Chapter 2  Literature Review

This study reports on judgments of information quality and cognitive authority. To provide a general framework for the study, several questions must be addressed regarding previous related literature.

- How is judgment of information quality and cognitive authority related to relevance judgment?
- What is “quality” in general and “information quality” in particular?
- What is “authority” in general and “cognitive authority” in particular?
- Are there any empirical studies on evaluating quality and authority in the Internet environments?
- What are the evaluation criteria for Web sites identified in previous literature?

The first question is addressed in Section 2.1, where brief reviews on the nature of relevance and relevance criteria studies are provided. Sections 2.2 and 2.3 deal with characterizing the concept of information quality and cognitive authority by reviewing general literature on quality and authority. In Section 2.4, several empirical studies on quality and authority in different electronic contexts are summarized with a focus on diverse methodologies that they have taken. The last section, 2.5, reviews the literature on evaluating Web sites and evaluation criteria from three different perspectives: information professionals, users, and systems.

2.1  Relevance, Relevance Judgment, and Relevance Criteria

2.1.1  Nature and Dimensions of Relevance
Traditionally, people’s evaluations of information objects have long been discussed in the context of relevance. Since the early 1960s, relevance has been the most central and complex concept in information retrieval (IR). It is not necessary to review a whole body of relevance literature here because there have been a number of fine review articles in which the nature, dimensions, approaches to the study of relevance are discussed substantially (e.g., Mizzaro, 1997; Saracevic, 1975; Schamber, 1994; Schamber, Eisenberg, & Nilan, 1990). Therefore, the aim here is to examine important issues and trends focusing on recent literature on relevance judgment and relevance criteria.

Recently, Saracevic (1996b) has identified four frameworks which emerged over time: system (e.g., Harman, 1995), communication (Saracevic, 1975), situational (Schamber, Eisenberg, & Nilan, 1990), psychological (Harter, 1992). He then adds a fifth framework, an interactive framework. Saracevic’s interactive framework fits the approach of this study by viewing interactions as levels or strata (see Saracevic (1996a) and Saracevic (1997) for a stratified model of IR interaction). In this framework, he defines relevance as “an attribute or criterion reflecting the effectiveness of interactive exchange of information between people (i.e. users), and information systems in a communicative contact (p. 210).” He then suggests that there is not one relevance at play, but that there exists an interdependent system of relevances, highly interacting within and between different strata or levels. Since there is no information retrieval without such a system of relevances, according to Saracevic, a major issue of information science is to address the problems of understanding the effectiveness of interaction between different elements in the system of relevances.
Furthermore, Saracevic (1996b) categorizes “relevance manifestations” as follows: system or algorithmic relevance; topical or subject relevance; cognitive relevance or pertinence; situational relevance or utility; motivational or affective relevance. He claims that the manifestations interact dynamically within and between themselves. He offers some examples, saying that topical relevance is most often inferred on the basis of retrieved items, and cognitive and situational relevance follow from and interact with others. Motivational relevance may govern inferences in the others.

In a follow-up study, Cosijn and Ingwersen (2000) propose a modification of Saracevic’s model in terms of the five different manifestations. They claim that affective relevance is not the same as motivational relevance, and that motivational relevance is already included in the “intention attribute” of relevance. Furthermore, they argue that affective relevance should be viewed as part of relevance, which influences all other types of relevance, rather than as a discrete category of manifestations of relevance. They say “judging from current literature it seems that the level of influence of affective relevance differs from those of the other subjective relevance types (p. 14).” However, Cosijn and Ingwersen did not provide either empirical evidence or sufficient discussions as to why affective relevance should not be classified as a separate category of relevance. It seems that affective aspects of information seeking in general, and affective relevance in particular, has been found as a core aspect of the interaction process that users are experiencing (e.g., Kuhlthau, 1994). Finally, Cosijn and Ingwersen suggest that the modified model includes a socio-cognitive type of relevance that is highly context dependent and associated with organizational strategies.
There have been only a few empirical studies which identify the nature and dimensions of relevance from users. T. K. Park (1993) presents a model that reflects the nature of the thought processes of users who are evaluating bibliographic citations. She identifies three major categories that affect relevance assessments: internal context, external context, and problem context. According to her, the internal context refers to a user’s interpretation of a citation based on his or her own prior experiences or perceptions of the information problem. The external context is associated with an individual’s perceptions and situations in relation to the search process, the stage of research, and the priority of information needs related to the present search. The problem context is a content-oriented context, and indicates uses of information in relation to expanding one’s thoughts and constructing one’s ideas in the problem area. She claims that these three categories are not discrete; rather, they are interconnected and influence one another.

H. Park (1997) identifies “dimensions of relevance,” focusing on the relationship of three “orientations:” problem, use and value. He has found that the following relevance dimensions covered 60% of the total number of uses and 37.3% of the total number of users in his study: “similar, related, useful, helpful, relevant, and interesting.” His results also reveal that when the next six dimensions of “new, important, good, studied, needed, and applicable” were added to the top six dimensions, these accounted for more than 85% of the total number of uses and more than 65% of the total number of users.

2.1.2 Relevance Criteria

In the 1990s, substantial empirical studies have been carried out to identify user-defined relevance criteria in various information seeking environments including
academic (Barry, 1994; Cool et al., 1993; T. K. Park, 1993; Wang & Soergel, 1998; 
Wang & Soergel, 1999), weather-related (Schamber, 1991), and serious life/health-
related (Nilan, Peek, & Snyder, 1988) situations. According to Wang (1997), the 
common features across these studies in terms of data collection procedures are: (a) 
participants were end users; (b) data were collected in natural settings; (c) participants 
were not give a predefined set of relevance criteria, nor was the term “relevance” used in 
instructions to them; (d) user criteria were derived from content analysis of verbal or 
written reports; (e) models and grounded theories were proposed. The user criteria 
identified by these studies are presented in Table 1.

There are two ways to summarize the findings from previous studies. One is 
Wang’s (1997) attempt to look at whether these studies yield a compatible criteria set. 
Regarding content of information, quality is a prominent criterion of relevance judgment. 
Except Nilan et al.’s (1988) study, the users in the remaining four studies mentioned 
some aspects of information quality. Schamber’s (1991) occupational users of weather-
related information mentioned “accuracy” and “currency” which are components of 
used criteria such as “goodness” and “usefulness” which are clearly related to the concept 
of quality. Academic users (graduate students and faculty members) in three other 
studies mentioned “accuracy/validity (Barry, 1994),” “recency (Barry; Wang & Soergel, 
“expected quality (Wang & Soergel).” With regard to source of information, authority 
and related terms constitute most of relevance criteria. Nilan et al., Cool et al., and Wang 
& Soergel classified criteria under the category of “authority” while Schamber used the
term “reliability” and “verifiability,” which seem to be related to authority as she listed “expertise,” “source confidence,” and “source agreement,” for those two. Other common criteria are “accessibility” and “availability” which can be classified in environment category.

