iR **information**research

Proceedings of the Seventh International Conference on Conceptions of Library and Information Science—"Unity in diversity" — Part 2

How do users evaluate individual documents? An analysis of dimensions of evaluation activities

Iris Xie, **Edward Benoit III** and **Huan Zhang**
School of Information Studies, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, WI 53201, USA

Abstract

Introduction. Evaluation plays an important role in users' information searching and retrieving processes. While previous research mainly focuses on applied criteria, less research is on other dimensions of evaluation. This study explores the dimensions of evaluation activities including criteria applied, elements examined, activity engaged in and time spent in evaluating individual documents.
Method. Conducting their own tasks, thirty-one participants with different demographic backgrounds, participated in the study. Multiple data collection methods were used: pre-questionnaires, interaction diaries, think-aloud protocols, transaction logs and post-questionnaires.
Analysis. Types of evaluation criteria, elements, evaluation activities and their associated pre- and post-activities were analysed by open-coding. Descriptive analysis was applied to analyse the most and leasdt time spent on individual documents with associated factors.
Results. The findings of this study reveal that evaluation activities are complicated and dynamic. Eighteen types of evaluation criteria, seven types of evaluation elements, six types of evaluation activities and their associated pre- and post-activities were identified. In addition, factors affecting time spent in evaluating individual documents were analysed.
Conclusions. The authors offer suggestions for information retrieval system design to support effective evaluation.

CHANGE FONT

Introduction

Evaluation is an essential component for information retrieval and searching. Evaluation can be investigated on two levels: search result and individual document evaluation. In this study, the authors focus on the evaluation of individual documents. While search result evaluation refers to the identification of useful or relevant documents from the retrieved results, document evaluation refers to the assessment of usefulness or relevance of an individual document retrieved or browsed. The notion of evaluation is fraught with layers of complexity including the cognitive elements required for evaluation. Often, users spend more time determining the value of a given document through evaluation than the other types of information search activities.

Additionally, evaluation can be characterized in multiple dimensions. Previous studies investigated criteria, in particular the relevance criterion, more than other dimensions of evaluation. Users' relevancy judgments maintain a focal point in current evaluation research (Bade 2007; Barry 1998; Borlund 2003; Fitzgerald and Galloway 2001; Schamber 1991; Vakkari and Hakala 2000). Recently, however, a handful of studies investigated the role of credibility in users' document evaluation (Jansen et al. 2009; Metzger et al. 2003; Rieh 2002). Moreover, previous research is limited largely to evaluation criteria. In addition, many of the researchers use simulated searching tasks in their studies, which cannot reveal the nature of evaluation activities. The limitations of previous research call for further exploration of users' evaluation of individual documents.

Research problem and research questions

Researchers do not understand fully the complex nature of evaluation. Previous research on evaluation mainly focuses on applied criteria. Less research is on other dimensions of evaluation, such as the elements or components that users examine for evaluation and their evaluation activities with associated pre- and post-activities. Finally, the amount of time users spend evaluating may also help researchers understand the nature of evaluation. The study addresses the following research questions:

- What criteria do users apply in evaluating an individual document?
- What components or elements do users examine in evaluating an individual document?
- What are the types of evaluation pre- and post-activities, if any?
- How much time do users spend in evaluating an individual document?

Literature review

As previously noted, the current evaluation research offers a solid foundation on the discussion of relevance as the key evaluation criterion. Several studies explored relevance judgments through a pre-internet lens, thereby shaping trends for future studies (Barry 1994; Park 1993; Saracevic 1969; Schamber 1991). More recent literature continues studying relevance. In doing so, several authors noted the limitation of existing relevance-based research. Barry (1994, 1998) suggested the need for more in-depth research and questions users' credibility evaluation, indicating a need for future research. Borlund's (2003) '*multidimensionality of relevance*' highlights the lack of consistency in defining relevance itself. Additional studies '*recommends the broadening of the perspectives*' (Vakkari and Hakala 2000) noting '*no literature addresses how relevance and evaluation judgments might coalesce into a decision*' (Fitzgerald and Galloway 2001: 992).

Some researchers (Rieh 2002; Savolainen and Kari 2006; Tombros *et al.* 2005) further proposed the need for a more complex understanding of evaluation, highlighting the current literature limitations. In this review, the authors summarized previous literature on evaluation, in particular evaluation criteria, elements, activities and time.

Criteria

As previously noted, research on evaluation criteria divides into two main areas, relevance and credibility and represents the bulk of evaluation literature. Relevance evaluation often focuses on the similarity measures used

by information retrieval systems rather than user perspective of how documents themselves are judged. Bade (2007) expressed his frustrations with the misinterpretation of similarity measures and relevancy evaluation. He argued for a multidimensional understanding, rather than a binary approach and objects to how relevancy is thought of as either objective or subjective. This thinking leads to two major flaws, as he states, '*The first is that any relevance judgment by any human being will always be subjective to some degree, no matter how objective that person may strive to be. The second is that nothing is relevant to anything for any machine*' (Bade 2007: 840). Borlund's (2003) examination of relevance also revealed how the relevance concept, especially in regard to the '*multidimensionality of relevance*', was many faceted and did not just refer to the various relevance criteria users might apply in the process of judging relevance of retrieved information objects, which explains why some argue that no consensus had been reached on the relevance concept.

One of the earlier studies of relevance criteria indicated the criteria employed by users including: tangible characteristics of documents (e.g., the information content of the document, the provision of references to other sources of information), subjective qualities (e.g., agreement with the information provided by the document) and situational factors (e.g., the time constraints under which the user was working). While the studies' interview process revealed participants' true and detailed thoughts, which proved its relatively complete list of criteria, the results are limited to the evaluation of printed textual materials (Barry 1994). Barry's (1998) later study further explored the evaluation of documents, in an attempt to identify which various document representations contain clues that allow users to determine the presence, or absence, of traits and/or qualities, which determine the relevance of the document to the user's situation. She concluded that document representations might differ in their effectiveness as indicators of potential relevance because different types of document representations vary in their ability to present clues for specific traits and/or qualities.

Further studies indicate the complicated and dynamic nature of users' relevance criteria. Greisdorf (2003) pointed out relevance judgment process is a problem-solving and decision-making process, in which topicality, pertinence, and utility of a retrieved item are considered. Vakkari and Hakala (2000) found a connection between an individual's changing understanding of his or her task and how the relevance of references and full texts is judged. In similar fashion, Fitzgerald and Galloway (2001) developed a complex theory of critical thinking regarding relevance and evaluation judging. Through their study, they found a high degree of intermingling between relevance judging and evaluation, along with a number of strategies participants used to assist them in both thinking processes. Additionally, affect and consideration of convenience likewise influenced decision making (Fitzgerald and Galloway 2001). Savolainen and Kari (2006) found eighteeen different user-defined relevance criteria by which people select hyperlinks and pages in Web searching. Of the individual criteria, specificity, topicality, familiarity and variety were used most frequently in relevance judgments. The study shows that despite the high number of individual criteria used in the judgments, a few criteria such as specificity and topicality tend to dominate.

