Human-Computer Interaction View on Information Retrieval Evaluation

Tiziana Catarci[1] and Stephen Kimani[2]

[1] Sapienza, University of Rome, Via Ariosto 25, 00185 Rome, Italy
catarci@dis.uniroma1.it
[2] Jomo Kenyatta University (JKUAT), P.O. Box 62000, 00200, Nairobi, Kenya
stephenkimani@gmail.com

Abstract. The field of information retrieval (IR) has experienced tremendous growth over the years. Researchers have however identified Human-Computer Interaction (HCI) aspects as important concerns in IR research. Incorporation of HCI techniques in IR can ensure that IR systems intended for human users are developed and evaluated in a way that is consistent with and reflects the needs of those users. The traditional methods of evaluating IR systems have for a long period been largely concerned with system-oriented measurements such as precision and recall, but not on the usability aspects of the IR system. There also are no well-established evaluation approaches for studying users and their interactions with IR systems. This chapter describes the role and place of HCI toward supporting and appropriating the evaluation of IR systems.

Keywords: Information Retrieval, Human-Computer Interaction, Evaluation.

1 Introduction

1.1 Motivation for HCI in IR

The field of information retrieval (IR) has experienced tremendous growth over the years. Researchers have however identified Human-Computer Interaction aspects as important concerns in IR research [16]. For instance: IR system design, evaluation, and the study of users' information search behaviours and interactions. Allen [2] indicates that there is a need to establish a link between research within IR and the design of user interfaces. According to the ACM Special Interest Group on Human-Computer Interaction (SIGCHI), Human-Computer Interaction (HCI) is a discipline concerned with the design, evaluation, and implementation of interactive computing systems, and the study of major phenomena surrounding them [37]. Marchionini [56] points out three developments that make it important to incorporate HCI in IR:

- Information Retrieval (IR) and Human-Computer Interaction (HCI) are related fields having strong traditions that have been challenged and energized by the World Wide Web.

M. Agosti et al. (Eds.): PROMISE Winter School 2012, LNCS 7757, pp. 48–75, 2013.

- The type and nature of content have evolved and changed e.g. type of content has moved beyond text to include statistics, multimedia, computer code, sensor streams and biochemical sequences.
- The type and nature of users have evolved. Data has become increasingly accessible to a large number of users with no or minimal training in information retrieval e.g. on the Internet through mobile devices, TVs, etc.

According to Marchionini [56], the foregoing three developments lead to the concept referred to as Human-Computer Information Retrieval (HCIR), whereby "*we think of information interaction from the perspective of an active human with information needs, information skills, powerful digital library resources (that include other humans) situated in global and local connected communities – all of which evolve over time.*" Marchionini [56] goes on to argue that the concept suggests systems that are characterized by:

- Systems should aim to get people closer to the information they need, especially to the meaning; that is, systems can no longer only deliver the relevant documents, but must also provide facilities for making meaning with those documents.
- Systems should increase user responsibility as well as control; that is, information systems require human intellectual effort, and good effort is rewarded.
- Systems should have flexible architectures so they may evolve and adapt to increasingly more demanding and knowledgeable installed bases of users over time.
- Systems should aim to be part of information ecology of personal and shared memories and tools rather than discrete standalone services.
- Systems should support the entire information life cycle (from creation to preservation) rather than only the dissemination or use phase.
- Systems should support tuning by end users and especially by information professionals who add value to information resources.
- Systems should be engaging and fun to use.

Incorporation of HCI techniques in IR can ensure that IR systems intended for human users are developed and evaluated in a way that is consistent with and reflects the needs of those users [57].

1.2 Motivation for HCI in IR Evaluation

The study of IR systems has prescribed and dominant evaluation methods that can be traced back to the work by Cleverdon [17]. The traditional methods of evaluating IR systems have over a long period of time been mainly concerned with system-oriented measurements such as precision and recall, but not on the usability aspects of the user interface such as how well users can accomplish their goals and tasks, interactive, and cognitive issues. There are no well-established evaluation approaches for studying users and their interactions with information retrieval systems [33], [48].

1.3 User Interface Techniques for IR

During the user interface design process, the primary focus is on who the users are and what the tasks are. The main role of the system is to support user in their tasks. A task could be some activity that involves achieving a particular goal or purpose. In general, the user interface of an IR system has the role of guiding, supporting and transforming user's information problems, goals or needs [33]. The user interface can be described as the elements that the user comes into contact with when using a computing system. According to Hix and Hartson [38], the user interface generally comprises two parts: the interaction part and the interface software part. The interaction part or the interaction component is concerned with how the user interface works and its behaviour in response to what the user does while performing a task. The interface software part is concerned with the implementation of the interaction component. In the sequel, is a description of various existing user interface techniques for supporting users to interact with and use information retrieval systems.

Query Formulation and Query Reformulation

Many search engines expect the user to formulate an initial information request in a manner understandable by the underlying search engine. The user interface for such search engines typically accepts the information request (i.e. query) in form of a keyword-based statement. Users often need to reformulate their query after an initial query has been executed. Most search engines support query reformulation features such as: spelling suggestions, spelling corrections, and automatic query term reformulation [36]. Such features support the user by suggesting potential search directions and paths that can yield results that are relevant to the user. The features strive to put control of selection and interpretation of results in the user's hands.

Browsing

Traditional user interfaces of information retrieval systems have been geared toward analytical searching rather than browsing. Analytical search approaches to information retrieval necessitate the systematic formulation of specific, well-structured queries. Browsing involves broad query terms and scanning larger sets of information in a relatively unstructured manner. Browsing is generally considered to virtually involve no planning, preparation or focus. For instance, Marchionini [55] notes that browsing does not involve planning and is often utilized as an alternative to an analytical search strategy. Many studies have been reported that show the benefits of browsing, for instance [11], [40], [53].

Faceted Search and Navigation

Unlike traditional taxonomies in which the hierarchy of categories is fixed and inflexible, faceted search enables users to decide how to navigate information hierarchically. For instance, users can decide how they will move from a category to its subcategories, and at the same time decide the order in which the categories are presented. Faceted navigation guides users by showing them available categories without requiring them to browse through hierarchies that may not suit their needs or way of thinking [35].

Lookahead

Lookahead [9] supports exploration with no penalty. For instance, some web applications automatically complete query terms and suggest popular searches such as shown in Fig. 1.

Fig. 1. Lookahead

Surrogates

It is important for the user to be able to assess search results. Objects such as images can be displayed in the results as complete objects. It is therefore relatively to assess such results. However, for other objects such as videos and documents, it is often not practical to display them in the results as complete objects. In the latter case, information about those objects is included e.g. key-frames for video objects; titles and abstracts for documents; thumbnails for Web pages; etc. This type of information is sometimes referred to as a surrogate [5].