The other way is to look at whether there are criteria which appeared in only some studies. For example, Schamber and Cool et al. found that users mentioned a number of criteria related to “presentation” and “format.” This is because the users in these two studies mainly interact with full texts while users in other studies (e.g., Barry, 1994; T. K. Park, 1993; Wang & Soergel, 1999) were interacting with the representation of documents. Barry and Schamber (1998) identified similarities and differences in the criteria mentioned by users in their own findings. They attribute the differences of criteria identified by Barry and Schamber to “situational contexts” and “research task requirements.” They interpret that the respondents in Barry’s study used only print documents and those in Schamber’s study used information in a “multimedia environment” including sources from interpersonal communication, mass media, weather instruments and computer systems. The common criteria identified by Schamber and Cool et al. suggest that we can extend Barry and Schamber’s argument. Differences or variations in relevance criteria may be due to the types of information resources and systems with which the users are interacting beyond “situational factors” or “contexts.”
The importance of two common criteria across six studies, quality and authority, were confirmed by Bateman (1999). She has used a somewhat different methodology from other previous studies, in which 210 graduate students in a mail survey were instructed to rate 40 relevance criteria by importance. Using 11 criteria that survey respondents rated as most important, she develops a model which contains three constructs: information quality, information credibility, and information completeness. In her study, confirmatory factor analysis was used to confirm the reliability of the criterion terms or sub-scale scores in measuring the constructs. Bateman’s model suggests that the following criteria contribute to the construct of “information quality:” current, accurate, understandable, consistent, and focused. The construct of “information credibility” contains the following criteria: about my topic, well-written, credible. The last construct, information completeness is composed of the following criteria: comprehensive, suitably general or specific. Second-order factor analysis was then used to determine how much variance of the concept of high relevance could be explained by construct. She has found that these three constructs, information quality, information credibility, and information completeness are correlated with each other, and that they explain 48% of the respondents’ concept of relevance.

2.1.3 Relevance Judgment and Criteria for Judgment

Reviewing numerous studies on relevance shows that in general relevance studies can be broken into two categories: studies on the nature and dimensions of relevance (Section 2.1.1); and studies on relevance criteria (Section 2.1.2). While most studies on relevance criteria (e.g., Barry, 1994; Cool et al., 1993; Schamber, 1991) are based on empirical findings, most studies on the nature and dimensions of relevance are based on
theoretical discussions (e.g., Cosijn & Ingwersen, 2000; Saracevic, 1996b) with a few exceptions (e.g., H. Park, 1997; T. K. Park, 1993). It appears that there is a gap in the literature on the study of relevance. In general, researchers have paid less attention to understanding the nature of judgment while they paid much more attention to the nature of relevance and criteria of relevance.

Consequently, previous studies on the dimensions of relevance and relevance criteria have not taken account of the judgment process which can bridge the concepts of relevance and relevance criteria. This lack of the judgment process makes an unclear differentiation between users’ utterances about the aspects of relevance judgment or the factors influencing such a judgment. The “dimensions” of relevance identified by H. Park (1997) are also characterized as “criteria” of relevance in the previous studies. For example, the terms, “good,” “useful,” and “important,” that were characterized as “factors” by Cool et al. (1993) are identified as “dimensions” in H. Park’s study. Another example can be found in the terms “new” and “not of interest,” which were discussed as “criteria” by T. K. Park (1993), and called “dimensions” by H. Park.

In this study, the aspects of “judgment” and “criteria for judgment” will be differentiated. Schamber and Bateman (1996) provide useful definitions for relevance, relevance judgment, and relevance criteria. They define relevance as “users’ perceptions of the potential of certain information to resolve their problems in the context of their information seeking and use situations” (p. 218), and define relevance judgment as “users’ decisions to accept or reject specific information items at a certain time” (p. 218). According to Schamber and Bateman, relevance criteria refers to the factors that lead people to select (accept) one thing rather than another.
Schamber and Bateman’s definitions indicate that relevance deals with relationships between people’s information problems and information objects, while relevance judgment is associated with people’s decisions made in certain contexts of situation, task, and intention. Their definitions also indicate that judgment is a guide to behaviors. Let us take an example. If people think that they have found good information, which involves their judgment, they make a decision to select the information. This leads to their behavior, such as picking up a book, opening a Web page, following a hyperlink in a Web page. These behaviors can be observed externally while the judgment and decision processes are taking place internally. When people make a judgment and make a decision based on that judgment, there exists a criterion (or criteria) for their judgment and decision. These criteria are factors and reasons influencing their judgment and decision.

Returning to the examples of relevance dimensions and criteria identified in previous studies, it can be argued that the aspects or facets of judgment are related to values that people perceive. For example, the values such as good, useful, interesting, important, and new relate to evaluation and anticipation. Therefore, they were categorized as the facets of judgment. Other examples of the facets of judgment include reliable, credible, accurate, and authoritative. The next question then would be: what are the factors and characteristics that lead people to make such a judgment? This requires a set of criteria. For instance, a person would believe that this information is good because it is organized clearly. The criterion here is organization or structure. Another example can be found in which a person thinks that the medical information is useful because it is
written by a physician. Usefulness is an aspect of judgment, and author affiliation is a criterion.

2.2 Information Quality

2.2.1 What is Quality?

Garvin’s (1988) book provides general concepts and definitions of quality which can be applied to diverse disciplines beyond quality management. He identifies five principal approaches to defining quality: transcendent, product-based, user-based, manufacturing-based, value-based.

- The transcendent view understands quality as something timeless, absolute, and universally recognizable, an essence that rises above changes in taste or style.

- The product-based approach views quality as a precise and measurable variable. This approach lends a vertical or hierarchical dimension to quality, since “good” can be ranked according to the amount of the desired attribute they posses.

- The user-based definitions start from the premise that quality “lies in the eyes of the beholder.” This is an idiosyncratic and personal view of quality, and one that is highly subjective.

- The manufacturing-based definitions of quality are rooted in consumer preferences, and are primarily concerned with engineering and manufacturing practices.
- The value-based approach defines quality in terms of cost and price, blending two related but distinct concepts: excellence and worth.

The coexistence of these different approaches to quality often results in unclear, imperfect, and chaotic interpretations of the concept. This problem was well noted by Wormell (1990) who edited conference proceedings from the seminar about “information quality” sponsored by NORDINFO (Nordic Council for Scientific Information and Research Libraries). Although the focus of this seminar was in “quality management” of information in business organizations and public administrations, at least two interesting papers on definitions and applications of information quality are included in this proceedings (e.g., Merchand, 1990; Olaisen, 1990). These two papers will be discussed later.

In order to define the term, “information quality,” we need to establish similarities and differences between information quality and other related concepts such as database quality, data quality, and quality management. Garvin’s (1988) identification of five types of approaches to the concept of quality seems to provide a number of useful points for clarifying the interpretation of these terms.

In the transcendent approach, quality remains a very elusive concept. It is difficult to operationalize the concept of quality. Johannsen (1995) argues that this approach can be used in libraries in the selection and acquisition of materials such as fiction, art, and music, since it considers quality to be equal to “fine craftsmanship.” Thus, this approach contributes the “aesthetic aspect” of information quality. Aesthetics are also difficult to operationalize, but people will know that when they see it.
Product-based definitions of quality are different from transcendent, as quality can be something measurable. Differences in quality thus reflect differences in the quantity of some ingredient or attribute possessed by a product. This approach has been used when some researchers identify attributes of “database quality.” For example, Jasco (1995) explores three types of quality problems in the context of CD-ROM databases: incompleteness, inaccuracies, and inconsistency. It indicates that if a CD-ROM database product is more complete, accurate, and consistent, it has higher quality. Tenopir (1995) uses the term “content quality,” and discusses content quality in databases. She identifies five factors that affect content quality as follows: scope, structure, accessibility, accuracy, and consistency. With each factor, she defines some precise directions to take. For instance, she claims that there must be a clear editorial policy for scope, and there should be certain key fields which are searchable for structure. Her arguments indicate that if the database has a clear editorial policy, certain key fields, accessibility, accuracy, and consistency, then we can ensure “content quality.” It means that these five factors, as evaluation measurements, are the inherent characteristics of a product of database.

