions of IR on the Web (Wang et al. 2000).

In order to capture something of the heterogeneous nature of the Web, its wide variety of users and the context in which information is sought; research methodologies used in IR and ISB investigations are becoming increasingly qualitative (Martzoukou 2005). However, analysis of large data-sets (Broder 2002; Huberman et al. 1998; Spink and Jansen 2004) of user transactional data has also been applied in order to examine users' interactions with Web-based search engines. The second method (log analysis) has become more common (Spink and Jansen 2004). While analysis of keywords, results, search histories and user-logs provides an interesting picture of user actions and ultimate choices, they struggle to capture a user's cognitive processes involved with those choices. They also provide little user-related data regarding how users scan the content of Web pages or 'browse' (navigate) hypertext links. In other words, they demonstrate "how", but not "why".

Experiment-based or observational methodology will produce the most accurate results only if variables between the users' and their information interaction can be identified and accounted for or controlled. As a result, many studies relating to Web IR and seeking or searching behavior are conducted using small groups of similar users. Studies that have adopted this methodology include:

1. Navarro-Prieto et al. (1999) ~ Twenty-three University of Sussex students from the School of Cognitive and Computer Science (ten Computer Science, thirteen Psychology)
2. Hölscher and Strube (2000) ~ Twelve "expert" participants
3. Choo et al. (2000) ~ Thirty-four IT specialists, managers, and research/marketing/consulting staff from seven organisations
4. Lazonder et al. (2000) ~ Eight "expert" and seventeen "novice" participants
5. Saito and Mirva (2001) ~ Ten participants with similar knowledge and experience
6. Ford et al. (2001) ~ Sixty-nine masters students using the AltaVista for prescribed searches
7. Choo and Marton (2003) ~ Twenty four women IT professionals

12.4.1 Choo: Behavioral Model for the Web

An important aspect of IR on the Web relates to how users navigate (called browsing) the hypertext links of a Web page (including the dynamic page/results of a search engine query) in order to meet their information need.

In their behavioral model for the Web, Choo et al. (2000) propose a model of information seeking behavior to capture some of the browsing related information seeking strategies (called moves) employed by users.

Table 12.4 Information seeking behaviors and web moves

	Starting	Chaining	Browsing	Differentiating	Monitoring	Extracting
Literature Search Moves (Ellis et al. 1989a; 1993; 1997)	Identifying sources of interest	Following up references found in given material	Scanning tables of contents or headings	Assessing or restricting information according to their usefulness	Receiving regular reports or summaries from selected sources	Systematically working a source to identify material of interest
Anticipated Web Moves (Choo et al. 2000; 2003)	Identifying Web sites/ pages containing or pointing to information of interest	Following links on starting pages to other content related sites	Scanning top-level pages: lists, headings, site maps	Selecting useful pages and sites by bookmarking, printing, copying and pasting, etc.; Choosing differentiated, preselected site	Receiving site updates using e.g. push, agents, or profiles; Revisiting 'favorite' sites	Systematically searches a local site to extract information of interest at that site

Table 12.4 illustrates the "Web moves" identified by Choo, and their comparison to the "actions" of Ellis' behavioral model.

Any framework developed to investigate or present how users interact with and retrieve information on the Web must take both browsing type and query type behaviors into account. In doing this question relating to users' personalities and individual differences has become a key focus in much of the contemporary academic literature.

12.4.2 Ford, Millerand Moss: Individual User Differences

Ford et al. (2001, 2005) identified a number of key characteristic differences between users that affected search strategies and performance. These include such dimensions as (1) cognitive style (2) prior experience (3) Internet perceptions (4) gender, and (5) age. Information seeking behavior, and individual user and system differences were categorized into pre-existing theoretical models from multiple research disciplines. Figure 12.12 illustrates the theoretical framework in which Ford et al. (2001, 2005) examined the information seeking behavior of sixty-nine masters level students engaging the AltaVista search engine in a prescribed IR task.