In addition to relevance criteria, creditability and authority criteria also get more attention. Rieh found users typically judge credibility and authority both while searching and evaluating specific documents. Furthermore, Rieh concluded, '*usefulness and goodness are the two primary facets of information quality*' (2002: 157). Fogg *et al.* (2003) used a large sample of users in real settings in their examination of creditability judgments. Accordingly, they discovered that site presentation had the biggest impact on creditability. Another examination of factors of credibility by Jansen *et al.* points out the effects of brand, '*Brand was found to have significant influence on number of all links examined, all links clicked and sponsored links clicked. Brand also appeared to have significant effects on all links and sponsored links relevance ratings*' (Jansen *et al.* 2009: 1590). The trust issue appeared to have the biggest impact, as users simply trusted a known brand over an unknown one, often being more willing to examine sponsored links of a known brand, such as Yahoo! or Google.

Different users apply their evaluation criteria in different ways. After comparing how college students and nonstudents evaluate Web-based resouces, Metzger, Flanagin and Zwarun (2003) found non-students viewed Internet resources in a vastly different way. While both students and nonstudents agreed on the variance of credibility of internet-based sources, '*the nonstudents indicated that they verified online information more than the students did, although both groups reported that they verify online information only rarely to occasionally*'

([Metzger*et al.*2003](#): 285). In attempting to explain this finding, the authors speculate students may use or find the Iinternet sources for convenience and, therefore, do not think of the quality of their work. While this is disconcerting at face value, the authors suggest the real numbers may be more shocking. They state '*Student participants reported that they verify online information only 'rarely' to 'occasionally,' and it is quite possible that these results are somewhat inflated due to the social desirability inherent in this measure*' ([Metzger *et al.* 2003](#): 287).

Criteria identified are not limited to relevance and credibility. Barry and Schamber compared their previous individual studies, thereby establishing several criteria beyond relevance and credibility. The study identified over twenty criteria categories within Barry's previous work, including '*the document as a physical entity; other information or sources within the environment... [and] the user's beliefs and preferences*' ([Barry and Schamber 1998](#): 224). Additionally, Schamber's prior studies offered additional criteria not previously mentioned ([Schamber and Eisenberg 1991](#); [Schamber *et al.* 1990](#)). These include, '*Geographic Proximity...Clarity, Dynamism and Presentation Quality*' ([Schamber em>et al. 1990](#): 225). Barry and Schamber ([1998b](#)) also found that criteria such as *access* and *affectiveness* appear within both researchers' previous works.

Elements

Although limited, researchers recently began their investigation on evaluation elements. The misuse of criteria for elements within a handful of studies illustrates the relative immaturity of the field ([Savolainen and Kari 2006](#); [Tombros *et al.* 2005](#)). Similar to criteria research, element-based research splits into two areas, content and format elements.

Through an investigation of the assessment criteria used by online searchers, Tombros *et al.* ([2005](#)) identified the features of Web pages that online searchers use when they assessed the utility of Web pages for information-seeking tasks; how the features used by the participants varied depending on the type of task; and the variation of Web page features as participants progressed along the course of tasks.

Kelly *et al.* ([2002](#)) identified patterns among relevant documents and their features, such as lists, tables, frequently asked question lists, forms, downloadable files, question terms present in special markup, links and length. They found a difference between the features of relevant pages for task and fact questions, i.e. lists occur more often in documents relevant to task questions, frequently asked questions are more common in task questions, links are more common in documents relevant to fact questions and documents relevant to task questions are longer, on average, than other documents ([Kelly *et al.* 2002](#)). Examinations regarding specific elements used to assess relevancy indicate these elements can be grouped into six categories: abstract, author, content, full text, journal/publisher and personal. They also indicated that multiple elements were examined when making relevant, partially relevant and not-relevant judgments and that most criteria could have either a positive or negative contribution to the relevance of a document. The elements most frequently mentioned by study participants were content, followed by criteria characterizing the full text document ([Maglaughlin and Sonnenwald 2002](#)).

After investigating which documents researchers choose for a research project, Wang and Soergel ([1998](#)) developed the concept of Document Information Elements, typically referred to as format elements. The study identified those elements that users applied as specific criteria (such as quality, topicality, authority, etc.) to make inclusion or rejection decisions. Such user evaluation combines both criteria and elements into a singular early study. While Wang and Soergel highlighted the criteria previously mentioned, their discussion of information elements remains a rare notation. The study indicated that the four major elements which users applied 'were title, abstract, journal and author' ([Wang and Soergel 1998](#): 124). The study's participants noted an additional fourteen elements including geographic location, publication date, document type, author's affiliation, descriptors, language, etc. Additionally, the authors found that specific elements corresponded to a specific criterion, such as 'Recency' pairing with 'Publication Date' or 'Novelty' with 'Title' and 'Author' ([Wang and Soergel 1998](#): 125).

Evaluation activities and time

The complexity of users' evaluation of documents extends beyond criteria and elements. As previously noted, time, evaluation activities and their pre- and post-activities may further illustrate the nature of user evaluation. Few researchers have examined activities related to user's evaluation. Some studies attempt to translate use behaviour into an algorithmic based quantitative study (Chi *et al.* 2001). Others rule out the effectiveness of using click-through data for evaluating the subject (Joachims *et al.* 2005). Similar to the scarcity of literature relating to user evaluation activities, few previous temporal studies of user interaction or evaluation of documents exist. Jansen and Spink (2006) found that the percentage of users who view only one page of results has dramatically risen from 29% in 1997 to 73% in 2002. He *et al.* (2002) on the other hand found an average session of twelve minutes and an earlier study by Jansen and Spink (2003) found an average session time of fifteen minutes. The latter study also examined the average time users spent examining a single document. Although the average time spent examining one document was sixteen minutes and two seconds, adjusted for the outlying data, '*over 75% of the users view the retrieved Web document for less than 15 minutes*' (Jansen and Spink 2003: 4). The authors noted nearly half of the users (40%) viewed documents for less than three minutes; therefore, users typically viewed retrieved documents between three and fifteen minutes.

Limitations of current research

The focus of existing literature itself reveals the limitation on evaluation research. A vast majority of studies examine relevance judging. Although some researchers investigated the different elements users examined during the evaluation process, few have addressed evaluation activities, or evaluation time. An unbalanced literature suppresses the full understanding of user evaluation behaviours, thereby limiting its application to system design. Additionally, most of the literature relies heavily on simulated environments, in which participants search for assigned search topics, rather than their own tasks.

These limitations reveal the need for further investigation of dimensions of evaluation activities. This study examines evaluation criteria applied, elements examined, evaluation activities engaged and time spent. Moreover, the use of real users with real problems, allow the study's findings better to represent the nature of evaluation activities.

Methodology

Sampling

A total of thirty-one participants, responding to fliers posted to different community centres and public places (e.g., local libraries, grocery stores, etc.), listserves (e.g., Craigslist), as well as newspaper advertisements, were recruited from the Greater Milwaukee area; they represented general users of information with different sex, race, ethnic backgrounds; education and literacy levels; computer skills; occupations; and other demographic characteristics. Each participant was paid $75 for their involvement in the study. Table 1 summarizes the participants in this study.