From the user-based perspective, the user is the “ultimate quality judge.” Here, it is assumed that individuals have different wants or needs, and the products or objects that “best satisfy their preferences are the ones they regard as having highest quality” (Garvin, 1988, p. 43). Garvin notes that a basic problem with the user-based approach is its equation of quality with maximum satisfaction. While the two are related, they are not identical. A product that maximizes satisfaction is certainly preferable to one that meets fewer needs, but is it necessarily better as well? The user-based approach is very complex since it is not a simple question of delivering the information the user requires.
Johannsen (1995) develops a taxonomy of quality in the user-based approach based on the distinction between requirements, needs, expectations, and excitement:

- **Expected quality:** emphasizes the expectations a user would take for granted and never mention unless directly asked.
- **One-dimensional quality:** refers to features actually mentioned by customers as being important for their degree of satisfaction.
- **Exciting quality:** is made up of the characteristics that a user does not require explicitly but that nevertheless will lead to satisfaction and success.

The fourth approach identified by Garvin (1988) is the manufacturing-based quality which is primarily concerned with processes. In this approach, quality means meeting specifications and “making it right the first time.” The manufacturing-based approach recognizes that a product or service that deviates from specifications is likely to be poorly made or unreliable, and it provides less satisfaction than one which is properly constructed or performed. Furthermore, according to the manufacturing-based approach, improvement in quality leads to lower costs, for preventing defects is viewed as less expensive than repairing or reworking them. According to Johannsen (1995), there were some works in the 1980s which emphasized methodologies for error and duplicate detection, automated authority control, and spelling correction in online databases. Johannsen notes that the quality concept behind these efforts is very similar to manufacturing-based views.

From the point of view of the value-based approach, a quality product is one that provides performance or conformity at an acceptable price or cost. Recently, Saracevic and Kantor (1997) conducted a study on the value of library and information services.
They define value as “an assessment by users of the qualities of an interaction with the service and the worth or benefits of the results of interaction, as related to the reasons for using the service” (p. 540). This definition indicates that the concept of quality is clearly related to the notion of value, and further to the “worth or benefits of the results.”

One thing we should note from the five types of quality approach is that there is no need for us to choose only one type of quality view to describe a product, service, or information object. These five types can coexist, and further can be integrated. For example, Marchand (1990) identifies 8 dimensions of information quality: actual value, features, reliability, relevance, learning over time, validity, aesthetics, and perceived value. He claims that the list of eight dimensions suggests why information quality is difficult to describe and measure, and why five different approaches are needed to define information quality. According to Marchand, the user-based approach focuses on perceived value and aesthetics, the product-based approach focuses on actual use, features, and meaning over time, the manufacturing-based approach on relevance and reliability, and the transcendent approach on meaning over time and validity.

Although there is no single conceptual framework for defining information quality, there is one for “data quality.” The difference between data quality and information quality lies in the system in which data or information is accumulated. Data quality is often discussed in the context of database systems which contain “highly structured and often numeric data with strict constraints on field formats, ranges, and size” (Fox, Levitin, & Redman, 1996, p. 115). On the other hand, information quality in this study has been discussed in the context of “information system” in which users are interacting with “texts” including images, sounds, text information. This difference is
important to note since in information quality there is much room for meaning and interpretation which is, by nature, subjective and user-based, while there is little room for it in data quality. Despite this difference, Fox, Levitin, and Redman’s framework for data quality is still worth analyzing.

Fox, Levitin, and Redman (1994) discuss the dimensions of data quality with respect to the data values in the following four categories:

- **Accuracy, precision, and reliability**: accuracy refers to the degree of closeness of its value to the correct value; precision refers to the measurement or classification used in specifying an attribute’s domain; reliability refers to the probability of correctness.

- **Currency, age, and timeliness**: A datum is said to be current or up-to-date at time $t$ if it is correct at time $t$. Currency can be measured by looking at how far out-of-date the datum’s value is.

- **Completeness and duplication**: A data collection can be incomplete because of missing values. Therefore, missing record incompleteness can be measured by finding the fraction of records missing from the data collection.

- **Consistency and integrity**: Correct data must be consistent, but consistency is necessary but not sufficient for correctness. Data integrity is sometimes used as a synonym for consistency, but sometimes it means data accuracy or correctness, and sometimes security and concurrency control.

Levitin and Redman (1995) discuss a list of characteristics (dimensions) that are crucial for data quality. In this model, they argue that when the characteristics of the
facts represented by the data have the following components, we believe the data (or facts represented by the data) has quality:

- **Content**: it should have relevance and obtainability of values
- **Scope**: it should be comprehensive and essential
- **Level of detail**: it should meet the level of attribute granularity and domain precision
- **Composition**: it concerns the structuring or grouping of facts represented by data
- **Consistency**: it should require semantic consistency and structural consistency
- **Reaction to change**: a good model must accommodate change

Fox, Levitin, & Redman (1996) are also concerned with quality dimensions of data representation. According to them, from a user’s perspective, the most important data representation issue is format. They claim that the format should be appropriate, unambiguous, universal, precise, and it also should make efficient use of the recording medium.

It is interesting to note that Fox, Levitin, & Redman’s quality dimensions of “data values” (accuracy, currency, completeness, consistency) have similarities with dimensions (facets) of judgment of information discussed in Section 2.1.3. This is because although data quality and information quality deal with different types of objects, in both frameworks, value is a key word which determines the dimensions.

### 2.2.2 Information Quality in the Value Added Model

Taylor’s (1986) concept of quality in his value-added model provides the most general framework which can be applied to the evaluation of information systems,
information, and data. His value-added model is based on the assumption that the judgment that users are making to choose particular information objects over others is giving value to the former, but not to the latter. That is, we are making judgments of value continuously while monitoring a variety of information systems and extracting what seems to be of value. To Taylor, the processes of storage and display on the system side and choice and use on the user side are throughout based on conscious and unconscious assumptions about the value of information.

Taylor’s value-added model is a three part information process: users, interface, and system. The user is defined as an agent who actively seeks information to achieve some objectives. According to him, the user has established a new information system which in miniature may have the same structure, values added, and other attributes of larger systems. He identifies six categories of user criteria of choice: ease of use, noise reduction, quality, adaptability, time-saving, and cost-saving.

Among them, Taylor defines quality as “a user criterion which has to do with excellence or in some cases truthfulness in labeling” (p. 62), and identifies five values included in quality: accuracy, comprehensiveness, currency, reliability, and validity. It is worth noting that quality values are for the most part intangible values. According to Taylor, “we tend to be suspicious of a system or a package which needs to advertise its reliability or its accuracy by words only” (p. 62). He further argues that these characteristics are earned over time and by reputation. His five quality values are defined as follows:
• Accuracy: the value added by system processes that assure error-free transfer of data and information as it flows through the system and is eventually displayed to a client.

• Comprehensiveness: the value added by the completeness of coverage of a particular subject or discipline (e.g. chemistry) or of a particular form of information (e.g. patents).

• Currency: the value added: (a) by the recency of the data acquired by the system; and (b) by the capability of the system to reflect current modes of thinking in its structure, organization, and access vocabularies.

• Reliability: the value added by the trust a system inspires in its users by its consistency of quality performance over time.

• Validity: the value added when the system provides signals about the degree to which data or information presented to users can be judged as sound.

Although Taylor’s concept and dimensions of quality are mainly based on an information system as a whole, it is inferred that these five values, accuracy, comprehensiveness, currency, reliability and validity are also associated with “data” or “information” or “outputs” from system performance.