Relevance Feedback

Relevance feedback enables users to guide an IR system by indicating whether they consider particular results to be more or less relevant [60]. Relevance feedback modifies an existing query based on available user-based relevance judgements for previously retrieved documents. It is worth pointing out that it is also possible to consider automatic relevance feedback, whereby the underlying information retrieval system is fully automated without user interaction, and with many relevance judgments [64].

Summarization, Analytics and Visual Presentation

Summarization and analytics can enable users digest query results. Summarization can be considered to encompass any means of aggregating or compressing the query results into a form that is less likely to lead to information overload on the part of the user. For instance, through clustering, etc. In fact and in general, faceted search, which was described previously, can also be viewed as a form of summarization. The representation of summarization or analytics can be presented using appropriate information visualization techniques.

In Table 1 is a categorization and summary of the user interface techniques for information retrieval.

Table 1. User interface techniques for IR

User interface technique for IR	Examples/References
Query formulation and query reformulation	[36]
Browsing	[11], [40], [53]
Faceted search and navigation	[35]
Lookahead	[9]
Surrogates	[5]
Relevance feedback	[60], [64]
Summarization, analytics and visual presentation	(Mani and Maybury, 1999)

2 HCI in IR Evaluation: Appropriate Evaluation Metrics and Models

2.1 Metrics

Existing literature reports on various metrics or measures regarding information retrieval (e.g., [15], [76], [69], [77]). Over time, four standard categories of measures have emerged: performance measures, interaction measures, usability measures, and contextual measures [48]. For each of the four categories, we in the sequel specifically discuss measures that are appropriate to Human-Computer Interaction in information retrieval evaluation.

Performance Measures

The traditional and classic evaluation measures of information retrieval system performance have been precision and recall. Such and other traditional IR measures can be found in [76]. Other measures include: F-measure, average precision (AP), mean average precision (MAP), and geometric average precision (GMAP). "*Since these measures are document-based, they measure only the performance of the system in retrieving items predetermined to be "relevant" to the information need. They do not consider how the information will be used, or whether, in the judgment of the user, the documents fulfill the information need*" [24].

Interactive Recall and Precision

The traditional IR performance measures are based on an evaluator's relevance judgments. The user's or subject's relevance judgments often do not agree with the evaluator's relevance judgments. It may also be that the evaluator has searched through hundreds of documents in order to provide relevance judgments. The user or subject may not search long enough to find all of these documents [48].

Toward addressing the mismatch between evaluator's relevance judgments and subjects' relevance judgments, *interactive recall and precision, and interactive TREC precision* [74-75] have been proposed.. For instance: interactive recall is the number of TREC relevant documents saved by the user divided by the number of TREC relevant documents in the corpus; interactive TREC precision is the number of TREC relevant documents viewed by the user divided by the total number of documents viewed.

Multi-level Relevance and Rank Measures

The traditional IR performance measures do not take into account that relevant documents appearing further down on the results list are likely to be less useful because users are less likely to view them. The user needs to put in some effort to get to those documents and by the time the user arrives at the document its content may be less valuable because of what the user has learned on the way to the document. Although MAP was created to address the ordering problem in systems-centered research, it still maintained some of the problematic assumptions of the traditional IR performance measures [48]. The following measures have consequently been proposed:

- J¨arvelin and Kek¨al¨ainen's *cumulated gain* measures [44-45].
- Borlund and Ingwersen's *ranked half-life* measures [14].
- Cooper's expected *search length* [12], [20].
- Dunlop's *expected search duration* [27].
- Losee's *average search length* [54].
- K¨aki and Aula's *immediate accuracy* [47].

Time-Based Measures

Time-based measures are often used as indicators of efficiency. It is worth noting that effectiveness (performance), efficiency and satisfaction are standard usability measures. Although the three measures are interrelated, they can also be looked at separately. Efficiency will be looked at again later when describing usability measures.

- K¨aki and Aula [47] describe two time-based IR measures that are relevant to Human-Computer Interaction, namely *search speed* and *qualified search speed*. Although the measures are based on answers not relevant documents, they can be extended to cover retrieval itself.
- Cleverdon et al. [18] describe the *response time of the system*.

Informativeness

Informativeness is a measure for evaluating search results by focusing on relative evaluations of relevance rather than absolute measures [70-72]. Although informativeness measure has not yet been validated, renewed interest in the measure will perhaps lead to its validation and adoption.

Cost and Utility Measures
A number of authors such as Cooper [21] and Salton [63], [61] have proposed cost and utility measures.

Contextual Measures
"*Much less attention has been paid to contextual aspects of end-user searching of electronic information systems, by either librarians or information scientists.*" [24]. There exists research evidence, such as seen in Saracevic and Kantor [67] and Dalrymple [23], acknowledging the importance of the user's context in information retrieval.

User Characteristics
Measuring user characteristics (or sometimes referred to as individual differences) separately from the search process can enable the researcher to use them to predict performance or to explain differences in performance [15].

- Fenichel [31] highlights common measures of user characteristics including: *sex of subject, age, college major, profession, level of computer experience, and level of search experience*.
- Ford et al. [32] propose Internet *perceptions* and *cognitive complexity* as additional measures of user characteristics.
- Kelly [48] proposes the following additional measures of user characteristics: intelligence, creativity, personality, memory, and cognitive style.

Measures of Information Needs
There are also IR measures that characterize the information need. For instance:

- Task-related measures (e.g. task-type, task familiarity, task difficulty and complexity)
- Topic-related measures (e.g. topic familiarity and domain expertise)
- Persistence of information need
- Immediacy of information need
- Information-seeking stage
- Purpose, goals and expected use of the results

Interaction Measures
Interaction measures are used to describe the activities and processes that subjects engage in during information retrieval. Interaction measures include:

- Number of queries
- Number of search results viewed
- Number of documents viewed
- Number of documents saved
- Query length

Since most interaction measures are counts, they can be combined to form other measures. For instance, the number of documents saved can be divided by the number of documents viewed [48].

Usability Measures

Usability is the extent to which users can use a system with effectiveness, efficiency and satisfaction to accomplish a task in a specified context of use [43]. Although there exist other definitions of usability, the ISO definition is one of the most commonly used.

Effectiveness

This is the extent to which the user is able to reach goal while using the system. The most common way for measuring effectiveness in HCI studies has been by measuring error rate and binary task completion [39]. In information retrieval, effectiveness can be measured by using appropriate measures from the performance measures that were described earlier (for instance: interactive precision and interactive recall), and also by eliciting self-reported data from subjects about their perceptions of performance.

Efficiency

Efficiency refers to how fast the user takes to finish tasks using the system. One of the most common ways for measuring efficiency is by recording the time it takes a subject to complete a task [39]. Efficiency can therefore include measures such as:

- The overall time the subject takes
- Amount of time the subject spends doing specific things
- Amount of time the subject spends in specific or different modes

In addition to the foregoing efficiency measures, and like with effectiveness, efficiency can also be measured by eliciting self-reported data from subjects about their perceptions of efficiency.