Demographic characteristics		Number	Percentage
Sex	Male	10	32.3%
	Female	21	67.7%
Age	18-20	1	3.2%
	21-30	13	41.9%
	31-40	5	16.1%
	41-50	7	22.6%
	51-60	5	16.1%
	61+	0	0.0%
Native Language	English	29	93.5%
	Non-English	2	6.5%
Ethnicity	Caucasian	29	93.5%
	Non-Caucasian	2	6.5%
Computer Skills	Expert	3	9.7%
	Advanced	21	67.7%
	Intermediate	7	22.6%
	Beginner	0	0.0%
Occupation	administrative assistant, marketing communications, librarian, student, programmer, tutor, school guidance secretary, social worker, portfolio specialist, trust associate, client relationship associate, software developer, self employed, buyer, nurse, academic advisor, painter, unemployed, etc.		

Table 1: Demographic characteristics of participants (n=31)

Data collection

Multiple methods were used to collect data. Participants were first asked to fill in prequestionnaire in which their demographic information and their experience in searching for information were requested. Participants were asked to keep an *information interaction diary* for two weeks to record how they achieved two search tasks, one work related and another personal-related. The diary consisted of information in relation to their tasks, source selections, search activities including evaluation activities and reasons associated with applying different types of search activities, among other pieces of information. After that, participants were also invited to come to the Information Intelligence and Architecture research laboratory to search for information for two additional work-related and personal search tasks. They were instructed to think aloud during their search process. Their information search processes were captured by Morae, usability testing software that not only records users' movements but also captures their thinking aloud, including their feelings and thoughts during the search process. Finally, participants were asked to fill in the post-questionnaire regarding their experience in their search activities, their problems and factors affecting their search activities.

Each participant was asked to conduct two self-generated tasks instead of assigned tasks. However, two out of sixty-two tasks were not able to be analysed because of the poor quality of the recorded data; thus the total of tasks being analysed in this study was sixty. Diary data were only used for offering explanations to evaluation activities because the data provided in the diaries was not as detailed as the log data, while data recorded by the Morae software were analysed mainly for this study. The recoded data were transcribed. A participant's every movement was transcribed and his or her related verbal protocols were also included in the transcription. The coding scheme applied in this study is presented in the Data Analysis section below.

Data analysis

The unit of analysis is each evaluation activity. As defined in the Introduction, evaluation refers assessment of the usefulness or relevance of individual documents retrieved or browsed. It starts from a user assesses an individual document and ends when the user moves to the next search activity. Types of evaluation criteria, elements and evaluation activities and their associate pre/post activities were analysed based on open coding (Strauss and Corbin 1990). In the end, the authors developed a coding scheme consisting of evaluation criteria, elements, evaluation activities and their associated pre- and post-activities, as well as time. Table 2 presents the coding scheme. To save space, the detailed discussion of different types of criteria, elements, evaluation activities and their associated pre- and post-activities and time with definitions and examples are presented in the Results section.

To test inter-coder reliability, two researchers independently coded twenty tasks from ten participants randomly selected from sixty tasks performed by thirty-one participants. The inter-coder reliability for criteria, elements and evaluation pre- and post-activities was 0.90, 0.95 and 0.95 respectively, according to Holsti's (1969) reliability formula. Reliability = 2M/(N1+N2), where M is the number of coding decisions on which two coders agree and N1 and N2 refer to the total number of coding decisions by the first and second coder, respectively.

Variables	Definitions	Examples
Criteria	Participants' judgments applied during the evaluation of an individual document.	Scope, 'This site was not useful at all it did not provide much info about it (s5).' Credibility, 'But this article has references so that might be reliable (s16).'
Elements	The individual components that participants examined during the evaluation of an individual document.	Text, 'Just reading what kind of a plant it is (s18).' Author, 'Oh, Zach Lau, I know him (s6).'
Evaluation pre- and post-activity	Participants' actions taken during, before and following the evaluation of an individual document.	Evaluation activity, 'I think I want to look at that... I am going to Control F it and look for my word, arsenic [using find function] (s19).' Pre-activity, [types www.babyname.com into URL] (s1). Post-activity, 'Ok, so I am going to print all course details (s24),'
Time	Time that participants spent in evaluating an individual document from beginning to the end in seconds.	130 seconds evaluating brides.com (s15) and 330 seconds evaluating en.wikipedia.org (s3)

Table 2: Coding scheme

Finally, descriptive analysis were applied to analyse time spent on individual documents including the most spent time and least spent time with associated factors.

Results

The findings of this study are presented based on the research questions proposed: 1) types of criteria applied 2) types of elements examined, 3) types of evaluation pre- and post-activities engaged in and 4) time spent in evaluating an individual document.

Criteria

Eighteen types of criteria emerged from the data. These criteria can be classified into the following four categories: 1) content coverage 2) quality, 3) eesign and 4) access.

Content coverage

The data show six criteria related to content coverage: scope, specificity, depth, document type, data type and intended use. *Scope* is one of the most important criteria for evaluation and refers to the extent to which information provided by the document is covered. Participants seemingly always prefer the individual documents that cover more aspects of one topic, as one participant stated,

> This site was very useful it gave a little explanation under the picture of each different model and gave little insights on what it has compared to other different models before and after it gave info on whether it has Bluetooth capability….it gave more info on the differences between each one than the actual Garmin site did, it was more beneficial than the actual Garmin site was (s5).

Thus if the document failed to cover enough information, participant felt dissatisfied, as another one commented,

> It does not look like a lot is going on in Bastogne…it does not really tell me much about how big the town is or anything (s23).

Comprehensive coverage is not enough. Specificity and depth also play essential roles in evaluation. *Specificity* refers to the extent to which information covered by the document is focused to match the user needs. Here are two typical examples,

> This whole page is devoted to body world info which is what I was looking for and I have ticket prices here which I am going to make a note of (s12),

and

> Qualifications…18 years old by August 1 2008…application deadline….this has basically answered all my questions right off the bat…(s31).

Depth refers to the extent to which information provided by the document is in detail. While most of the participants liked in-depth information, some did not, which was determined by their information needs. For instance, one participant said,

> This 6th site I feel was the best out of all of them. It went into a lot of detail, it gave personal accounts of this man's experience having a dog with it and it gave a lot of detail how he dealt with it what kind of symptoms and treatments that the other sites did not offer. Overall this one gave me the most info (s5).

However, another one commented,

> it supports using drugs and they go to more specifics which I am not interested in (s7).

Document type and data type are the two key criteria applied for document evaluation. *Document type* refers to the formats of the individual document such as PDF, video, etc. Participants preferred some specific types of documents to present their needed information. According to participant 12,

> I am trying to figure out what this is about because we have not done any…always looking for ways to mix up lesson and make it less boring, if I had a video it might be helpful [teaching the class]). (s.12)

Participant 21 liked the slide show,

> they have a slide show of healthy desserts so you get to see a picture which I always appreciate. (s.21)

Data type refers to the formats of the information provided by individual documents including but not limited to pictures, numbers, etc. One participant explained,

> There are pictures of the hotels so I can get an idea of what they look like…the Ramada looks good (s28).' Another one stressed, 'This one caught my eye because it looks like it had some statistics (s12).