2.3 Cognitive Authority

2.3.1 What is Authority?

The term “authority” is found in many different forms throughout society, cutting across many disciplines, including philosophy, education, psychology, political science, law, religion, and information science. De George (1976) discusses a number of different
types of authority. According to him, we refer to a person being an authority, by saying “He is an authority on Hegel;” we can speak of a person having authority, by saying “He has the authority to perform marriages;” we can speak of things embodying authority, when we speak of “the authority of the law.” He points out that in all cases “authority is either a relation or a relational quality attributable to a person or office or document or set of rules” (p. 77). In all cases there is a bearer of authority to those persons (or functions or things) from whom (or over which) he or she is the authority. De George notes that authority is not simply a diadic relation between the bearer and the subject. It is impossible to be an authority in general; authority is always and necessarily related to some field or area of competence or applicability over which the authority is exercised. Thus, “all authority is a relation among a bearer, a subject, and a field, in virtue of a particular quality, attribute, or context” (p. 77).

De George (1976) divides authority into two broad classifications: “epistemic authority” (declarative-emotive) and “deontic authority” (performatory), describing that the former corresponds to being “an authority” and the later refers to being “in authority.” He explains epistemic authority as one person being an authority for another in a given field in which he/she holds superiority over that other on knowledge. Regarding deontic authority, he distinguishes imperial authority, exercised by a state through its government and its various organs; paternalistic authority, exercised by a parent over his children; and, operative authority, vested in any designated leader or office by a group.

Watt (1982) claims that authority is always associated with inequality of some kind, and it is more obvious in authority of knowledge. Some people know more than others in knowing “how” as well as “about” something. According to Watt, it is often
rational to defer to the judgment of those who know more than we do, and irrational to
prefer our own judgment. What he calls “learned authority” is worthy of note. He
discusses that learned authority is specific to its subject, and normally an authority on a
subject is accepted by those who know that subject, as knowing even more than they do.
He describes a book as authoritative even if those who consult it make no use of the
information they find there. He also says that a person may be an authority on a subject
even if the advice he gives is not followed. According to Watt, learned authority answers
the question, “Why should I agree or believe?” and not the question, “Why should I
obey?” Watt’s concept of “learned authority” seems to be similar to that of “epistemic
authority” described by De George (1976).

The discussion of authority should not be confused with discussions of
authenticity. Judgments about authenticity are based on assessments of the origins,
completeness, and internal integrity of a document (Bearman & Trant, 1998). According
to Bearman and Trant, authenticity addresses: what the resource purports to be; how it
was created; whether it is unaltered from the original; and/or whether its representation is
transparent. Thus, it is grounded in methods of identification and verification of the
resource. In the context of information retrieval, users’ concern about authenticity is
“How do I know that it is what it purports to be?” rather than “Can I trust this
information” The latter question involves the judgment of authority.

2.3.2 Cognitive Authority in Second-Hand Knowledge

Wilson (1983) provides a useful basis for conceptualizing cognitive authority in
his book, Second-hand Knowledge. Wilson’s fundamental assumption in his concept of
cognitive authority is that we know of the world through two different ways: one way is
based on our first-hand experience and the other is based on what we have learned second-hand from others. Wilson discusses first-hand experience and second-hand knowledge as follows. What we can find out from first-hand experience itself depends on the stock of ideas we bring to the interpretation and understanding of our encounters with the world. If we had to depend entirely on ideas that we ourselves invested, we would make little sense of the world. As he describes, we mostly depend on others for ideas as well as for information about things outside the range of our own direct experience. Much of what we think about the world is what we have second-hand from others. He proposes that the phrase “second hand” is especially appropriate in suggesting second best, which is not as good as first hand, but quite good enough.

Wilson explains the concept of cognitive authority within the framework of two kinds of knowledge. According to him, all we know of the world beyond the narrow range of our own personal experience is what others have told us. It is all “hearsay.” But, we do not count all hearsay as equally reliable. Some people know what they are talking about, others do not. In Wilson’s terms, “Those who do [know] are my cognitive authorities.” Wilson notes that cognitive authority is different from the other kind of authority, that of the person who is in a position to tell others what to do. He calls it administrative authority. Administrative authority seems to be the same kind as “deontic or performatory authority” (De George, 1976). Wilson defines cognitive authority as “influence on one’s thoughts that one would consciously recognize as proper” (p. 15). Therefore, it is “legitimate influence.” He points out that cognitive authority is clearly related to credibility. According to him, “the authority’s influence on us is thought proper because he is thought credible, worthy of belief” (p. 15). He claims that the
notion of credibility has two main components: competence and trustworthiness. The cognitive authorities are clearly among those we think of as credible sources. In other words, “Those we think credible constitute the potential pool of cognitive authorities on which we might draw” (Wilson, p. 16).

Wilson further explains the “bases of authority.” According to him, when people acquire such influence from “cognitive authorities,” there should be answers for “what makes you think so?” Whatever our reasons for thinking the others deserving of cognitive authority, it is not that we have conducted a direct test of their knowledge. Rather, we have to cite indirect tests of indexes of credibility. He illustrates the situation in which we may be faced with a number of different people all claiming to be knowledgeable but all having different things to say on the same subject. Given that we ourselves are not knowledgeable on the subject, Wilson raises a question, “how can we choose among them, or how can we defend our choice once made?” (p. 21).

Let us take this situation to information retrieval. As Wilson points out, it is not only individuals in whom we recognize cognitive authority; we also recognize it in books, instruments, organizations, and institutions. That is to say that a newspaper, a magazine, a reference book can have cognitive authority independent of the particular individuals who produce the publication. Wilson discusses various external tests for recognizing the cognitive authority of a text as follows:

- Personal authority (author)
- Institutional authority (publisher)
- Textual type authority (document type)
- Intrinsic plausibility authority (content of text)
The first obvious consideration would be recognition of its author. According to Wilson, “we can trust a text if it is the work of an individual or group of individuals whom we can trust” (p. 166). He notes that the element of time affects people’s recognition of personal cognitive authority. That is, personal cognitive authority applies to a person “as he is now, at the time the tests of cognitive authority are applied: present reputation and accomplishments up to now” (p. 167). Wilson claims that an old reputation is not enough to establish the present authority of old texts. He then presents a question: what do we do when present reputation is unknown or irrelevant and we don’t have personal recommendation from others? He suggests that we may rely on a rule of recency: the newer the better, the older the worse.

The second sort of cognitive authority is associated with a publisher and publication history. For example, a publishing house can acquire a kind of cognitive authority as it is thought to be good at publishing their work. Therefore, publication by a house we respect constitutes a kind of personal recommendation. A single journal can have the same kind of authority, which transfers to the articles it publishes. He points out that other sorts of institutional endorsements are sometimes used as tests of authority: sponsorship of a publication by a professional organization; publication by a governmental agency or state printer; prizes and award given to the text or to its author for this text. Also, published reviews furnish a special indirect test. If a reviewer is known, and the reviewer has cognitive authority for us, his review constitutes a personal recommendation.

The third kind of cognitive authority can be found in the text itself. For example, some reference work such as a standard dictionary has cognitive authority. People are
not very concerned about the names of compilers in these reference books. A standard
reference work that is repeatedly revised may be thought of as an institution in its own
right.