In contextualizing the observed behaviors of users into pre-existing theoretical frameworks Ford et al. (2001) were able to develop initial findings regarding the

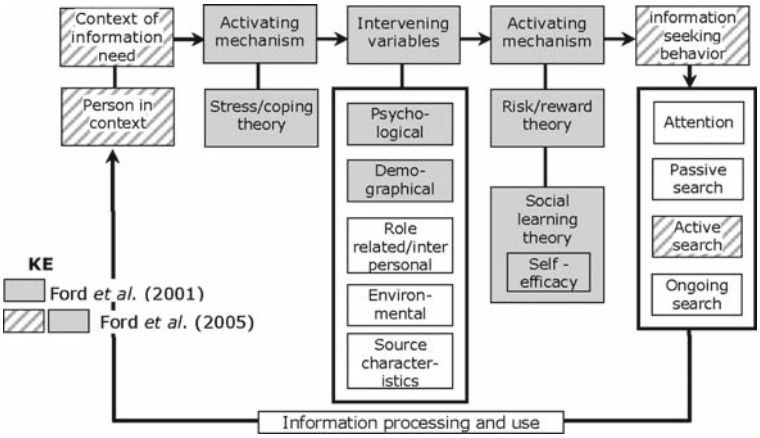

Fig. 12.12 Ford et al. (2001, 2005)

effect of identified individual differences in users on IR strategies and performance. For example, when examining Internet perceptions, it was found that poor IR performance was linked to perceptions "that the Internet is too unstructured, of not being in control, failing to keep on target, failing to find one's way around and getting lost" (Ford et al. 2001, p. 1060). A similar approach has been taken in the current research project, of which this paper is a literature review component. Pre-existing models such as the technology acceptance model (TAM) have been integrated into an interdisciplinary investigation of the impact of user perceptions of information quality on IR strategies.

Because the study investigates such cognitive processes as individual and/or groups of user perceptions, a hybrid methodology has been selected, using quantitative data collection strategies and qualitative analysis of the user results. The small sized user-groups employed in some previous qualitative studies of Web ISB – twelve 'expert' participants in Hölscher and Strube (2000), eight 'expert' and seventeen 'novice' participants in Lazonder et al. (2000), ten participants in Saito and Mirva (2001), and only five participants in Hale and Moss (1999a,b) – typically presented with limitations regarding generalisability of research findings. To address this issue, a minimum target of fifty participants was set when data collection started in March 2006. Data was collected over a thirteen month period (March 2006 to March 2007) to allow the user-group time to grow, with the final number of usable data-sets being eighty (from 123 registrations) when data collection stopped in March 2007.

For a participant's data to be considered "usable", a completed data-set of four on-line surveys had to be submitted in the specified order, and within a six-month time frame from a participant's submitted registration. With each survey designed as a stand-alone data-capturing tool however, users who only completed two or three of the four surveys have still provided valuable statistical data relating to

specific topics identified in the research. The data collected includes two technology acceptance model (TAM) surveys, incorporated to measure users' perceptions and expectations of their own ability to find information on the Web, as well as their perceptions and expectations of the actual information they retrieve and the Web's ability to provide relevant information. These perceptions are seen as a fundamental variable in the user's judgments (berry-picking model, Bates 1989) and user responses to the system feedback (search process model, Spink 1997) from the search engines they most often choose to use. Data collection also includes an ISB Strategies survey designed to map out typical user/search-engine interaction, and a final survey that establishes user perceptions of quality within the context of the specific types of information they look for on the Web. Figure 12.13 provides a framework to guide the theoretical structure of the current research.

The framework has been adjusted with descriptions of specific variables as they pertain to the current research. For example; "Role" is described as "Academic Role", representing one of the user-variables upon which four sub-classes within the user-group can be identified, and results compared. In this application then, Wilson's (1994) model doesn't so much describe expected user behaviors, but provides a theoretical backdrop where synergy between the various disciplines and parts of the investigation can be identified and used to better understand the user-group results, and therefore Web-based IR behaviors. We separate information seeking behavior into information seeking and searching behavior. While the authors agree, in principle, that this is true, a significant number of Web-users begin their interaction with the Web with "search" type behaviors such as a Web search engine query, and then shift to "seeking" type moves (Choo et al. 2000, 2003) such as scanning or browsing. Essentially, in an episode such as this, it could be inferred that information seeking becomes a sub-set of the information search process. For this reason, information seeking and search behavior are classified in the current research as different user information behaviors that users can iteratively swap between.