Another important evaluation criterion, *intended use* refers to the targeted audience of individual documents. Participants normally quickly check whether they were the intended audience of the document. One participant said,

> It kind of gives like a sales pitch not really an informative description (s5)

and another one mentioned,

> These are for hotels I think I am getting away from (s29).

Document quality

The findings of this study show five criteria related to document quality: reputation, currency, unique information, credibility and accuracy/validity. *Reputation* of the source of the document is the first criterion selected by participants to indicate the document quality. Reputation refers to the extent to which the source of a document is well known or reputable. Most of the participants showed considerably trust on the documents from reputable and authoritative sources. Reputable sources include education sites, news agencies, governmental sites, popular commercial sites, etc., illustrated by the following two examples,

> My topic is very recent and it is business related one of my first choices is cnn.com. I have used that site before and has [*sic*] a good variety of indexes and tabs to follow that I am sure something this huge would definitely be front and centre (s2).

> I am searching for books on classroom management in a college setting and I am going to start with amazon.com because I know that one is specific to books and more accurate than Google (s28).

At the same time currency and unique information are the key criteria indicating document quality. *Currency* refers to the extent to which information provided is timely, recent or up-to-date. While one participant praised the currency of the document,

I don't want that...this is not a current project (s19)

another one expressed dissatisfaction for an out-of-date document,

it's 1998 so it might be old info (s16)

Unique information refers to the extent to which information provides new viewpoints or ideas. Although unique information could be listed under content, the study's participants applied the criterion as a quality measure. For example, participant 2 discussed the unique finding from a document,

they kind of reiterate what I found in CNN…one of the big differences from what the White House says and what the industry says are quite different and this gives actual percentages (s2).

Participant 3 highlighted what Wikipedia offers,

Wikipedia does a good job of formulating the amount…maximum scores are 45T …there are no calculators…ok that was informative (s3).

Both credibility and accuracy or validity are undoubtedly the indicators for document quality. *Credibility* refers to the extent to which information provided is reliable. Participants are consistently looking for trustworthy information, as one participant indicated,

This looks like people are just writing posts. It's a personal thing, not reliable (s12)

.

At the same time, they rejected the information that seemed credible, as another participant said,

I want to find an apartment agency that I trust, yet I don't trust a lot of what I see on the Web… so something that I really care about, then I want to switch to a different type of source (s29).

Accuracy or validity refers to the extent to which information provided is seen as correct or convincing. Participants in general made judgments based on their own knowledge and experience. The following examples reveal participants' concerns on the credibility and accuracy of the provided information.

This does not seem right 23% lower wow…this is not right…I don't know... oh my I can't believe it (s17)

and '

$63,000 ….that's a little unrealistic I am going back to Google (s21).

Design

Design is the third aspect of evaluation criteria. The results of this study show three criteria related to design: specific features, organization/ layout and ease of use. *Specific features* of information retrieval systems relate to both criteria and elements. Participants considered these features as part of their evaluation criteria. Table 3 presents the types of features participants mentioned with examples.

Feature types	Examples
References	'I am scanning because I know this Website usually has some good references in their subheadings they have related articles that some would be relevant to this topic too so I am going to chose one of them (s2).'
Document description	'...used the document description feature to select location on map (s4).'
Customer review	'It also gives a link with customer reviews so you get more insight on how other people feel about each unit and that might give a better idea of how good the product is actually based on other people's experiences (s5).'

Table 3: Specific features used as evaluation criteria

For evaluation, participants also cared about *organization and layout*, as one participant described,

> I don't like the way they have this laid out. When I clicked on this one I immediately looked to see how it is organized (S12).

Another preferred a well-organized layout of prices,

> but the results that they give are nice and clear and the prices are clear which I appreciate (s21).

In addition, participants appreciated the systems that are *easy to use*. For instance, one participant claimed,

> It looks like a good source but then went you look at it…I don't know how to use this Website I am not going to bother with it (s7).

Accessibility

Accessibility is another area that participants considered in the evaluation process. There are four criteria related to accessibility: availability, cost, language and speed. Availability and cost were the main concerns for evaluation. *Availability* refers to the extent to which effort is required to obtain information. Obviously, participants hope all the desired information is available but that is not always the case, as one participant described,

> This search will not work because the 1880s are not available full text online (s14).

Another problem for availability is that sometimes participants were required to log in to view some information. While they could not or unwilling to do so, they just left, as one of the participants experienced,

> I am unable to search the library catalogue because I don't have a barcode or a pin number from the University of Chicago (s14)

echoed another one,

> oh you have to register…ok I will have to come back it seems like interesting info (s17).

Cost refers to the extent to which access to information is dependent on payment. Understandably, participants preferred those individual documents that provide free information. Some examples are as follows,

> These [study guides] are free which is nice…this is a really good Website for practice (s3), '

> I am going on their Website to see free patterns (s8).

Language is also an issue in evaluating the documents. If a document is in a foreign language that the participant did not understand, he/she would choose to give up.

> It's all in French that's not really helpful …this Website has a list of actual books, research guides that could help me in my search but it is all in French and I don't speak French so it is not helpful. (s16)

Another participant had the same experience,

> It is in German so I won't be able to use this site…. It's all in German so let me go out of here.

At the same time, *speed* is another concern. Participants were not willing to wait too long for the individual documents to load, no matter how good the information is. One participant joked,

Here's a site from Germany, probably has a server running in his basement (s6).

Another complained,

This Website takes a long time to load (s16).

Both of them left immediately.

Elements

This study also identifies seven elements that users examined when evaluating the individual documents. These elements consist of: (1) body of text, (2) sate, (3) author or source, (4) document type, (5) title, (6) abstract and (7) table of contents.

Body of text refers to the actual text of the individual documents. Within body of text, participants mainly checked the following: text, link, picture, number and others. Table 4 presents different types of elements in body of text that participants viewed for evaluation.

Types of elements	Examples
Text	'I am finding some of my problems because it talks about cultivation diseases and I see there is something about overwatering maybe I am overwatering my pansies...talks about stem rot...that is exactly what happened so it says the plant should not be overwatered I that might be my problem...so that is one of my answers (s20).' 'Let's see it says Cincinnati Reds at Milwaukee Brewers tickets 7/13. It gives a list of what tickets are available (s26).' 'They reviewed like twenty-one different shoes and it gives prices, width, weight and where to buy the shoes (s31).
Links, i.e. the hyperlinks embedded in the individual documents	'At the bottom another thing I like is that they have a lot of links to other articles in the database so for instance interview questions and answers, illegal interview questions, really good (s12).' 'I am going to click on the 'let them eat cake' link I like cake. At the bottom there are more recipes like this so I want to see what other recipes I could get (s21).'
Pictures	'There is a pretty picture of the pea shoots...this also says they are harvested from the snow peas but they can also be from most garden pea varieties (s18).' 'This is city data...these are photos of Columbus Georgia (s23).'
Numbers	'I see some statistics I am quickly going to scan and see (s12).'
Marks	'Some have stars * what does that mean? Oh you have to sign up for these activities (s3)'
Tables	'Here they do have a table (s25)'

Table 4: Types of elements in body of text viewed

Date is one of the key elements that participants examined first to see whether the date matches his/her information need. As indicated in the 'criteria' section, participants paid attention to the date of individual documents. In the following examples, one participant preferred the old map while another one liked the new dictionary:

> We've got really old maps. Lake Michigan territory 1778, cool map, I like old maps (s4).