Satisfaction

Satisfaction assesses how much the user is satisfied with the system. Satisfaction can be viewed as the contentment, fulfilment or gratification that users experience when they accomplish particular goals or desires.

Other User-Relevant Measures

Besides the standard usability measures, there are other possible user-oriented measures that are relevant to information retrieval evaluation. They include:

- Preference
- Mental effort and cognitive load
- Flow and engagement: Flow is a "*mental state of operation in which a person is fully immersed in what he she is doing, characterized by a feeling of energized focus, full involvement, and success in the process of the activity.*" [22], and engagement is "*a quality of user experiences with technology that*

is characterized by challenge, aesthetic and sensory appeal, feedback, novelty, interactivity, perceived control and time, awareness, motivation, and interest and affect" [58].

- Learning and cognitive transformation: The focus here is on the extent to which the system helps users learn about a particular topic.

In line with the foregoing discussion, the metrics that are appropriate to Human-Computer Interaction in Information Retrieval evaluation can be categorized as seen in Table 2.

Table 2. Categorization of measures appropriate to HCI in IR evaluation

Categorization of measures appropriate to HCI in IR evaluation	
Performance measures	• Traditional IR performance measures [76] ➢ Recall ➢ Precision ➢ F-measure ➢ Average precision (AP) ➢ Mean average precision (MAP) ➢ Geometric average precision (GMAP) ➢ Precision at *n* ➢ Mean reciprocal rank (MRR) • Performance measures for interactive information retrieval [74-75] ➢ Interactive recall ➢ Interactive user precision ➢ Interactive TREC precision ➢ Relative relevance (RR) • Multi-level relevance and rank measures ➢ Cumulated gain measures [44-45] ➢ Ranked half-life measures [14] ➢ Expected search length [12], [20] ➢ Expected search duration [27]

	➢ Average search length [54] ➢ Immediate accuracy [47] • Time measures ➢ Search speed [47] ➢ Qualified search speed [47] ➢ Response time of the system [18] • Informativeness [70-72] • Cost and utility measures [21], [63], [61]
Contextual measures	• User characteristics ➢ Sex of subject, age, college major, profession, level of computer experience, and level of search experience [31] ➢ Internet perceptions and cognitive complexity [32] ➢ Intelligence, creativity, personality, memory, and cognitive style [48] • Measures of information needs [48] ➢ Task-related measures (e.g. task-type, task familiarity, task difficulty and complexity) ➢ Topic-related measures (e.g. topic familiarity and domain expertise) ➢ Persistence of information need ➢ Immediacy of information need ➢ Information-seeking stage ➢ Purpose, goals and expected use of the results
Interaction measures	• Number of queries • Number of search results viewed • Number of documents viewed

	• Number of documents saved • Query length • Combinations of such measures
Usability measures	• Effectiveness [39] ➢ (Note it can be measured using appropriate performance measures e.g. interactive recall, interactive precision, interactive TREC precision, informativeness, cost, utility, etc) • Efficiency [39] ➢ Overall time the subject takes ➢ Amount of time the subject spends doing specific things ➢ Amount of time the subject spends in specific or different modes ➢ Etc • Satisfaction • Other relevant measures ➢ Preference ➢ Mental effort and cognitive load ➢ Flow [22] ➢ Engagement ➢ Learning and cognitive transformation [58]

2.2 Models and Theories

An information retrieval system can in general be viewed as one that consists of a "*device interposed between a potential user of information and the information collection itself*" [34], containing three major components:

1. Database
2. Communication channel or interface between the user and the database, and which has:

    - A physical component for facilitating interaction.

- A conceptual component that guides the user on how to interact with the information structure and search mechanisms.

3. User

According to Hansen [33], IR research is moving from text representations and related techniques to also include studies of the users and their information needs, behaviour and strategies, and interaction processes.

Information Foraging Theory

Information foraging theory is a theory proposed by Pirolli and Card [59] that describes information retrieval behaviour. The theory is derived from the evolutionary ecological explanations of food-foraging strategies in anthropology and behavioral ecology. It is based on the analogy of an animal deciding what to eat, where it can be found, the best way to obtain it and how much "energy" the meal will provide (how filling the meal will be) as illustrated below:

> "*Imagine a predator, such as a bird of prey, that faces the recurrent problem of deciding what to eat, and we assume that its fitness, in terms of reproductive success, is dependent on energy intake. Energy flows into the environment and comes to be stored in different forms. For the bird of prey, different types of habitat and prey will yield different amounts of net energy (energetic profitability) if included in the diet. Furthermore, the different food-source types will have different distributions over the environment. For the bird of prey, this means that the different habitats or prey will have different access or navigation costs. Different species of birds of prey might be compared on their ability to extract energy from the environment. Birds are better adapted if they have evolved strategies that better solve the problem of maximizing the amount of energy returned per amount of effort. Conceptually, the optimal forager finds the best solution to the problem of maximizing the rate of net energy returned per effort expended, given the constraints of the environment in which it lives.*" [59, p. 8)

Humans may be considered to be "informavores" that constantly make decisions on what kind of information to look for, whether to stay at the current site/place to try to find additional information or whether they should move on to another site/place, which path or link to follow to the next information site/place, and when to finally stop the search. Central to the information foraging theory is the concept of "information scent". Just like animals rely on scents to indicate the chances of finding prey in current area and guide them to other promising patches, humans rely on various cues in the information environment to get similar answers. Humans estimate how much useful information they are likely to get on a given path or direction, and after seeking information they compare the actual outcome with their predictions. When the information scent stops getting stronger (i.e., when users no longer expect to find useful additional information), the human users move to a different information source.

Berrypicking Model

The berrypicking model [6] acknowledges that searches are evolving and occur bit by bit. Users constantly change their search terms in response to the results returned from the IR system. The very act of searching gives feedback which may cause users to modify their cognitive model of the information being searched for. Moreover, information retrieval can be bit by bit. Therefore, the query is satisfied not by a single final retrieved set of results, but by a series of selections of individual references and bits of information at each stage of the ever-modifying search. The model therefore uses the analogy of picking huckleberries or blueberries in the forest. The berries tend to be scattered on the bushes and do not often come in bunches. They need to be picked one at a time.

Ingwersen's Cognitive Model

The traditional model of IR systems represents IR as a two prong set (system and user) of elements and processes converging on comparison or matching. One attempt to improve on the traditional IR model is made by Peter Ingwersen in his cognitive model [41]. IR interaction is viewed as a set of cognitive processes, which involves system characteristics (representational and retrieval techniques), the user's situational characteristics and the functionalities of the user interface/intermediary. The cognitive viewpoint of IR embraces the complexity inherent in IR when users are involved and focuses attention on the cognitive activities that take place during information seeking and retrieval, and user information, user-system interactions [41]. Ingwersen and J¨arvelin [42] identify five central and interrelated dimensions of the cognitive viewpoint:

1. Information processing takes place in senders and recipients of messages;
2. Processing takes place at different levels;
3. During communication of information any actor is influenced by its past and present experiences (time) and its social, organizational and cultural environment;
4. Individual actors influence the environment or domain; and
5. Information is situational and contextual.