The interdisciplinary framework (Fig. 12.13) is being used to:

1. Identify multi-disciplinary theories that can be applied to better understand human information behaviors.
2. Contextualise how and where the various identified theories contribute to the process of data collection, comparison and analysis.
3. Help map-out patterns of information behavior of the user-group, and therefore identify if relationships exist between various data-sets.

The framework is not, therefore, a predictive model for human IR on the Web, although clearly there are some predictive elements associated with it. It is a tool used to map-out patterns of participants' information behavior within multi-disciplinary constructs helping to identify what (if any) types of relationships exist between participants' data-sets. The framework is used in conjunction with a proposed theoretically-based macro information behavior model, which will be discussed in the following section.

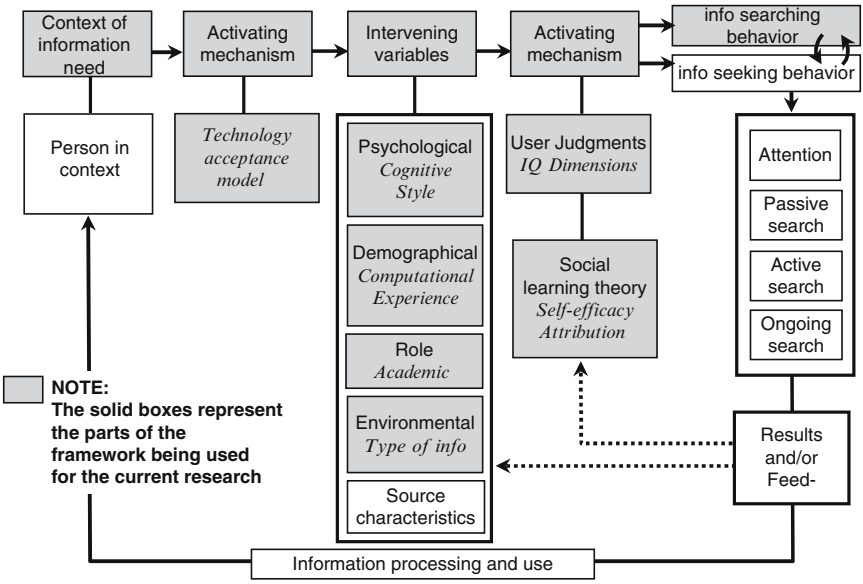

Fig. 12.13 An interdisciplinary framework for the current research project

12.4.3 Toward a Web IR Model

A theoretically-based, contextual, macro model for investigating Web-based information behavior is proposed (Fig. 12.14). The proposed model contends that user information behavior begins with an information need, which, if influenced by the user's cognitive style (Kim and Allen 2002), manifests itself in the use of specific information seeking or searching strategies. Cognitive style relates not so much to intellectual ability, but to preferred methods of operation on the part of the user. In the context of the current research, preferred modes of operation can be identified at a number of levels within the model. Research findings consistently advocate that a major influencing factor on user IR strategies is the user's pre-existing cognitive style (Ford et al. 2001; Kim 2000; Navarro-Prieto et al. 1999). In the proposed model, a user's cognitive style is seen as influencing their system-entry IR strategies, with users entering the IR process with a pre-existing preference to browse-seek (information seeking behavior) or search-seek (information searching behavior).

In this way, the two types of system interaction are classified as different sets of behavior, even though (1) there is likely to be common behaviors shared by each; and (2) users may periodically swap between the two behavior classifications. Unless a user already knows the URL of where they expect to find their target information, they are usually forced into a search-style strategy as their initial system interaction, regardless of their own cognitive preference. For this reason, user perceptions of self, the system, and expected interactions between their self and the system are also seen as having an influence on initial strategies. A better