> So this online etymology dictionary is from 2001. I think I am going to take that as my answer (s21)

Author/Source is related to authority of individual documents, therefore participants looked at the author or source, as described in these examples,

> This is interesting. This is actually a page for a government document on Bastogne (s23)

> Oh, Zach Lau, I know him (s6).

Document type is another element that participants checked in the evaluation process. Different from the 'Document type' in the 'criteria' section, here it refers to the pure document type participants looked at, without any implication. For example,

> I am going to click on some of the photos here (s23)

and another stated,

> I will watch this video for just a second…this video is pretty much a report of people who have already been there so I don't put much stock into that I want to see it for myself (s12).

Title, *abstract* and *table of contents* highlight the content of the documents. That is why participants in general checked title, abstract and table of contents when evaluating a document. Since *titles* of the individual documents had been evaluated when participant looked at the page of results beforehand, they did not catch so much attention, while still some participants adverted to it. One participant mentioned,

> Some of these are definitely scholarly journals. So far I do not understand this title because I don't have the science background (s19).

In particular, when the documents are too long, participants would read the *abstract* or *table of contents* instead of the full documents as shown in the example,

> They have an abstract here I can get it at home (s6)

Here the participant suggests the shorter abstract will be easier to review quickly later. Another participant stated,

> Because the document is 226 pages long I am just going to look at the table of contents (s14).

Evaluation activities and their pre- and post-activities

In order to characterize and understand evaluation activity better, it is important to examine evaluation activities as well as their pre- and post-activities, i.e., the actions that participants take during, before and after the evaluation assessment of an individual document. Evaluation activities and their associated pre- and post-activities are summarized based on participants' thinking-aloud and log data. Figure 1 presents types of evaluation activities and their pre- and post-activities.
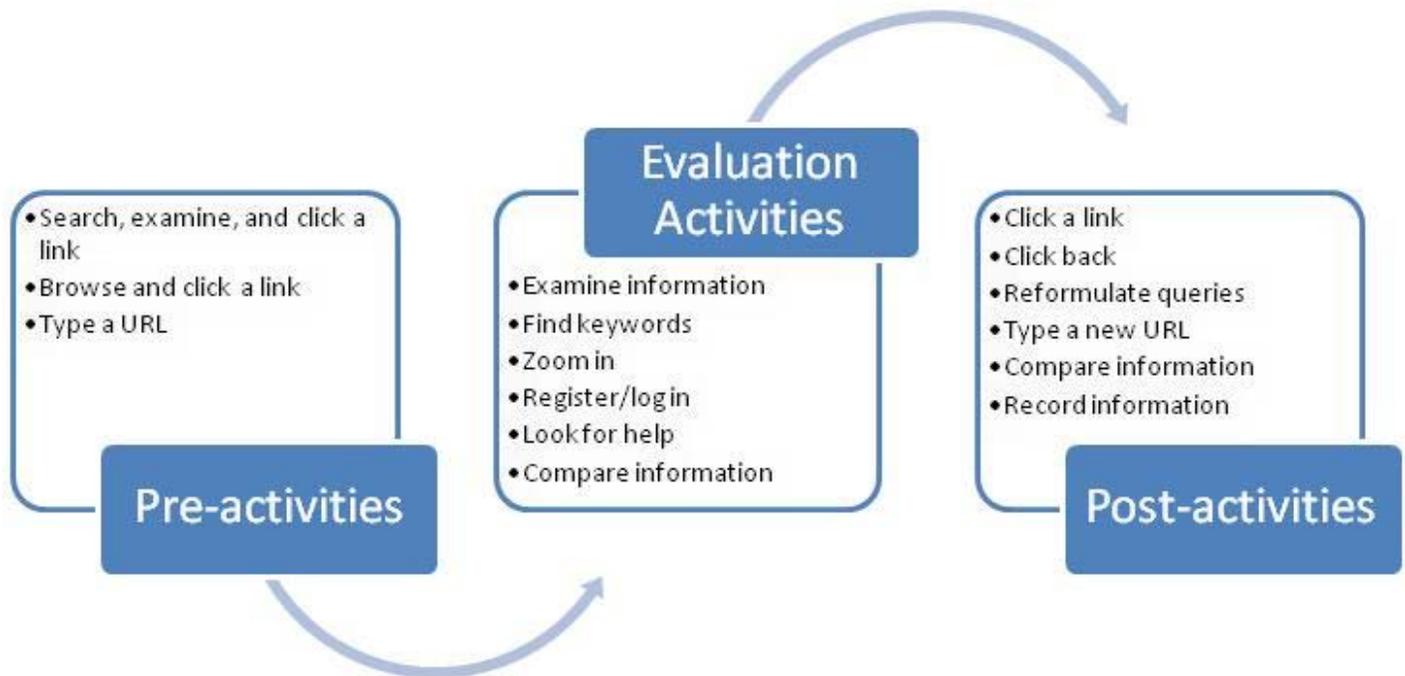
Figure 1: Evaluation activities and their pre- and post-activities

Evaluation activities

Throughout the information searching processes, participants followed various paths or pre- and post-activities. Evaluation activities can be summarized as follows: examine information, find keywords, zoom in, register or log in, look for help and compare information. The act of evaluation inherently requires users' *examination of information*, such as text, author, abstract, etc., presented in the Elements section, however, participants also undertook more specialized evaluating activities. Occasionally evaluation of long textual documents frustrates participants with time constraints. The incorporation of searching features or finding functions within Web browsers allow users to *find keywords* in an individual document. For example, one participant said,

> I am going to use the 'find function' to search for the word 'election' on this homepage and I see politics of Cook County elections campaigns but it is not covering the year that I need (s14).

Similarly, some participants used *zoom* functions to evaluate some specific information closer. One participant noted,

> Clicked 2nd result: Lake Michigan territory 1778 and zoomed in and out of map by clicking the zoom function (s4).

Another evaluation activity often occurs when Website require either *registration or logging in* to access elements of individual documents. If a piece of information is really valuable for participants, they would be willing to register or log in. Typical examples are,

> You need to log in. I think I have been here before, I think I will register [fills the registration form] (s3)

> First I have to put in my user name and password, you have to be a member [logs in] (s15)

Other participants need system assistance to properly access documents for evaluation. Sometimes, participants had to access or *look for help* functions to better use an individual document, as one participant said,

> I am also looking to see if they have a help menu [clicks link: help] (s8).