Finally, Wilson addresses the recognition of a text’s contents as plausible or
implausible and which gives or withholds authority accordingly. He is particularly
concerned with the “instant recognition” of a work because “a text usually has one
chance to capture our attention; reading a few words of it may be enough to discourage us
from continuing on to reading the whole thing” (p. 169). If the sample of text we read
strikes us as nonsense, we are unlikely to continue; if it seems eminently sensible, we
may read on.

Wilson makes a couple of more interesting points about recognizing cognitive
authority in information retrieval, which seem to be applicable in information searching
in the Web as well. His discussions begin with the situation when we find a single source
that appears to tell us what we want to know. Then the question is: Do I need to look
further, or can I take this source as at least provisionally settling the matter? According to
Wilson, if we are already convinced of the authority of the source (i.e., it is the
dictionary), the question is already answered, but if we are unfamiliar with the source, the
question is likely to arise explicitly. Wilson suggests that we need not only to find
reasons for taking the single source seriously but also reasons for thinking that there are
no other sources deserving to be taken still more seriously.

Another example that he discusses is related to the situation in which we do not
find a single source that appears to tell us what we want to know. Then we may have to
search for a number of texts from which we can find what we seek. Finding the right text
is neither simple nor straightforward. As Wilson notes, a text can be of high or low quality in many different ways: well and clearly written but unfortunately inaccurate; imaginative and stimulating but unsound; and so on. According to Wilson, the chief aspect of quality is *credibility*: “can one believe what the text says, or can one at least take it seriously?” (p. 171). He argues that other good or bad points about the text are of subordinate interest. He also claims that the question of cognitive authority can be rephrased as one of quality control.

Wilson’s discussions on the relations of several crucial concepts can be summarized as follows. Cognitive authority is a kind of influence, but not just any kind of influence; but influence on one’s thoughts that one would recognize as proper. The authority’s influence is thought proper because he/she is thought credible. Credibility, composed of competence and trustworthiness, can be considered a potential pool for cognitive authority, and the chief aspect of quality. Cognitive authority, consequently, is one of the quality control components in information retrieval.

### 2.4 Empirical Studies on Quality and Authority

There are only a few empirical studies which directly address the issues of quality and authority in information use environments. Olaisen (1990) investigates the use of electronic information with respect to quality and cognitive authority among managers in Norwegian financial institutions. Klobas (1995) examines the Internet use associated with information quality and usefulness among faculty and staff members in Australian universities. Jeong’s (1998) study is the most recent and comprehensive one in which she develops and tests a framework for measuring information quality on lodging Web sites. Each study will be discussed in detail below.
Olaisen (1990) analyzes the use of electronic information in a business setting to propose a model for determinants of information quality. Based on a literature review on philosophical consideration, including Wilson’s (1983) conceptual framework of cognitive authority, Olaisen divides quality factors into two categories: cognitive authority factors and technical user-friendliness factors. He defines “cognitive authority” quality factors as something dependent on how the information is perceived by users. “Technical user-friendliness” quality factors are dependent on what the user is offered.

Based on questionnaires with 327 responses and interviews with 50 managers, Olaisen (1990) has found that: (a) the “knowledgeable person” is the most important source for both daily administrative decision and strategic long-term decisions due to credibility, influence, reliability and relevance; (b) electronic information is emerging as an important source, scoring highly in relevance, perceived value, accessibility, actual value, flexibility and browsing possibilities, while scoring low concerning credibility, form, and user-friendliness; (c) the more experience the user has with computers and telecommunications the more she/he will be able to use different sources (i.e. printed, electronic, internal, and external source); (d) bank managers believe most in information produced by banks or credit-evaluation companies.

Olaisen (1990) is concerned with how people “rank” the quality factors in the two categories (cognitive authority and user-friendliness) of information quality. In the category of cognitive authority factors, the respondents ranked: (a) credibility, (b) reliability, (c) relevance, (d) validity, (e) meaning over time. In technical user-friendliness factors, the respondents ranked the following: (a) accessibility, (b) form, (c) flexibility, (d) actual value, (e) selectivity, (f) completeness.
Klobas (1995) has conducted an empirical study in which she examines the Internet use associated with information quality and usefulness in three Australian universities. She reports survey results from 299 faculty and staff members on the factors that influence the use of CWIS (Campus-Wide Information Systems). The factors used in Klobas’s study are organized as follows:

- Information quality: accuracy, currency, novelty, and authority
- Ease of use: ease of use, interface, and ease of learning
- Convenience: ease of access, availability, and convenience
- Usefulness: quality of work, control of work, work performance, speed of work, ease of work, relevance, and usefulness

A seven-point scale was used to measure the factors in terms of “importance” and “perceived quality.” The analyses of data show that the respondents believe that accuracy (M=6.55) and currency (M=6.34) are more important than novelty (M=5.69) and authority (M=5.54). On the other hand, when they were asked about perceived quality, novelty (M=5.09) and currency (M=5.08) are scored higher than accuracy (M=4.77) and authority (M=4.77). It can be assumed that authority scored lowest in both importance and perceived quality categories because respondents speak of internal campus information systems. The results reveal that information quality has no direct influence on intended use, however it has a strong influence on perceived usefulness (therefore indirect influence on intended use) whereas technical quality (ease of use) has no significant direct or indirect influence on intended use.

Recently, Jeong (1998) examined people’s perception toward information quality based on hypothetical lodging Web sites. She develops four constructs consisting of
perceived usefulness, perceived ease of use, perceived accessibility, and attitude in order to measure information quality. Eight hypothetical Web sites were developed, and each site was designed differently to meet the attributes of each constructed at very high and a very low level. 240 subjects (conference attendees) were randomly assigned to one of the eight Web sites. Each subject completed a questionnaire related to Web use, information needs, and evaluation of a lodging Web site.

Jeong found that users’ perceived usefulness of the lodging Web site had significant mean differences in intention to use information, actual information use, and recommendation to family or friends. The results of her study also show that users’ perceived ease of use was significantly different from intention to use information and information use. Their attitudes toward the lodging Web site had significant mean differences on intention to use information and recommendation. Jeong concludes that perceived ease of use and attitudes were found to be significant indicators to predict the users’ purchasing behavior using the lodging Web site.

The three empirical studies reviewed above collected data through self-administered survey questionnaires. Olaisen (1990) used a semi-structured interview which is somewhat different from the other two studies. However, he also had a pre-defined set of quality factors before he collected the data, and simply asked the users to rank them. Simply speaking, the authors in these previous studies identified potential factors based on the literature, and asked users to respond to “information use” with respect to information quality. For example, Klobas’s questionnaire used importance ratings in “usefulness,” “convenience,” “information quality,” and “ease of use.” Jeong also asked the users to rate the importance of each item including: “guaranteed accurate
information,” “timely information,” “helpful guidelines to navigate the site,” “easy to read information,” and “quality appearance.”

In these studies, a limitation is that users are perceived as passive recipients of objective information, answering “yes” or “no” as to whether they use electronic information (in the Internet or the Web), trust information, perceive quality, and so on. This assumption corresponds to the “traditional” information retrieval model in which users’ interaction with information is not reflected. As Belkin (1993) points out, the traditional model ignores users’ various activities which are dependent upon people’s goals, intentions, knowledge, and other information problem situations. Information retrieval interaction models (e.g., Belkin, Marchetti, & Cool, 1993; Belkin, Cool, Stein, & Thiel, 1995; Ingwersen, 1996; Saracevic, 1996a; Saracevic, 1997), which were proposed as an “alternative” view, perceive users as active seekers of information. Users look for information of potential interest, make judgments about usefulness or interest of information by engaging with it, and interpret information. These processes were not considered in studies of Klobas (1995), Jeong (1998), and Olaisen (1990). Since information quality and cognitive authority are related to judgments that users made, an alternative approach in which users’ interaction with information is reflected is needed.