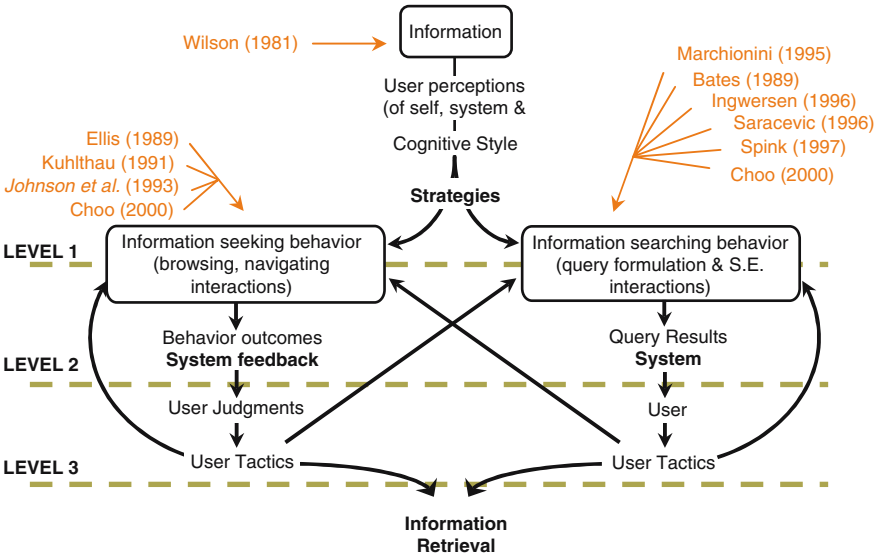

Fig. 12.14 A macro model of human IR behavior on the Web

understanding of the impact this forced step has on a user's (1) adoption of search engines, and (2) perception of the value of the search engine's results to their query is an expected outcome of the current research's data analysis.

12.5 Information Seeking and Searching Behavior

Within the next phase of the macro model (LEVEL 2) fall the many observable characteristics of previous models. Considering these models within the context of two types of IR behavior-sets allows for a measure of synergy between them not yet captured in the literature. Behavioral models such as Ellis' (1989a; Figs. 12.3 and 12.4) and Johnson and Meischke's (1993; Fig. 2.5) would fall predominantly into information seeking behavior, while the more query oriented interactive models of Spink (1997) and Bates (1989) would fall predominantly into information search behavior. The need to distinguish between information seeking and searching is recognised by researchers like Spink and Cole (2005), whose integrated information behavior model – a macro model – nests information searching behavior as a sub-set of information seeking behavior. While logically, "searching", that is, query formulation type information behaviors, is but one aspect of information seeking behavior. The problem with applying this concept to Web IR is that, more often than not, users experience the "search" and its associated tactics as their first information interaction with the system. For this reason, the current macro model seeks to classify information seeking and information searching behavior as alternative entry level strategies to IR on the Web.

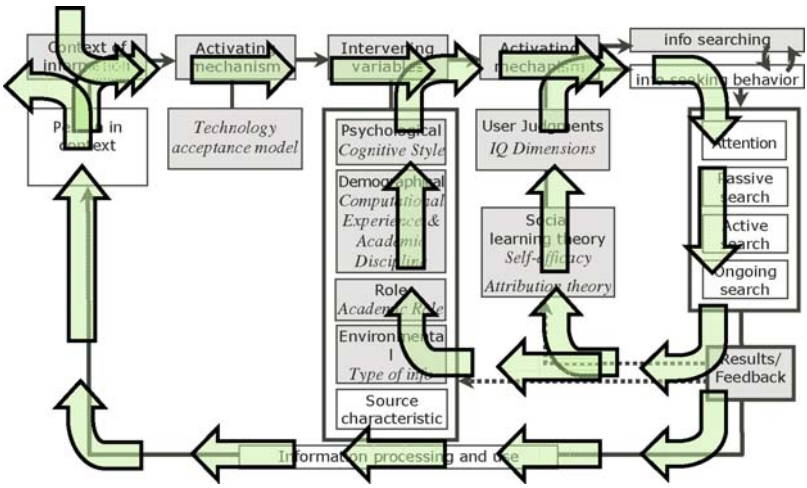

Fig. 12.15 The flow of IR (using the adapted interdisciplinary framework)

Between (LEVEL 2) and (LEVEL 3) of the macro model come any results of a user's information behavior. The result (described by some researchers as a system's "feedback") is used by the information seeker to make value judgments regarding (1) the system (search engine or Webpage) they have engaged; and (2) whether the content now presented to them will meet their information need. The value judgments of results are seen as being influenced by the intervening variables and activating mechanisms included in the adapted interdisciplinary framework (Fig. 12.13). IR is a highly iterative process, with the activating mechanisms and intervening variables imposing themselves into the IR process at any stage (see Fig. 12.15).