Finally, although not immediately apparent as an evaluation activity, in some cases, participants appreciated a Website offering the capability to *compare information*, such as airfares. One participant noted, 'In Tripadvisor if you put in your dates, they give you different discounts for comparison. So now go to each one and you can compare which gives you the best deals (s7).'

Pre-activities

As shown in Figure 1, there are three types of pre-activities: 1) search, examine and click a link 2) browse and click a link and 3) type a URL. *Search, examine and click a link* is one of the most frequently occurring activities before evaluation. Right before evaluating a particular individual document, participants searched in an information retrieval system, examined the results and clicked one of the search results. Participants either searched from an external search engine (i.e. Google) or an internal search engine (i.e. the one within a Website). Respective examples of log data are:

> clicks the first result from Google.com (s11)

> fills in fields to find flights, clicks 'find flights' link (s17).

Browse and click a link is another frequently occuring pre-activity. Participants browsed a site and clicked a link, including external links (i.e. link belonging to another Website) or internal links (i.e. link within the same Website). Respective examples of log data are:

> clicks a link 'biomedical news' from SpaceWar.com to go to 'Hospital and Medical News' (s27),'

> clicks: 'about missile test launches' (s9)

In some cases, participants *typed a URL* to directly go to an individual document before evaluating it. For instance, one participant

> types URL of 'cnn.com' (s2)

to go to that Website.

Post-activities

As presented in Figure 1, there are six types of post-activities: 1) click a link 2) click back, 3) reformulate queries, 4) type a new URL, 5) compare information and 6) record information. Table 5 presents the six types of post-activities with definitions and associated examples.

Post-activities	Definitions	Examples
Click a link	Participants decided to click a link, including internal links (i.e. a link pointing to an individual document also within the same Website) and external links.	'There is a bibliography so I am going to the bibliography to see [clicks link: Bibliography] if they have anything about the 1920s (s16).' 'There is a link to 'This document' and I click on that it brings me to US Department of State Website, this is a source I can trust (s29).'
Click back	Participants decided to click back to look at more of other similar results, including the results within the same Website and belonging to another Website.	'Click back, at the bottom there are more recipes like this so I want to see what other recipes I could get (s21).' 'I am going back to my Google search and find a different Website (s16).'
Reformulate queries	Participants gradually gained more knowledge about their topics which helped them to reformulate queries.	'Oh so soft baby blanket. It's kind of pretty, I kind of like the pattern on that one, so what I am going to do is copy 'red heart soft baby' into Google search (s8).' 'Alright I am going back to Google and try to search 'autism, behavioural modification techniques' (s15).'
Type a new URL	Participants typed a new URL to go to another Website.	'I am going to try a different Website; I am going to try Wikipedia [types URL] (s11).' 'This Website is not what I was looking for, maybe I will go to mpl.org [types in URL] to see if I can find an actual current book because the resources I have found are too specific or old (s16).'
Compare information	Participants accessed external information for verification or comparison.	'Ok so now it's looking like $859 is the cheapest airfare so last time I took United which I really liked so

		here it is $990... I don't know if this is the best Website I am going to try something else [types Tripadvisor.com to go to another flight Website] (s7).'
Record information	If participants found some good information, they always recorded it. Four types of recording were emerged from the data: (a) write on paper, (b) print out, (c) bookmark and (d) e-mail. The examples correlate to each type.	(a) 'It kind of just gives five different topics, bouquets, best blooms, favourite colours so I am going to write down a few of the different flower colours and things you can add to it (s15);' (b) 'Character and object animation I think that is the first class I would be interested in. Ok so I am going to print all course details (s24);' (c) 'I am going to bookmark this page and come back to that, bookmark that for later (s6);' (d) 'This looks good to me so I am going to email it to myself (s21).'

Time

Time is an important indicator for the efforts that users make in their information searching process. This study shows that participants spent most of their time engaging in evaluation activities. The mean time participants spent on evaluating an individual document is 133.65 seconds with standard deviation 197.20. Figure 2 presents the percentage of cases within each subgroup of time spent for individual document evaluation.
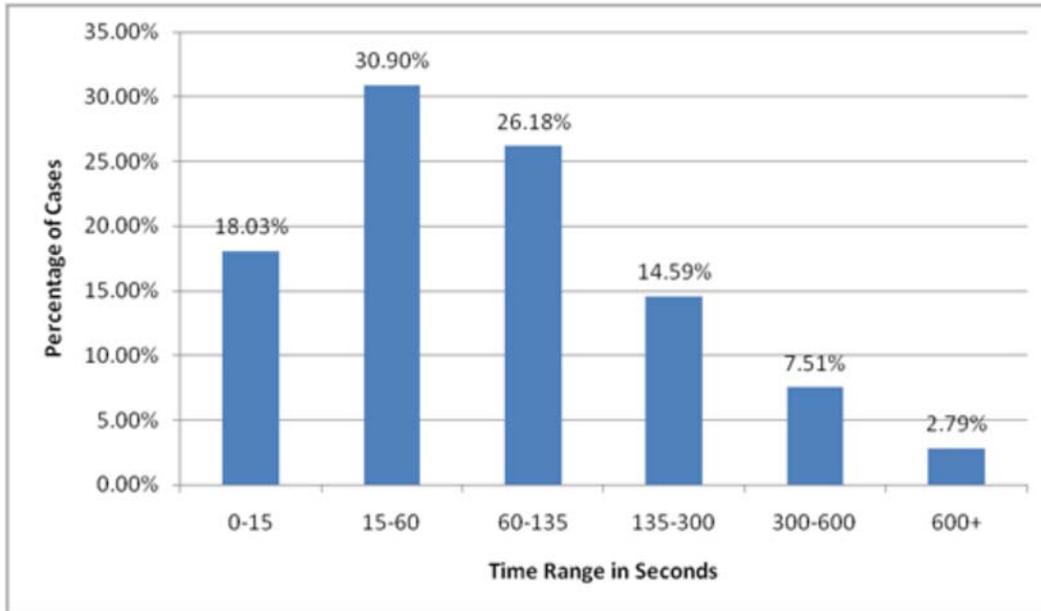
Figure 2: Time spent on individual document evaluation

Figure 2 presents time spent for individual document evaluation. Here a case refers to an individual document that participants examined. More than half of the cases (57.08%) are within the range of 15-135 seconds. At the same time 18.03% of cases spent less than 15 seconds in evaluating an individual document, while 2.79% spent more than 600 seconds. Different factors led to the differences in time spent in evaluating individual documents. The authors identified the factors that caused the two extreme situations in evaluating an individual document: time spent less than 15 second and time spent more than 600 seconds.

Time spent less than fifteen seconds

In many cases, participants quickly finished evaluating an individual document because of irrelevant information, requiring extra efforts, advertisement, duplication, non-authoritative, out-of-date and disorganized information as well as foreign language. In other words, the factors that lead to the quick evaluation are mainly associated with the documents. Table 6 presents types of factors for quick evaluation with examples.