While it is clear in viewing these dimensions that the cognitive viewpoint focuses on the user, Ingwersen and J¨arvelin [42] are careful to point out that the cognitive viewpoint is not just about users' cognitive structures, but also about the numerous other cognitive structures represented in the IR system. For instance, cognitive structures represented by document authors and IR system developers.

Belkin's Episodes Model

This model concentrates on what happens in interaction as a process. Ingwerson's model focused on elements. Belkin's episodes model [7] views interaction as a series of episodes where a number of different things happen over time. For instance:

- Processes of judgement, use, interpretation, etc depending on user's goals, tasks.
- Processes of navigation, comparison, summarization, etc.
- Involving different aspects of information and information objects.

The user's interaction with the information system is the central process, which should be understood as interaction, especially as human-computer interaction.

Belkin's Evaluation Model for IRR

Belkin et al. [8] suggest an evaluation model and methodology grounded in the nature of information seeking and centred on usefulness. The model assumes that in accomplishing the general work task and achieving the general goal, the user engaged in information seeking goes through a sequence of information interactions, each having its own short term goal that contributes to achieving the general goal. The model is illustrated in Fig. 2.

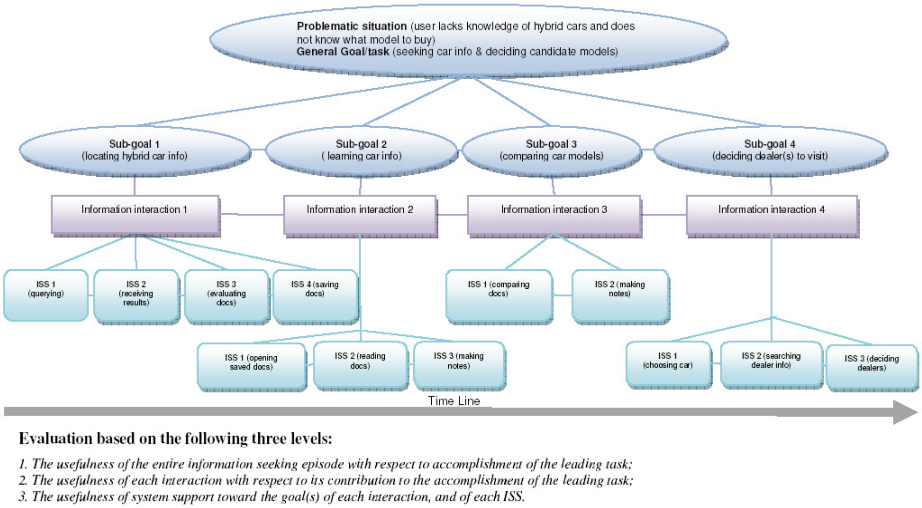

Fig. 2. Belkin's evaluation model for IRR

Saracevic's Stratified Model of IR Interaction

The stratified model starts with assumptions that:

- Users interact with IR systems in order to use information
- The use of information is connected with cognition and then situational application.

Saracevic [65-66] proposed and enhanced the stratified interaction model whereby interaction of the interplay among different levels of users and systems is the central component. While users engage in cognitive, affective, and situational levels of interaction, system involvement includes engineering, processing, and content-level. The complexity and dynamic interaction process requires changes and adaptations from both the user and system side.

Ellis' Model of Information-Seeking Behaviours

Ellis' model [28-29] concentrates on the behavior instead of on cognitive activities. The model has six key components which correspond to types of information-seeking characteristics: 1) starting, 2) chaining, 3) browsing, 4) differencing, 5) monitoring, and 6) extracting. In a more recent work, Ellis and Haugan [30] further modeled the information-seeking patterns of engineers and research scientists in relation to their research activities in different phases and types of projects, and identified similar behavior patterns.

Kuhlthau's Model

Kuhlthau [50-51] has proposed a model that describes the tasks involved in the information seeking process from a psychological perspective, containing affective/feelings, cognitive/thoughts, and physical/action activities. The model actually complements Ellis' model by attaching to stages of the 'information search process' the associated feelings, thoughts and actions, and the appropriate information tasks. The stages of Kuhlthau's model are initiation, selection, exploration, formulation, collection and presentation.

Other Specific Evaluation Models and Frameworks

Allen's Model

This model [3] which is shown in Table 3 offers a framework that can be used to support and guide IR evaluation.

Table 3. Allen's model

Component	Method	Task
Resource Analysis	Description of information system functionality	Describe resources used to complete the tasks
User Needs Analysis	• Questionnaire (qualitative and quantitative data) • Log statistics (quantitative data)	• Users goals, purpose, objectives, actions, individual preferences • Measures like time, number of actions, and type of actions
Task analysis	Hierarchical Task Analysis	Users tasks, goals and activities that they accomplish when meeting their needs
User Modeling		Merging needs, user tasks and goals, and system tasks
Designing for usability	Requirement lists (qualitative data)	Requirements for user interface redesign

Ahmed et al.'s User-Centred Approach to the Design and Evaluation of IR Interfaces

Ahmed et al. [1] have proposed a user-centred approach for designing IR interfaces. The approach is based on performing the following:

1. A competitive analysis of an existing IR system to perform usability testing.
2. A user task analysis based on activities during usability test.
3. An initial prototype design drawn from task analysis.
4. A heuristic evaluation of the initial prototype design.
5. An interactive prototype design, incorporating input from heuristic evaluation.

6. A formative evaluation of the interactive prototype using task scenarios.
7. A revised prototype design based on formative evaluations, and finally.
8. A summative evaluation of the final prototype design and a comparison of the results with the results of competitive analysis for performing the same tasks.

IIR (Interactive Information Retrieval) Evaluation Model
Borlund [13] proposes the IIR evaluation model whose key elements are the use of realistic scenarios (referred to as simulated work task situations), and the (call for) alternative performance measures such as the ones that were described earlier.

The information retrieval theories and models can be categorized as summarized in Table 4.

Table 4. Categorization of IR theories and models

General approach	**Specific view of IR**	**Examples of theories/models**
Cognitive	Elements	• Ingwersen's cognitive model [41]: IR viewed as a set of cognitive processes (i.e. elements of cognitive processes).
Interaction	User's interaction with system as episodes	• Belkin's episodes model [7]: Focuses on user's interaction with IR system, where interaction is viewed as a series of episodes. • Belkin's evaluation model for IRR [8]: IR interaction viewed as a series information interactions each with a short term goal that contributes to achieving the general goal.
	Complex and dynamic interplay of users and systems	• Saracevic's stratified model of IR interaction [65-66]: IR interaction viewed as the interplay among different levels of users and systems requiring changes and adaptations.