Generally speaking, intervening variables include such elements as users' cognitive style, level of system experience, knowledge of topic, and other "individual" characteristics associated with each user. In order for a researcher to make valid observations or develop meaningful theory in regards to those observations, a 'sample-group' of users must share a number of key intervening variables. If the user-group is large enough, then sub-groups who share different common variables can also provide a rich picture of the phenomenon being investigated. In the case of the current research, post-graduate level students and academics were identified as the target user-group. They were chosen specifically because it was assumed they would be high-end information users, and therefore posses (individually and collectively) discerning value judgments regarding the quality of any information they retrieve from the Web. The variables within the target group members that will assist in categorizing observed user information behaviors have been identified as; (1) Cognitive style; (2) Computational experience; (3) Academic discipline; (4) Academic role; and (5) Type of information most often sought. Other variables that could be investigated include age, gender, level of qualification, and geographic location. The two activating mechanisms that will be investigated most prominently

are the TAM (Davis, 1989) and an aggregate list of theoretically accepted information quality (IQ) dimensions developed from nineteen widely accepted IQ frameworks (Knight and Burn 2005; Knight 2007).

The users' response to system feedback is classified in the macro model as their tactics. It is assumed that at the broadest level, the tactics chosen by users are most directly influenced by the value judgments made of the system feedback. However, because of the feedback/loop nature of the model, user tactics will see the user return to behaviors associated with information seeking, or searching; or if the user is satisfied with the content presented to them, IR. Therefore, the tactics stage is one of the stages where users may swap between or stay within classified sets of behaviors. It is hoped that by examining users' changes in behavior within the context of the activating mechanisms and intervening variables of the interdisciplinary framework, that a better understanding of why users make specific information behavior choices can developed. In essence, research into information seeking behaviors is an attempt to understand how user's link language/ communication construct to meaning. This is particularly pertinent given that the act of research itself is also an attempt to find meaning. To that end, the terminologies used to describe the various human information behaviors are themselves imposed constructs developed by researchers to help contextualize and understand the behavior being examined.

12.6 Conclusion

This chapter has provided a move towards a comprehensive model of Web interaction. Such a model needs to include the motivating human aspect (information need) that begins any Web search episode and its close connection to the cognitive make up of the individual (a searcher), and the setting (information environment) in which the individual attempts to fulfill their need. The adding of a fourth required element (to the three required elements already named) in order for Web IR to take place is the actual interaction between the user and the Web system environment, and the user and the information. Virtually all the research covered adheres to these four basic required elements of information seeking behavior, without one of them, Web IR cannot take place.

References

Bates MJ (1989) The design of browsing and berrypicking techniques for the online search interface. On-line Review 13: 407–431

Broder A (2002) A taxonomy of web search. ACM SIGIR Forum 36: 3–10

Brooks TA (2003) Web search: how the web has changed information retrieval. Information Research 8(3)

Choo CW, Marton C (2003) Information seeking on the web by women in IT professions. Internet Research 13: 267–280