Types of factors	Examples
Irrelevant information	'That's not what I want I found a type of disclaimer that's not what I want; I want a definition (s11),' and 'Seal the deal in seven seconds, this is about sales I don't want that go back right away (s12).'
Extra efforts	'This one I have to re-put-in my information; I am not going to bother with that I am going to Orbitz. Most of them you don't have to re-enter your flight information (s7).'
Advertisement	'That's a product catalogue that's not going to help (s6),' and 'That is not at all what I want; it's a bike forum but has ads (s19).'
Duplication	'It takes me back to the IRS page that I found on my own (s8),' and 'It's saying the same thing (s20).'
Not authoritative	'This looks like people are just writing posts; it's a personal thing, not reputable (s12).'
Out-dated information	'Very, very old, it is not the one I want (s17).'
Disorganized information	'I actually like the other Website better because it straights out listed things (s25).'
Foreign language	'It is in German so I won't be able to use this site. It's all in German so let me go out of here (s30).''It is in German so I won't be able to use this site. It's all in German so let me go out of here (s30).'

Table 6: Types of factors for quick evaluation

Time spent more than 600 seconds

While factors in relation to the individual document are the main reasons for quick evaluation, it is more complicated in terms of the types of factors that lead to participants engaging more time for the evaluation. In general, search topics have impact on time spent for individual documents. When participants have to read the documents more carefully to understand the information, the evaluation process is longer. For example, participants spent more time to evaluate individual documents that results of the following search topics, '*find information about nuclear weapons* (s1)' and '*Bush admin's plan to assist at-risk homeowners from foreclosure* (s2).' Simultaneously, research topics also prolong the evaluation process, such as '*peer-review literature on silicosis or exposure to silica in the last four months* (s4),' etc.

Normally personal tasks in relation to everyday life require less time for the evaluation. However, there are exceptions. In particular, the characteristics of the document also influence the evaluation time. If one individual document is long and contains many similar items for comparison, it might force a participant spent more time to find needed information. For instance, participant 31 needed information about new running shoes and he found a document that '*is seven pages long; there is a lot of information here*'. He spent 850 seconds to assess the document.

It is interesting to find that the majority of the documents that participants spent more than 600 seconds to evaluate are from authoritative and reputable sources in a specific area. Here are some typical examples,

> Because my topic is very recent and it is business related one of my first choices is cnn.com. I have used that site before and it has a good variety of indexes and tabs to follow that I am sure something this huge would definitely be front and centre. Well that is interesting I am looking more into the how this plan is really going to be implemented what the net fallout will be (s2 spent 675 seconds on the document.)

> There is a magazine called Runner's World, it is supposed to be experts on running (s31 spent 855 seconds on the document.)

Moreover, if a document is from the only or official source for specific information, participants undoubtedly would stay on it. For example,

> I would like get to more info on MCAT, [Medical College Admission Test] in general about the MCAT, when it is taken, scores and the best ways to practice. I am on the Medical College Admission Test official Website right now [aamc.org] (s3 spent 1080 seconds on the document.)

> Here is a Columbus Georgia homepage, I will see what they got from here (s23 spent 615 seconds on the document.).

Discussion

The results of this study not only validate findings from previous research but also introduce several new dimensions to help better understand the nature of document evaluation. The examination of criteria reiterated the importance of relevance judgments reasserting the findings of previous research (Bade 2007; Barry 1994 and 1998; Borlund 2003; Fitzgerald and Galloway, 2001; Vakkari and Hakala 2000). Moreover, the complex nature of relevance judgment expands to the multidimensionality of criteria to content coverage, quality of documents, system design and document access based on the findings, as users apply differing criteria to meet their personal understanding of relevancy. In particular, this study demonstrates that credibility remains an important criterion as seen through the aspects of reputation, currency, unique information, credibility and accuracy, thus validating previous research (Jansen *et al.* 2009; Metzger *et al.* 2003; Rieh 2002). The validity nature of the study extends

beyond criteria, however. Previous research findings relating to elements evaluated are also echoed throughout the study's findings. The findings of Kelly *et al.* (2002) connecting relevancy and document features, for example, are re-affirmed by this study. Additionally, the evaluation elements within full-text documents discussed by Maglaughlin and Sonnenwald (2002) also appear here.

The main contribution of this study also lies in to the revealing new aspects of document evaluation. Within its findings regarding criteria, the study indicates a high degree of complexity and dynamic nature. Unlike previous studies which focus on either relevance or credibility, the emergence of eighteen criteria within content coverage, quality, design and access, offers a richer view of participants' judgments. Participants apply criteria from multiple aspects depending on their personal preference and their tasks. Furthermore, the study suggests participants applied this wide array of criteria to an equally diverse element set. This study, unlike previous ones, indentified a total of seven separate elements including: body of text, date, author or source, document type, title, abstract and table of contents. The expansion of the element set indicates a broader application of evaluation criteria by users as well as the difficulty to make evaluation judgments. Through both the criteria and element findings, the study indicates user behaviour varies widely, with a high degree of personal preference.

Perhaps the most important addition to the current research is this study's findings within the areas of participants' evaluation activities and pre- and post-activities. The participants' application of specialized evaluation activities goes beyond the basic examination of information. Evaluation activities, to some extent, is a mini version of the search process which includes searching, browsing, examining, comparing, looking for help and registering to access fulltext, etc. While the pre-activities appear straight forward, the post-activities again indicate a complex variety of user behaviour depending mainly on the outcomes of the evaluation. Additionally, the evaluation activities highlight a fluid, non-linear information search process. Moreover, the findings of this study on time spent on each individual document explains why users spend short period in evaluating documents as identified by Jansen and Spink (2003). It reveals that uselessness and inaccessibility of documents are the key factors leading to quick evaluation while the factors behind longer evaluation are more related to users' interest in the topic, their tasks as well as the documents themselves.

The findings of this study have significant implications for information retrieval system design to support effective evaluation. The reiteration of current literature strengthens the findings as a strong foundation for user-based evaluation systems. Furthermore, the addition of a broader understanding of the complex nature of user document evaluation implies a need for supporting such evaluation through better system design. Table 7 presents some of the system design suggestions based on the findings of this study.

Findings of the study	Design suggestions
Types of criteria	The identified criteria that participants applied in the evaluation process require information retrieval systems to make the information associated with content coverage, quality of documents, system design and document access available or transparent to users, therefore users can quickly find the information and make their judgments. For example, an information retrieval system needs to assist users to judge the quality of documents by presenting information in relation to its reputation, unique information, currency, credibility and accuracy/validity. Additionally, adding more system features, such as offering customer reviews, document descriptions will also help the evaluation.