Table 4. (*Continued*)

Behaviour	Standard	• Ellis' model [28-29]: IR viewed as some specific information-seeking behaviour or activities.
	Extended	• Kuhlthau's model [50-51]: complements Ellis' model by including the associated feelings, thoughts and actions.
	Ecological	• Information foraging theory [59]: IR viewed analogically as evolutionary ecological food-foraging behavioral strategies. • Berrypicking model [6]: IR viewed as evolving and occurring bit by bit, analogous to picking huckleberries or blueberries in the forest.

3 Framework for Usability Evaluation in Information Retrieval

A framework for the usability evaluation of an Information Retrieval system would entail aspects or parameters such as described in the sequel.

3.1 Participants

HCI Experts

It is common in HCI to involve HCI experts in evaluating interactive systems. This is normally done during the early phases of the design process. Evaluation methods that involve HCI experts are referred to as expert-based methods. They include: heuristic evaluation and cognitive walkthrough. This approach to evaluation where the participants are HCI experts is relevant also specifically to IR systems.

Users

It is also useful to conduct evaluations whose participants are the intended users of the IR system. According to Siatri [68], it is interesting to note that the first user studies were investigating people's information seeking needs [73], [10]. There are many different types of evaluations where the participants are the intended users, for instance usability tests, observational methods (e.g., think-aloud and stimulated recall), query techniques (e.g., questionnaires and interviews), physiological monitoring methods (e.g., eye tracking, measuring skin conductance, measuring heart rate), etc. It is worth noting that involving the intended users in the evaluation of the IR system makes the evaluation set up more closely resemble the actual information retrieval processes and settings users would experience in the real world.

Surrogate Users

Sometimes it is extremely difficult to find and recruit actual users to participate in an evaluation. For example: high-powered individuals, national intelligence personnel, etc. In such cases, it is often better to involve surrogate users as a proxy for the actual users than not to conduct a user-based evaluation at all. It is however important to appropriately manage risks associated with surrogate users [52]. For instance: surrogate users should as much as possible resemble the actual users i.e. they should share key and relevant characteristics with the actual users.

There exist many methods for recruiting participants for IR evaluations, including newspaper advertising, posting signs, sending solicitations to mailing list, online advertising, using market research companies, etc. An interesting development in this area is the use of crowdsourcing, for instance through Mechanical Turk [4].

3.2 Tasks

During the evaluation of information retrieval systems, it is important to ensure that the set up is close to the actual information retrieval processes and contextual aspects users would experience in the real world. One way of introducing this realism is by involving the potential users in the evaluation of the IR system, as was mentioned earlier. Another way is by appropriately incorporating user tasks in the evaluation. The tasks can take many forms, for instance: standard tasks in information retrieval, real world work tasks, and simulated work tasks.

Standard Tasks in Information Retrieval

During information retrieval, users primarily engage in the following typical tasks [62],[5]:

- Formulation and submission of a query,
- Examination of the results, with a
- Possible feedback loop to re-formulate the query, and
- Integration of search results and evaluation of the whole search.

Each task or step indicates some statement of user requirement i.e. what the goal-directed user is trying to do with the system [46].

Simulated Work Task Situations

Borlund [13] proposes evaluation model for interactive information retrieval systems, whose key elements are the use of realistic scenarios that simulate real world work task situations. The scenarios are referred to as simulated work task situations. A simulated work task situation is a semantically rather open description of a scenario of a given IR requiring situation. A simulated work task situation is aimed at triggering and developing a simulated information need by allowing for user interpretations of the situation, leading to cognitively individual information need interpretations as in real-life. Research suggests that a simulated work task situation is more time consuming for the participants because it requires them to complete an additional and more complex task beyond finding relevant documents [48]. The additional task is to actually use the information in a manner that matches the user model behind the search task. It is therefore important to ensure that the simulated work task situations that the participants are presented with are not overly involving or tedious.

3.3 Measures

We had previously, in Section 2.1, described in detail the metrics or measures that are appropriate to Human-Computer Interaction in Information Retrieval evaluation. In the sequel we focus on the measures that would be relevant specifically to the usability evaluation of IR systems.

Standard Usability Measures

Effectiveness
In information retrieval evaluation, effectiveness can be measured by eliciting self-reported data from users about their perceptions of performance. Effectiveness in information retrieval can also be measured by using appropriate measures from the performance measures that were described previously. In particular:

- *interactive recall*
- *interactive precision*
- *interactive TREC precision*

Efficiency
In information retrieval evaluation, efficiency can be assessed using such measures as:

- The overall time the user takes
- The time the user takes doing specific things
- The time the user takes in specific or different modes

Efficiency can also be measured in information retrieval by eliciting self-reported data from users about their perceptions of efficiency.

Satisfaction
In information retrieval evaluation, satisfaction can be measured by eliciting self-reported data from users about their level of contentment, fulfilment or gratification as a result of using or interacting with the information retrieval system.

Interaction Measures

Interaction measures are relevant to usability evaluation of information retrieval systems. They include:

- Number of queries
- Number of search results viewed
- Number of documents viewed
- Number of documents saved
- Query length
- Appropriate combinations of the above measures

It is worth pointing out that interaction measures can be resourceful when assessing effectiveness.

User Characteristic Measures

It is important to measure user characteristics when conducting usability evaluation of information retrieval systems. Such information can for instance enable the researcher explain differences (such as in effectiveness) between different users. Measures of user characteristics could include: *sex, age, profession, computer experience, search experience*, Internet *perceptions*, *cognitive style,* etc.

Information Need Measures

Measures of information need are important when conducting usability evaluation of information retrieval systems. Such measures can for instance enable the researcher to predict or explain efficiency and effectiveness regarding particular topics. Information retrieval measures that characterize the information need include:

- Task-related measures (e.g. task-type, task familiarity, task difficulty and complexity)
- Topic-related measures (e.g. topic familiarity and domain expertise)
- Persistence of information need
- Immediacy of information need
- Information-seeking stage
- Purpose, goals and expected use of the results

Other User-Relevant Measures

There are also other measures that are closely related to the standard usability measures, and are therefore relevant to the usability evaluation of information retrieval systems. They include:

- Preference
- Mental effort and cognitive load
- Flow and engagement
- Learning and cognitive transformation

3.4 Evaluation Method(s)

There exist many methods that can be used for usability evaluation. Although there are several different ways of classifying them, they can generally be categorized as expert-based evaluation methods and user-based evaluation methods. User-based evaluation methods include: usability tests, observational methods (e.g. think aloud, stimulated recall/post-task walkthrough, transaction logging), query techniques (e.g., questionnaires and interviews), and physiological monitoring methods (e.g., eye tracking, measuring skin conductance, measuring heart rate). Expert-based evaluation methods include: heuristic evaluation and cognitive walkthrough.