2.5 Evaluation Criteria for Web Pages

2.5.1 Criteria from the Perspectives of Information Professionals

A number of librarians and researchers in library and information science have been concerned with the problem of evaluation of information resources in general and information quality in particular on the Web. The approaches taken in these studies are different from those in this study in two ways. First, the approaches taken in the previous
studies are related to teaching critical evaluation skills for library users by developing guidelines from bibliographic perspectives as opposed to understanding judgments of users from the perspectives of information retrieval interaction employed in this study. Second, the previous studies discuss evaluation criteria for Web sites as a whole rather than individual information items on the Web. In this study, the focus is on an individual Web page that users are interacting with. However, since information quality and authority are the primary concern of evaluating Web sites in these “guidelines,” it is worthy of reviewing how “information professionals” identified criteria for evaluating Web sites. For example, Smith (1997) states that while most authors on evaluation of Internet resources agree that traditional criteria apply, there are particular aspects and concerns in the Internet environment. He notices that authority is of particular concern. He also says that the problem in evaluating authority on the Internet is even more difficult since there is no “metainformation” such as introductory notes for scope on the Internet.

While discussing the challenges of Web use in the context of library instruction, Arnold and Jayne (1998) have noticed that students in general subscribe to the myth that computers have more authority and that information from a machine is superior to that found in a printed source. Consequently, they are often willing to accept the information without questioning the source. Despite the lack of quality control of the Web, students expect it to have better and more current information "on everything," and obtain immediate gratification, since "you get something on the screen right away," (p. 46) plus glitz and graphics. Arnold and Jayne emphasize that it is important to raise students'
awareness of the filtering processes of evaluation and selection that occur behind the scenes for library materials, which is not present on the Web.

Most of the literature written by information professionals shares similar objectives. The purposes include offering indicators of “how to make a quality assessment when potentially relevant materials have been located” (Cooke, 1999, p. 3). In her recent book, Cooke attempts to provide “extensive and detailed guidelines on the wide range of factors which could influence users’ perceptions of the quality of the sources they find” (p. 16). The assumptions underlying the literature on evaluation criteria include:

- The Web has a lot to offer, but not all sources are equally valuable or reliable; so, think critically about Web resources (Grassian, 1998).
- Users may lack the knowledge and skills to make an assessment of quality (Cooke, 1999).
- Librarians, who have provided filtering for information in the library, seek to adapt traditional information evaluation techniques to the Internet environment (Bradt, 1996).
- Libraries now maintain Web sites that have lists of Internet information resources; without explicit criteria for selection, libraries risk wasting their time (Smith, 1997).
- Scholars in most fields have evaluated print information, and the same criteria can be used to assess information found on the Internet (Kirk, 1999).
Table 2. Dimensions and Criteria of Web Site Evaluation

<table>
<thead>
<tr>
<th>Dimensions of evaluation</th>
<th>Evaluation criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accessibility of a source(C)</td>
<td>Audience(K)</td>
</tr>
<tr>
<td>Accuracy (C, S, T)</td>
<td>Cost and copyright(P)</td>
</tr>
<tr>
<td>Accuracy and balance of content(W)</td>
<td>Cost(K, S)</td>
</tr>
<tr>
<td>Accuracy/Reliability(K)</td>
<td>Coverage(C, T)</td>
</tr>
<tr>
<td>Authority and reputation(C)</td>
<td>Format/Access/Retrieval approaches(K)</td>
</tr>
<tr>
<td>Authority of author(W)</td>
<td>Graphic and Multimedia design(S)</td>
</tr>
<tr>
<td>Authority(S, T)</td>
<td>Hardware and software(P)</td>
</tr>
<tr>
<td>Authority/Credibility (K)</td>
<td>Information structure and design(W)</td>
</tr>
<tr>
<td>Browsability and organization(S)</td>
<td>Links made to other resources(S)</td>
</tr>
<tr>
<td>Comparison with other source(C)</td>
<td>Presentation and arrangement of information(C)</td>
</tr>
<tr>
<td>Connectivity(S)</td>
<td>Purpose (K)</td>
</tr>
<tr>
<td>Currency and maintenance of a source(C)</td>
<td>Purpose and audience(S)</td>
</tr>
<tr>
<td>Currency(S, T)</td>
<td>Purpose of a source (C)</td>
</tr>
<tr>
<td>Currency/Meaning over time(K)</td>
<td>Required computing environment(S)</td>
</tr>
<tr>
<td>Ease of use (C, P)</td>
<td>Resource identification and documentation(W)</td>
</tr>
<tr>
<td>Interactivity(S)</td>
<td>Reviews(S)</td>
</tr>
<tr>
<td>Navigation within the document(W)</td>
<td>Scope (S)</td>
</tr>
<tr>
<td>Objectivity(T)</td>
<td>Scope/Comprehensive of Coverage/ Uniqueness(K)</td>
</tr>
<tr>
<td>Overall quality of a source(C)</td>
<td>Searching(S)</td>
</tr>
<tr>
<td>Quality and content (P)</td>
<td></td>
</tr>
<tr>
<td>Quality of links (W)</td>
<td></td>
</tr>
<tr>
<td>Quality of writing(S)</td>
<td></td>
</tr>
<tr>
<td>Relevance and scope of content(W)</td>
<td></td>
</tr>
<tr>
<td>Relevancy (P)</td>
<td></td>
</tr>
<tr>
<td>Relevanace/Value(K)</td>
<td></td>
</tr>
<tr>
<td>Reliability and stability(P)</td>
<td></td>
</tr>
<tr>
<td>Site access and usability (W)</td>
<td></td>
</tr>
<tr>
<td>Uniqueness(S)</td>
<td></td>
</tr>
<tr>
<td>User friendliness(S)</td>
<td></td>
</tr>
<tr>
<td>Validity of content(W)</td>
<td></td>
</tr>
</tbody>
</table>

Note: The original literature included here did not differentiate judgment from criteria. All of them were listed under the label of evaluation criteria; C = Cooke (1999); K = Kjartansdottir & Widenius (1995); P = Pratt et al. (1996); S = Smith (1997); T = Tate & Alexander (1996); W = Wilkinson, Bennett, & Oliver (1998).

Table 2 presents evaluation criteria identified in the previous studies (Cooke, 1999; Kjartansdottir & Widenius, 1995; Pratt et al., 1996; Smith, 1997; Tate & Alexander, 1996; Wilkinson, Bennett, & Oliver, 1998). They were all claimed as evaluation criteria, which can be applied by librarians selecting Internet information sources for a resource guide (Smith), by the authors. Following the discussions made in Section 2.1.3, in which the researcher in this study argued that the aspects of judgment and factors (criteria) for judgment should be differentiated, Table 2 contains two
categories of Web site evaluation: dimensions of evaluation and criteria for evaluation. While the dimensions of evaluation involve a number of different value judgments, criteria for judgment refers to various factors that influence their value judgment.

The following 5 categories appear more than two times from six different studies with regard to the dimensions of evaluation.

- Accuracy, accuracy/reliability, accuracy and balance of content (in 5 studies)
- Authority, authority and reputation, authority of author, authority/credibility (in 5 studies)
- Currency, currency and maintenance of a source, currency/meaning over time (in 4 studies)
- Quality and content, quality of links, quality of writings, overall quality of a source (in 4 studies)
- Relevancy, relevance/value, relevance and scope of content (in 3 studies).