Choo CW, Detlor B, Turnbull D (2000) Information seeking on the web: an integrated model of browsing and searching. First Monday 5.
Davis FD (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 13: 319
Doctorow C (2001) Metacrap: putting the torch to seven straw-men of the meta-utopia. http://www.well.com/~doctorow/metacrap.htm
Ellis D (1989a) A behavioral model for information retrieval system design. Journal of Information Science 15: 237–247
Ellis D (1989b) A behavioral approach to information retrieval system design. Journal of Documentation 45: 171–212
Ellis D, Cox D, Hall K (1993) A comparison of the information seeking patterns of researchers in the physical and social sciences. Journal of Documentation 49: 356–369
Ford N, Miller D, Moss N (2001) The role of individual differences in Internet searching: an empirical study. Journal of the American Society for Information Science and Technology 52.
Ford N, Miller D, Moss N (2005) Web search strategies and human individual differences: cognitive and demographic factors, Internet attitudes and approaches. Journal of the American Society for Information Science and Technology 56: 741–756
Heinström J (2000) The impact of personality and approaches to learning on information behavior. Information Research 5: 78
Hölscher C, Strube G (2000) Web search behavior of Internet experts and newbies. Proceedings of the 9th Conference on World Wide Web, pp. 81–101
Huberman B, Pirolli P, Pitkow J, Lukose R (1998) Strong regularities in world wide web surfing. Science 280: 94–97
Hsieh-Yee I (2001) Research on web search behavior. Library and Information Science Research 23: 167–185
Ingwersen P (1992) Information retrieval interaction. Taylor Graham, London
Ingwersen P (1996) Cognitive perspectives of information retrieval interaction: elements of a cognitive ir theory. Journal of Documentation 52: 3–50
Johnson JD (2003) On context of information seeking. Information Processing and Management 39: 735–760
Johnson JD, Meischke H (1991) Women's preferences for cancer information: from specific communication channels. American Behavioral Scientist 34: 742–755
Johnson JD, Meischke H (1993) A comprehensive model of cancer-related information seeking applied to magazines. Human Communications Research 19: 343–367
Kim KS (2000) Individual differences and information retrieval: implications on web design. Proceedings of RIAO.
Kim K, Allen B (2002) Cognitive and task influences on web searching behavior. Journal of the American Society for Information Science and Technology 53: 109–119
Knight SA (2007) The impact of user perceptions of information quality on world wide web information retrieval studies. Doctoral dissertation, School of MIS, Edith Cowan University
Knight SA, Burn JM (2005) Developing a framework for assessing information quality on the world wide web. Informing Science Journal 18: 159–172
Kuhlthau CC (1991) Inside the search process: information seeking from the user's perspective. Journal of the American Society for Information Science and Technology 42: 361–371
Lazonder AW, Biemans JA, Wopereis IG (2000) Differences between novice and experienced users in searching information on the world wide web. Journal of the American Society for Information Science 51: 576–581
Marchionini G (1995) Information seeking in electronic environments. Cambridge Series on human computer interaction. Cambridge University Press, Cambridge, MA
Martzoukou K (2005) A review of web information seeking research: considerations of method and foci of interest. Information Research 10: 215
Moss N, Hale G (1999b) Cognitive style and its effect on Internet searching: a quantitative investigation. ECER 99: European Conference on Educational Research, 1189

Navarro-Prieto R, Scaife M, Rogers Y (1999) Cognitive strategies in web searching. HFWEB'99: Conference on Human Factors and the Web, July 5, 1999, Washington

Ojala M (1986) Views on end-user searching. Journal of the American Society for Information Science 37: 197–203

Rieh SY (2004) On the web at home: information seeking and web searching in the home environment. Journal of the American Society for Information Science and Technology 55: 743–753

Saito H, Miwa K (2001) A cognitive study of information seeking processes in the WWW: the effects of searcher's knowledge and experience. In M.T. Özsu, H.J. Schek, K. Tanaka, Y. Zhang, Y. Kambayashi (Eds.) Proceedings of the 2nd International Conference on Web Information Systems Engineering, Kyoto, Japan (pp. 321–333)

Saracevic T (1996) Modeling interaction in information retrieval (IR): a review and proposal. Proceedings of the 59th Annual Meeting of the American Society for Information Science 33: 3–9

Spink A (1997) Study of interactive feedback during mediated information retrieval. Journal of the American Society for Information Science and Technology 48: 382–394

Spink A, Cole CB (2006) Human information behavior: integrating diverse approaches and information use. Journal of the American Society for Information Science and Technology 57: 25–35

Spink A, Jansen BJ (2004) Web search: public searching of the web. Springer

Spink A, Saracevic T (1997) Interactive information retrieval: sources and effectiveness of search terms during mediated online searching. Journal of the American Society for Information Science 48: 741–761

Sullivan D (2002) Nielsen//NetRatings search engine ratings. Search Engine Watch. http://www.searchenginewatch.com/reports/netratings.html (20 March 2003)

Wang P, Hawk WB, Tenopir C (2000) Users' interaction with world wide web resources: an exploratory study using a holistic approach. Information Processing and Management 36: 229–251

Wilson TD (1981) On user studies and information needs. Journal of Documentation 37: 3–15

Wilson TD (1994) Information needs and uses: fifty years of progress? In B.C. Vickery (Ed.) Fifty years of information progress. Journal of Documentation Review (pp. 15–51) aslib, London