After collecting the evaluation data, there are some basic things one should do before embarking on the actual data analysis [25], including:

- Looking at the data: A simple glance at the data could be all that is necessary.
- Saving the data: One might need to do more analysis in the future.

According to Dix et al. [25], the choice of the statistical technique for data analysis depends on factors such as:

- The questions we want to answer e.g., "Is there a difference?" (e.g., is one IR system better than another?), "how big is the difference?", "how accurate is the estimate?"
- Type of data/variables e.g. discrete data vs. continuous data; number of independent variables vs. number of dependent variables.

In the following we highlight important considerations when using some specific methods in the usability evaluation of information retrieval systems.

Heuristic Evaluation vs. Usability Testing

Doubleday et al. [26] compared heuristic evaluation with user testing on an IR system. The expert evaluators identified 86 usability problems whereas 38 problems were identified in the user testing. However, not all of the 38 problems found by user testing were identified by the expert evaluators. Some genuine problems would therefore have gone undetected if there had been no user testing. Another example is reported by Cogdill [19], where the expert evaluators identified 27 usability problems compared to 21 problems found in the usability test. Cogdill also noted that using both heuristic evaluation and usability testing resulted in a high degree of comprehensiveness in the study. It is therefore worth pointing out that expert-based and user-based evaluation methods can play a complementary role in evaluating information retrieval systems.

Transaction Logging

Although transaction logging is one of the oldest and most common methods for collecting data when evaluating interactive information retrieval systems, the recent explosion of studies using Web transaction log data has re-popularized the approach. The method relies on computer and Web monitoring tools in order to collect logs characterizing user's interaction with the system. There are various types of logging including: system, proxy, server, and client logging. The researcher however needs to be aware of the following main challenges when using transaction log data: ensuring the validity and reliability of the logger, extracting and preparing data generated by the logger, and interpreting the data [48]. Most transaction logging tools can run in the background while the user interacts with the information retrieval system, without causing any distractions or disruption. Transaction logging is therefore a potentially useful observational method because it can capture users' natural search behaviours without interrupting them. Transaction logs can also be resourceful in providing an objective dimension to the information retrieval evaluation measures that the research is interested in.

Questionnaires

Questionnaires are common in information retrieval evaluations. Questionnaires can be used at various points during an evaluation of an IR system. Consequently there are several types of questionnaire, for instance: screening questionnaire, pre-study questionnaire, and post-study questionnaire. Questionnaires can be administered electronically or manually (pen-and-paper). Kelly et al. [49] found that in the context of interactive information retrieval evaluation, subjects' responses to closed-questions were significantly more positive when elicited electronically, than manually.

Interviews

Research suggests that in information retrieval evaluation, interviews are more appropriate when one is asking complex, abstract questions than when one is asking relatively easy questions [49]. Interviews can also be useful in information retrieval evaluation during simulated recall/post-task walkthrough [48].

Think-Aloud

In think-aloud the user is expected to perform an information retrieval task and at the same time articulate their thoughts as they carry out the task. One of the challenges with the think-aloud is that users may have a difficult time simultaneously articulating their thoughts and carrying out the information retrieval task that they have been given. In many evaluations, the information retrieval system is novel. Users may therefore not be able to handle the additional cognitive demands placed by think-aloud, while they are also learning how to interact with the system. Some researchers have proposed that subjects complete a short training task before they start searching in order to get accustomed to think-aloud [48]. It is worth noting that stimulated recall (i.e. post-task walkthrough) can serve as an alternative to think-aloud. In stimulated recall, the researcher records the screen of the computer as the user performs the searching task. After the searching task is complete, the recording is played back to the user who is then asked to articulate their thoughts and decision-making as the recording is played.

The kinds of parameters that would be typically expected in a framework for the usability evaluation of an Information Retrieval system are summarized in Table 5.

Table 5. Parameters expected in a framework for the usability evaluation of IR systems

Evaluation aspect	**Parameters**
Participants	• HCI experts • Users • Surrogate users
Tasks	• Standard tasks in information retrieval e.g. ➢ Formulation and submission of a query, ➢ Examination of the results, with a ➢ Possible feedback loop to re-formulate the query, and ➢ Integration of search results and evaluation of the whole search [62], [5] • Simulated work task situations [13]
Measures	• Standard usability measures ➢ Effectiveness (interactive recall, interactive precision, interactive TREC precision, informativeness, cost, utility, etc) ➢ Efficiency (overall time user takes, time user takes doing specific things, time user takes in specific or different modes, etc) ➢ Satisfaction • Interaction measures e.g. ➢ Number of queries ➢ Number of search results viewed ➢ Number of documents viewed ➢ Number of documents saved ➢ Query length ➢ Appropriate combinations • User characteristic measures e.g. ➢ Sex ➢ Age ➢ Profession ➢ Computer experience ➢ Search experience ➢ Internet perceptions ➢ Cognitive style • Information need measures e.g. ➢ Task-related measures (task-type, task familiarity, task difficulty, complexity, etc) ➢ Topic-related measures (topic familiarity and domain expertise, etc) ➢ Persistence of information need ➢ Immediacy of information need ➢ Information-seeking stage

	➢ Purpose, goals and expected use of the results • Other relevant measures e.g. ➢ Preference ➢ Mental effort and cognitive load ➢ Flow and engagement ➢ Learning and cognitive transformation
Evaluation method(s)	• User-based evaluation methods e.g. ➢ Usability tests ➢ Observational methods (think aloud, stimulated recall/post-task walkthrough, transaction logging, etc) ➢ Query techniques (questionnaires and interviews) ➢ Physiological monitoring methods (eye tracking, measuring skin conductance, measuring heart rate, etc) • Expert-based evaluation methods e.g. ➢ Heuristic evaluation ➢ Cognitive walkthrough

4 Conclusions

In this chapter, we have observed that the traditional methods of evaluating IR systems have over a long period of time been primarily concerned with system-oriented measurements such as precision and recall, but not on the usability aspects of the IR system. Moreover, there are no well-established evaluation approaches for studying users and their interactions with IR systems. It is therefore important to consider, appropriately adjust and invent user interface techniques that can support the user in their information retrieval tasks by guiding, supporting and transforming the user's information problems, goals or needs. Human Computer Interaction researchers and designers should also endeavour to appropriately use, revise and propose user-information seeking models and evaluation techniques for information retrieval systems. In line with that, we have in this chapter described: existing user interface techniques for supporting users to interact with and use information retrieval systems, measures that are appropriate to Human-Computer Interaction in information retrieval evaluation, existing IR information seeking theories and models, and IR evaluation frameworks. We have also described the typical elements that would constitute a framework for the usability evaluation of an Information Retrieval system.