It is particularly interesting to note that the authors in previous studies describe “quality” in diverse contexts including source, content, links, and writing. Authority is also discussed with respect to different facets such as reputation, author, and credibility. This indicates that quality and authority are the two most multiple and complicated concepts in the evaluation of Web sites.

The following evaluation criteria appear two times among six studies summarized in Table 2: audience, cost, coverage, purpose, and scope. It seems that there is less consensus in evaluation criteria than in dimensions of evaluation. While some dimensions of evaluation repeatedly appear in four or five studies, these common criteria emerged only in two studies. As discussed earlier in this section, criteria identified in the
previous studies are based on evaluation of a Web site as a whole. It becomes obvious when we look at the criteria such as purpose, scope, and coverage. Therefore, there is a need to identify evaluation criteria based on individual items retrieved in the Web site. The criteria that users use for evaluating individual items could be different from those used in evaluating a Web site.

It should be noted that all the dimensions of evaluation and criteria of evaluation discussed above are “proposed” ones. This means that the list is based on the authors’ suggested dimensions and criteria, and not verified based on empirical findings. The evaluation criteria studies conducted with real users will be discussed in the next section.

2.5.2 Criteria from the Perspectives of Users

There has been little research on user-based criteria for evaluating Web sites while there are a number of studies on criteria suggested by professionals as discussed in the previous section. Only two of them carried out by Abels, White, and Hahn (1997, 1998) are empirical studies. In Phase I (1997), Abels, White, and Hahn identify the criteria that influence the use of the Web and analyze them within the context of users’ overall information-gathering behavior. In Phase II (1998), they implement user-based design criteria in a Web page.

In their Phase I study, Abels, White, and Hahn (1997) collected data from 9 faculty members from business programs through a questionnaire distributed electronically and a half-day focus group session held in an electronic environment. Through “brainwriting,” the participants identified positive features and negative features of Web pages which would influence subsequent use or non-use of Web pages. The items were then ranked within the positive and negative groups according to the average
ranking. The median rank was used to determine the most influential positive factors and negative factors. Six clusters emerged in the analysis: appearance, content, linkage, special feature, structure, and use.

What makes Abels, White, and Hahn’s (1997) research interesting is that they have identified both positive and negative features of Web pages for each cluster. For example, with respect to “content,” usefulness of information, currency of information, concise information are the features considered to make a decision to use. When users found superficial information, uninformative content, advertisements, repetitious content, boring text from a Web page, these factors negatively influence decision, and lead them not to use a page. Regarding the “structure” category, the users in Abels et al.’s study considered positive that they found a Web page is well organized, straightforward with an intelligible structure including the ability to get an overview of the structure, innovative presentation and organization, they were likely to use this Web page. On the other hand, if they found a Web page has long text and poorly named subsections, they were not likely to use it. The authors also identified types of information normally used in work activities, and methods of finding this information within the current information service environment.

In Phase II of study, Abels, White, and Hahn (1998) implement a test site for an academic business community. As an alternative to existing “ad hoc” design processes, the authors developed a four-stage user-based design process, composed of an information-gathering stage, a development stage, an evaluation stage, and an implementation stage. The designers used the criteria identified in the Phase I study, which corresponds to the information-gathering stage. In the development stage, the
criteria need to be operationalized, and ranked based on feasibility. The designers translate the criteria into Web site features. According to the authors, for example, “content” was rated highly by the focus group. A brainstorming session took place on content, and it permitted the construction of categories of content that would be useful for the development of the Web site. The authors note that authority is intertwined with content and is implicit. They describe that “when a user states that information must be useful, they are referring not only to topic coverage but also to the source or producer of the information” (p. 42). Their “development” plan for each user criteria category can be summarized as follows:

- Content: users like useful and current information which has novelty, in the sense that information is not easily accessible elsewhere.
- Structure: users expect a clear structure in which categories are easy to understand rather than incidental or arbitrary.
- Linkage: users want to explore and discover information serendipitously; therefore linkage should put together related information.
- Use: users expect that the content indicators and organizing structures should be integrated into the navigational system so that they can be easy to use.
- Search: users want responsive and effective searching.
- Appearance: users consider unattractive graphics, excessively numerous graphics, over-large graphics, and dark graphics all annoying.

Following the development stage, the evaluation stage will be started as the site is being developed initially. In the evaluation stage, the design is quite flexible and evaluation efforts can be targeted to specific content and/or design features. Then, as the
site becomes operational and available, real users begin to incorporate the site into their normal information behavior, which is what they call the implementation stage. According to the authors, implementation does not indicate the end of the process. Evaluation and feedback are incorporated into the implementation phase so that users can make suggestions about new resources, organization, links, or appearance. As their project is in the stage of evaluation, specific results about evaluation and implementation were not be reported. Abels, White, and Hahn’s two-part project is a good exemplary study which shows a way of integrating user-based criteria with system development and implementation. It involves operationalizing definitions of the criteria and translating the criteria into Web page features. During the design process, the authors emphasized the importance of having user input at several different stages.

2.5.3 Criteria from the Perspectives of Systems

Recently researchers in computer science have started paying attention to the problems of quality and authority with respect to the effectiveness of a search engine (e.g., Amento, Terveen, & Hill, 2000; Kleinberg, J. M., 1999; Price & Hersh, 1999; Zhu & Gauch, 2000). The common problems that they discovered are: (a) the number of relevant pages in the Web is too large for a user to digest; (b) Web documents vary widely in quality and authority; (c) a user needs a way to filter information from a huge collection of relevant pages.

Kleinberg (1999) is one of the first authors who suggested a way to “filter” a small set of the most “authoritative” or “definitive” pages from a collection of relevant pages. He begins his discussion with the problem of current text-based ranking methods. The example that he has taken is that although www.harvard.edu is the most authoritative
page for the query “harvard,” www.harvard.edu is not the one that uses the term harvard most often. He claims that analyzing the hyperlink structure among Web pages provides a way to address these difficult problems. According to Kleinberg, hyperlinks encode a considerable amount of latent human judgment, and this type of judgment is precisely what is needed to formulate a notion of authority. In other words, the creator of page \( p \) has some “conferred authority” on page \( q \), and decides to include a link to page \( q \). His model is based on the relationship that exists between the authorities for a topic and those pages that link to many related authorities (hubs). He claims that a good hub is a page that points to many good authorities; a good authority is a page that is pointed to by many good hubs.

The algorithm devised by Kleinberg has been implemented in “Clever,” a prototype search engine developed at IBM (Chakrabarti, et al., 1999). According to the members of the Clever project, for any query of a topic, Clever first obtains a list of 200 pages from a standard text index such as AltaVista. The system then augments these by adding all pages that link “to” and “from” the 200 pages. The result collection typically contains between 1000 – 5000 pages. For each of these, Clever assigns initial numerical hub and authority scores. The system then refines the values: a page that has many high-scoring hubs pointing to it earns a higher authority score; a location that points to many high-scoring authorities garners a higher hub score. Clever repeats these calculations until the scores have more or less settled on their final values. According to the project members, the results are generally independent of the initial estimates of scores used to start the process. In addition, the members claim that Clever’s algorithm reveals the underlying structure of the Web.
The Clever project can be compared with the Google search engine which also implements a link-based ranking measure. Brin and Page (1998) state that the Google search engine has two important features that produce high precision results: it makes use of the link structure of the Web to calculate a quality ranking for each Web page (PageRank); and it utilizes links to improve search results. The core idea of Google is PageRank. It is related to the notion that we have in academic literature: to recognize the importance and quality of a paper, we might count citations to the paper. They extended this idea by not counting links from all pages equally, and by normalizing the number of links on a page. The main difference between Clever and Google is that Google’s basic philosophy is to look only in the forward direction, “from link to link.” In contrast, Clever also looks backward from an authoritative page to see what locations are pointing there (Chakrabarti et al., 1999).