Table 7: System design suggestions

Types of elements	The seven elements that participants examined in assessing the documents need to be highlighted so users can easily identify them. In particular, users have to examine different elements in the body of text, such as tables, figures, pictures, numbers, etc. In that case, system design needs to extract the information out of the fulltext so users can effectively evaluate the data. It is also useful to provide key elements in visual forms so users can easily identify them. Offering overviews of individual documents is another option for design.
Evaluation activities and their associated pre- and post-activities	The evaluation process represents a miniature of the search process itself. That indicates it is not enough just to highlight the key words in the document. It is important to offer search function as well as explicit and implicit help in relation to how to find specific information and how to access part or the full documents. It is also helpful to make it easy for users to access previous documents for comparison purpose.
Time spent on evaluating individual documents	Information retrieval systems need to play a more active role in facilitating users effectively evaluating individual documents. In many cases, document problems are the main reasons for users quickly giving up their evaluation on individual documents. System design needs to focus on how to generate more relevant, accurate, updated and high quality documents and reducing the efforts users need to take in examining documents as well as offering translation tools for documents in foreign languages. For lengthy documents, information retrieval systems need to offer table of contents, best passages, abstracts, etc. for effective evaluation.

Conclusion

This study highlights the complex nature of individual document evaluation. While echoing previous research's findings on evaluation criteria, the authors discovered a richer blend of applied criteria beyond relevance. Similarly, the document elements evaluated offers a full complement of options, rather than the previous findings suggest. The time spent on individual document evaluation reveals the problems of information retrieval system design. Finally, the exploration of evaluation activities, in combination with criteria, elements and time spent for evaluation offers a strong foundation for information retrieval system design applications.

Although this study offers insight information to help further the understanding of document evaluation, it does have several limitations. While this study used real users with real problems, the data were not collected in real environments. The diary method recorded search activities in the real environment, the recorded data were not as detailed as the log data; therefore they were not used for this paper. Additionally, despite the wide variety of participants, the sample of 31 participants limits the generalizability of the findings and the identifications of relationships among different dimension of evaluation.

Future studies should explore the dynamic nature of evaluation, in particular increasing the sample of participants to represent different types of users with different types of tasks and further analyse the relationships among dimensions of evaluation. Additionally, further research is needed to better understand the role of interfaces on document evaluation, implement the design principles derived from user studies and further test the enhanced interfaces. Through expanding the understanding of how users evaluate, enhanced information retrieval systems can better connect real users with real problems to real solutions.

Acknowledgements

About the authors

Iris Xie is a Professor in the School of Information Studies, University of Wisconsin-Milwaukee. She received her Bachelor's degree in Library and Information Studies from East China Normal University, Master of Information Science from Shanghai Academy of Social Sciences, Master of Library and Information Science from the University of Alabama and her PhD from Rutgers University. She can be contacted at hiris@uwm.edu

Edward Benoit III is a Doctoral Student in Information Science at the School of Information Studies, University of Wisconsin-Milwaukee. He received his Bachelor's degree and Master of Arts in History and Master of Library and Information Science from the University of Wisconsin-Milwaukee. He can be contacted at eabenoit@uwm.edu.

Huan Zhang is a Master's Student in Library and Information Science at the School of Information Studies, University of Wisconsin-Milwaukee. She received a Bachelor's degree and Masters of Infromation Management from the School of Information Management, Wuhan University. She can be contacted at huan@uwm.edu.

References

- Bade, D. (2007). Relevance ranking is not relevance ranking or, when the user is not the user, the search results are not search results. *Online Information Review*, *31*(6), 831-844.
- Balatsoukas, P., Morris, A. & O'Brien, A. (2009). An evaluation framework of user interaction with metadata surrogates. *Journal of Information Science*, *35*(3), 321-339.
- Barry, C.L. (1994). User-defined relevance criteria: an exploratory study, *Journal of the American Society for Information Science*. **45**(3) 150-159.
- Barry, C.L. (1998). Document representations and clues to document relevance. *Journal of the American Society for Information Science,* **49**(14), 1293-1303.
- Barry, C.L. & Schamber, L. (1998). Users' criteria for relevance evaluation: A cross-situational comparison. *Information Processing & Management*, 34(2/3) 219-236.
- Borlund, P. (2003). The concept of relevance in IR. *Journal of the American Society for Information Science and Technology*, 54(10), 913-925.
- Cool, C., Belkin, N.J., Frieder, O. & Kantor, P. (1993). Characteristics of texts affecting relevance judgments. In M.E. Williams, (Ed.). *Proceedings of the 14th National Online Meeting*, (pp. 77-84). Medford, NJ: Learned Information
- Fitzgerald, M.A. & Galloway, C. (2001). Relevance judging, evaluation and decision making in virtual libraries: a descriptive study. *Journal of the American Society for Information Science and Technology*, **52**(12), 989-1010.
- Fogg, B.J., Soohoo, C., Danielson, D.R., Marable, L., Stanford, J. & Trauber, E.R. (2003). How do users evaluate the credibility of web sites? A study with over 2,500 participants. In *DUX'03 Proceedings of the Conference on Designing for User Experiences*, San Francisco. New York, NY: ACM Press.
- Greisdorf, H. (2003). Relevance thresholds: a multi-stage predictive model of how users evaluate information. *Information Processing and Management*, **39**(3), 403-423.
- He, D., Göker, A. & Harper, D.J. (2002). Combining evidence for automatic web session identification. *Information Processing & Management*, **38**(5), 727-742.
- Chi, E.H., Pirolli, P., Chen, K. & Pitkow, J. (2001). Using information scent to model user information needs and actions on the web. In *Proceedings of ACM CHI 2001 Conference on Human Factors in Computing Systems*, Seattle, WA, USA — March 31 - April 05, 2001. (pp. 490-497). New York, NY: ACM Press.
- Holsti, O.R. (1969). *Content analysis for the social sciences and humanities*. Reading, MA: Addison-Wesley.
- Jansen, B. & Spink, A. (2003). An analysis of web information seeking and use: documents retrieved versus documents viewed. *Proceedings of the 4th International Conference on Internet Computing, Las Vegas, Nevada. 23 - 26 June*, (pp. 65-69). Athens: CSREA Press.
- Jansen, B. & Spink, A. (2006). How are we searching the World Wide Web? A comparison of nine search engine transaction logs. *Information Processing & Management*, *42*(1), 248-263.
- Jansen, B., Mimi, Z. & Schultz, C. (2009). Brand and its effect on user perception of search engine performance. *Journal of the American Society for Information Science & Technology*, *60*(8), 1572-1595.
- Joachims, T., Granka, L., Pan, B., Hembrooke, H. & Gay, G. (2005). Accurately interpreting clickthrough data as implicit feedback. In *SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, (pp. 15-19). New York, NY: ACM Press.
- Kelly, D., Murdock, V., Yuan X. J., Croft, W. B. & Belkin, N. J. (2002). Features of documents relevant to task- and dact- oriented questions. In *Proceedings of the Eleventh International Conference on Information and Knowledge Management (CIKM'02)*, (pp. 645-647). New York, NY: ACM Press.
- Maglaughlin, K.L. & Sonnenwald, D.H. (2002). User perspectives on relevance criteria: A comparison among relevant, partially relevant and not-relevant judgments. *Journal of the American Society for Information Science and Technology*, **53**(5), 327-342.

Find other papers on this subject

Scholar Search   Google Search   Bing

Bookmark This Page