It is worth noting that there are some trends that are not only posing unique challenges but also providing tremendous opportunities to the IR community and other communities including the HCI community. For instance: IR of massive user-generated content (e.g. microblogs, social network discussion forums data, user-generated multimedia, etc), user-participation and crowdsourcing in IR, dynamic or continuously evolving and growing data (e.g. sensor data), etc. There is a need to realize user interface techniques that can support users in information retrieval tasks in the context of such trends. It is also worth noting that such trends challenge standard approaches to the usability evaluation of IR systems. For instance, crowdsourcing introduces aspects such as collaboration, trust, etc to the standard IR evaluation measures.

All in all, Human-Computer Interaction aspects are important in information retrieval. Efforts aimed at appropriately incorporating HCI techniques in IR can realize IR systems that meet and possibly exceed the needs of the intended users.

References

1. Ahmed, S.M.Z., McKnight, C., Oppenheim, C.: A user-centred design and evaluation of IR interfaces. Journal of Librarianship and Information Science 38(2), 157–172 (2006)
2. Allen, B.: From research to design: A user-centered approach. In: Ingwersen, P., Pors, N.O. (eds.) CoLIS 2. Second International Conference on Conceptions of Library and Information Science: Integration in Perspective, pp. 45–59. The Royal School of Librarianship, Copehagen (1996)
3. Allen, B.: Information tasks. Towards a user-centered approach to information systems. Academic Press, San Diego (1996)
4. Alonso, O., Rose, D.E., Stewart, B.: Crowdsourcing for relevance evaluation. SIGIR Forum 42, 10–16 (2008)
5. Baeza-Yates, R., Ribeiro-Neto, B.: Modern information retrieval. Addison-Wesley, Reading (1999)
6. Bates, M.J.: The design of browsing and berrypicking techniques for the online search interface. Online Review 13(5), 407–424 (1989)
7. Belkin, N.J., Cool, C., Stein, A., Thiel, U.: Cases, scripts and information seeking strategies: on the design of interactive information retrieval systems. Expert Systems with Applications 9, 379–395 (1995)
8. Belkin, N., J., Cole, M., Liu, J.: A Model for Evaluation of Interactive Information Retrieval. In: SIGIR Workshop on the Future of IR Evaluation (2009)
9. Berger, L.: Look ahead caching process for improved information retrieval response time by caching bodies of information before they are requested by the user. United States Patent (1999)
10. Bernal, J.D.: Preliminary Analysis of Pilot Questionnaires on the Use of Scientific literature. In: The Royal Society Scientific Information Conference, pp. 589–637 (1948)
11. Borgman, C.L., Hirsh, S.G., Walter, V.A., Gallagher, A.L.: Children's searching behaviour on browsing and keyword online catalog: the Science Library Catalog Project. Journal of the American Society for Information Science 46(9), 663–684 (1995)
12. Borlund, P.: The concept of relevance in IR. Journal of the American Society for Information Science 54, 913–925 (2003)
13. Borlund, P.: The IIR evaluation model: a framework for evaluation of interactive information retrieval systems. Information Research 8(3) (2003)
14. Borlund, P., Ingwersen, P.: Measure of relative relevance and ranked halflife: Performance indicators for interactive information retrieval. In: Proceedings of the 21st ACM SIGIR Conference on Research and Development of Information Retrieval (SIGIR 1998), Melbourne, Australia, pp. 324–331 (1998)

15. Boyce, B.R., Meadow, C.T., Kraft, D.H.: Measurement in Information Science. Academic Press, Inc., London (1994)
16. Callan, J., Allan, J., Clarke, J.L.A., Dumais, S., Evans, D.A., Sanderson, M., Zhai, C.: Meeting of the MINDS: An information retrieval research agenda. SIGIR Forum 41, 25–34 (2007)
17. Cleverdon, C.W.: The Cranfield tests on index language devices. In: Spark-Jones, K., Willett, P. (eds.) Readings in Information Retrieval. Reprinted from Aslib Proceedings, pp. 173–192. Morgan Kaufman Publishers, San Francisco (1967)
18. Cleverdon, C. W., Mills, L., Keen, M.: Factors Determining the Performance of Indexing Systems. Design, vol. 1. Aslib Cranfield Research Project, Cranfield (1996)
19. Cogdill, K.: MEDLINEplus Interface Evaluation: Final Report, University of Maryland, Human-Computer Interaction Lab (HCIL), College Park, MD (1999)
20. Cooper, W.S.: Expected search length: A single measure of retrieval effectiveness based on the weak ordering action of retrieval systems. American Documentation 19, 30–41 (1968)
21. Cooper, W.S.: On selecting a measure of retrieval effectiveness, part 1: The "subjective" philosophy of evaluation. Journal of the American Society for Information Science 24, 87–100 (1973)
22. Csikszentmihalyi, M.: Finding Flow: The Psychology of Engagement with Everyday Life. Basic Books, New York (1997)
23. Dalrymple, P.W.: Retrieval by reformulation in two university library catalogs: Toward a cognitive model of searching behavior. Journal of the American Society (1990)
24. Dalrymple, P.W.: User-Centered Evaluation of Information Retrieval. In: Allen, B. (ed.) Evaluation of Public Services and Public Services Personnel, pp. 85–102. University of Illinois, Urbana (1991)
25. Dix, A., Finlay, J., Abowd, G., Beale, R.: Human-Computer Interaction. Prentice Hall (2003)
26. Doubleday, A.R., Ryan, M., Springett, M., Sutcliffe, A.: A comparison of usability techniques for evaluating design. In: Proceedings of the Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, August 18-20, pp. 101–110. ACM, Amsterdam (1997)
27. Dunlop, M.: Time, relevance and interaction modeling for information retrieval. In: Proceedings of the 20th ACM Conference on Research and Development in Information Retrieval (SIGIR 1997), Philadelphia, PA, pp. 206–213 (1997)
28. Ellis, D.: The derivation of a behavioral model for information system design. Unpublished doctoral dissertation. University of Sheffield, England (1987)
29. Ellis, D.: A behavioural approach to information retrieval system design. Journal of Documentation 45(3), 171–212 (1989)
30. Ellis, D., Haugan, M.: Modeling the information seeking patterns of engineers and research scientists in an industrial environment. Journal of Documentation 53(4), 384–403 (1997)
31. Fenichel, C.H.: Online searching: Measures that discriminate among users with different types of experience. Journal of the American Society for Information Science 32, 23–32 (1981)
32. Ford, N., Miller, D., Moss, N.: The role of individual differences in Internet searching: An empirical study. Journal of the American Society for Information Science and Technology 52, 1049–1066 (2001)
33. Hansen, P.: Evaluation of IR User Interface. Implications for user interface design. Human IT (2), 28–41 (1998)

34. Harter, S.: Online information retrieval. Concepts, principles, and techniques. Academic Press, Orlando (1986)
35. Hearst, M.: User Interfaces and Visualization. In: Baeza-Yates, R., Ribeiro-Neto, B. (eds.) Modern Information Retrieval, ch.10 (1999)
36. Hearst, M.: Query Reformulation. In: Search User Interfaces, ch. 6. Cambridge University Press (2009)
37. Hewett, T., Baecker, R., Card, S., Carey, T., Gasen, J., Mantei, M., Perlman, G., Strong, G., Verplank, W.: ACM SIGCHI Curricula for Human-Computer Interaction (1992)
38. Hix, D., Hartson, H.R.: Developing user interfaces. Ensuring usability through product and process. Wiley, New York (1993)
39. Hornbaek, K.: Current practice in measuring usability: Challenges to usability studies and research. International Journal of Human–Computer Studies 64, 79–102 (2005)
40. Hutchinson, H.B., Drunin, A., Bederson, B.B.: Supporting elementary-age children's searching and browsing: design and evaluation using the International Children's Digital Library. Journal of the American Society for Information Science 58(11), 1618–1630 (2007)
41. Ingwersen, P.: Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory. Journal of Documentation 52, 3–50 (1996)
42. Ingwersen, P., Järvelin, K.: The Turn: Integration of Information Seeking and Retrieval in Context. Springer, Dordrecht (2005)
43. ISO: Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs): Part II, Guidance on Usability, ISO 9241- 11:1998 (1998)
44. Järvelin, K., Kekäläinen, J.: IR evaluation methods for retrieving highly relevant documents. In: Proceedings of the 23rd ACM SIGIR Conference on Research and Development of Information Retrieval (SIGIR 2000), Athens, Greece, pp. 41–48 (2000)
45. Järvelin, K., Kekäläinen, J.: Cumulated gain-based evaluation of IR techniques. ACM Transactions on Information Systems (TOIS) 20, 422–446 (2002)
46. Johnson, F.C., Griffiths, J.R., Hartley, R.J.: Task dimensions of user evaluations of information retrieval systems. Information Research 8(4) (2003)
47. Käki, M., Aula, A.: Controlling the complexity in comparing search user interfaces via user studies. Information Processing and Management 44, 82–91 (2008)
48. Kelly, D.: Methods for evaluating interactive information retrieval systems with users. Foundations and Trends in Information Retrieval 3(1-2), 1–224 (2009)
49. Kelly, D., Harper, J., Landau, B.: Questionnaire mode effects in interactive information retrieval experiments. Information Processing and Management (2008)
50. Kuhlthau, C.C.: Inside the search process: information seeking from the user's perspective. Journal of the American Society for Information Science 42, 361–371 (1991)
51. Kuhlthau, C.C.: Seeking meaning: a process approach to library and information services. Ablex Publishing, Norwood (1994)
52. Lievesley, M.A., Yee, J.S.R.: Surrogate Users – A Pragmatic Approach to Defining User Needs. In: Conference Proceedings and Extended Abstracts, CHI 2007, San Jose. ACM Press (2007)
53. Lin, X.: Map displays for information retrieval. Journal of the American Society for Information Science 48(1), 40–54 (1997)
54. Losee, R.M.: Evaluating retrieval performance given database and query characteristics: Analytical determination of performance surfaces. Journal of the American Society for Information Science 47, 95–105 (1996)
55. Marchionini, G.: An invitation to browse: designing full text systems for novice users. Canadian Journal of Information Science 12(3), 69–79 (1987)

56. Marchionini, G.: Toward Human-Computer Information Retrieval Bulletin. Bulletin of the American Society for Information Science (June/July 2006), http://www.asis.org/Bulletin/Jun-06/marchionini.html
57. Mira working group: Evaluation Frameworks for Interactive Multimedia Information Retrieval Applications (1996), http://www.dcs.gla.ac.uk/mira
58. O'Brien, H., Toms, E.: What is user engagement? A conceptual framework for defining user engagement with technology. Journal of the American Society for Information Science and Technology 59, 938–955 (2008)
59. Pirolli, P., Card, S.K.: Information foraging. Psychological Review 106(4), 643–675 (1999)
60. Rocchio, J.: Relevance feedback in information retrieval. In: Salton, G. (ed.) The SMART Retrieval System (1971)
61. Salton, G.: Evaluation problems in interactive information retrieval. Information Storage and Retrieval 6, 29–44 (1970)
62. Salton, G.: Automatic text processing: the transformation, analysis and retrieval of information by computer. Addison-Wesley, Reading (1989)
63. Salton, G.: The state of retrieval system evaluation. Information Processing and Management 28, 441–449 (1992)
64. Salton, G., Buckley, C.: Improving retrieval performance by relevance feedback. Journal of the American Society for Information Science 41(4), 288–297 (1990)
65. Saracevic, T.: Modeling interaction in information retrieval (IR): A review and proposal. In: Proceedings of the 59th ASIS Annual Meeting 1996, vol. 33, pp. 3–9 (1996)
66. Saracevic, T.: The stratified model of information retrieval interaction: Extension and applications. In: Proceedings of the 60th ASIS Annual Meeting 1997, vol. 34, pp. 313–327 (1997)
67. Saracevic, T., Kantor, P.: A study of information seeking and retrieving. Journal of the American Society for Information Science 39(3), 177–216 (1988)
68. Siatri, R.: The Evolution of User Studies. Libri 49, 132–141 (1999)
69. Su, L.T.: Evaluation measures for interactive information retrieval. Information Processing and Management 28, 503–516 (1992)
70. Tague, J.: Informativeness as an ordinal utility function for information retrieval. SIGIR Forum 21, 10–17 (1987)
71. Tague-Sutcliffe, J.M.: The pragmatics of information retrieval experimentation, revisted. Information Processing and Management 28, 467–490 (1992)
72. Tague-Sutcliffe, J.M.: Measuring Information: An Information Services Perspective. Academic Press, San Diego (1995)
73. Urquhart, D.J.: The Distribution and Use of Scientific and Technical Information. In: The Royal Society Scientific Information Conference, pp. 408–419 (1948)
74. Veerasamy, A., Belkin, N.J.: Evaluation of a tool for visualization of information retrieval results. In: Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Zurich, Switzerland, pp. 85–92 (1996)
75. Veerasamy, A., Heikes, R.: Effectiveness of a graphical display of retrieval results. SIGIR Forum 31, 236–245 (1997)
76. Voorhees, E.M., Harman, D.K.: TREC: Experiment and Evaluation in Information Retrieval. MIT Press, Cambridge (2005)
77. Yuan, W., Meadow, C.T.: A study of the use of variables in information retrieval user studies. Journal of the American Society for Information Science 50, 140–150 (1999)