The research by Kleinberg and members of the Clever Project certainly has a number of interesting points in terms of link-based analysis in general and particularly the relationship between a set of relevant authoritative pages and a set of “hub pages.” What they overlook, however, is that authority is a relation among a bearer, a subject, a field, in relation to a particular quality, attribute, or context (De George, 1976). For instance, a document which is authoritative to user A might not be so to user B because A and B have different knowledge levels. Another example is that a document P from a famous computer company can be recognized as an authoritative information source when a user is looking for information on the topic of computer; however, the very same document might not be recognized so for the task of medicine. Kleinberg and members
of the Clever Project do not discuss how they would deal with the problems of subjectivity and the situational factors of authority.

Amento, Terveen, and Hill (2000) believe that link analysis algorithms have the potential to solve the problem of identifying a small subset of documents worthy of quality in the Web. In this sense, their study is in the same line with Kleinberg’s (1999) and Brin and Page’s (1998) studies. The difference is that Amento et al.’s research focuses on experimental evaluations of link analysis algorithms. The topics used in their study were related to popular entertainment TV shows. Their study was conducted in two phases: the first phase was a study with 40 users; the second phase was a study with 16 experts. In the first phase, the task was to select the 15 “best” items for the topic. The “best” was defined as a set of items that collectively provided a “useful and comprehensive overview.” The items selected by the subjects were used as an initial set. In the second phase, the final items obtained were evaluated by experts on a scale from 1 (worst) to 7 (best). Then these items were used to test how well the rankings induced by each of the metrics (both content-based and link-based) matched expert quality judgments. In other words, the authors defined the “good items” as those that a majority of experts rated as good.

With regard to link-based features, Amento, Terveen, and Hill computed five metrics: in and out degree (Kleinberg, 1999), authority and hub scores (Kleinberg), and the PageRank score (Brin & Page, 1998). Regarding content-based features, page size and the number of images and audio files are recorded. What they found was that five link-based metrics are more or less similar in terms of precision, particularly in-degree and authority had pretty similar precision. They also found that simply counting the
number of pages on a site gives as good an estimate of quality as any of the link-based computations. In other words, according to them, the number of pages on a site is an indicator of higher quality. It appears that their results do not support the arguments made by Kleinberg’s and Brin and Page’s research.

On the other hand, Price and Hersh (1999) focus on quality of pages rather than authoritative information sources in the Web. They explore the idea of using automatic filtering techniques to identify pages likely to be of “high quality.” They have developed a prototype system that responds to a health query by returning a list of Web pages that are ranked according to the likely “quality of the page contents.” Their system examines Web pages and assigns a score that indicates the likelihood that each page will meet quality criteria such as relevance, credibility, absence of bias, content, currency, and value of links. Each indicator of quality is calculated as follows:

- Likely relevance: the algorithm is designed to discover pages that are less likely to be useful, such as chat room or bulletin board postings.
- Likely credibility: The module contains subroutines that inspect the URL, look for authorship of information, determine whether the site displays the HONcode logo, and search for particular words or phrases, such as “miracle cure.”
- Likely bias: The module identifies specific words and phrases such as “mastercard” and “visa” that suggest a commercial basis may be present.
- Content: The program determines how much text is displayed, and the ratio of hyperlinks to text.
• Currency: The algorithm searches for evidence indicating when the page was published or last updated.

• Value of links: The algorithm calculates an average link score for each page.

According to Price and Hersh, their software tool successfully separated the desirable from the undesirable pages, assigning higher scores to the “desirable pages” than to the undesirable pages. They then claim that preliminary results demonstrate that given a set of criteria to evaluate the quality of consumer health information, it is possible to automatically rank Web pages in order of likely quality.

Zhu and Gauch (2000) describe an approach that combines similarity ranking with quality ranking in three different search environments – centralized retrieval, distributed retrieval, and information fusion - in the Web. The quality metrics selected were currency, availability, information-to-noise ratio, authority, popularity, and cohesiveness. These metrics were measured as follows:

• Currency: the time stamp of the last modification of the document.

• Availability: the number of broken links on a page divided by the total numbers of links it contains.

• Information-to-Noise Ratio: the total length of the tokens after preprocessing divided by the size of the document.

• Authority: the Yahoo Internet reviews which assign a score from 2 to 4.

• Popularity: the number of links pointing to a particular Web page obtained from the AltaVista site.

• Cohesiveness: determined by how closely related the major topics in the Web page are.
The results from their experiments revealed that four of the metrics - currency, availability, information-to-noise ratio, and cohesiveness - significantly improved the search effectiveness in centralized information retrieval up to 24.8%. When the quality metrics were incorporated in site selection (distributed retrieval), availability, information-to-noise ratio, popularity, and cohesiveness significantly improved the precision. With respect to the information fusion experiment, only the popularity metric was significant. Zhu and Gauch noticed that the authority metric did not have a significant impact on either centralized search, site selection, or information fusion. They argue that this is because the authority metric is not related to search effectiveness at all, or the authority ratings they used are not accurate enough.

The limitations of Price and Hersh’s study as well as Zhu and Gauch’s study are associated with internal validity, which often refers to the extent to which the study accurately measures what it purports to measure. The first problem is associated with the measurement of quality: on what basis do the authors in each study select criteria (indicators, metrics) for assessment of quality? The criteria used in their studies are neither grounded in empirical findings nor based on exhaustive literature review. The second problem is related to the measurement of each criterion. For instance, to measure “content,” Price and Hersh use a program which determines “how much text is displayed, and the ratio of hyperlinks to text” (p. 913). Although this may be applicable in some cases of Web searching, this definition does not seem to be sufficient to deal with various cases of evaluation based on content.

Another example of quality indicators which seems to be problematic is Zhu and Gauch’s measurement of “authority” which was based on the Yahoo Internet Life (YIL)
reviews assigning a score range from 2 to 4 to a reviewed site. The most serious limitation of the measurement of authority is that if a Web site was not reviewed by YIL, its authority was assumed to be 0. This method of scoring authority does not appear to be appropriate. That’s because, as discussed in Section 2.3, authority is a much more complicated, dynamic, situational, subjective, and relative concept which cannot be relied on by only one review site.

The limitations and problems found in these two studies have close ties to the differences between the approaches taken in the previous studies and those in this present study. The authors in previous studies view quality and authority as objective, external and static concepts, whereas this study understands them as highly subjective, internal, and dynamic notions. Furthermore, the previous studies attempted to solve the problems of quality and authority in the “Web documents” from the perspective of systems. Thus, their approaches are closely associated with search effectiveness and precision of the results. The present study deals with the problem from the perspective of information interactions of users. Therefore, we focus on the process of judgment making about quality and authority based on the users’ own terms in Web searching environments.

2.6 Summary

There has been a substantial body of empirical studies in the 1990s which have revealed that people use diverse criteria beyond topicality to make a judgment of relevance. On the one hand, the relevance criteria identified in these studies yield a compatible criteria set, although users made relevance judgments in different information seeking environments. On the other hand, there are some differences and variations in
relevance criteria across the six studies reviewed, and they are due to the types of information resources and systems with which users are interacting.

One of the problems in these previous studies is that the authors in the studies did not differentiate the aspects of relevance judgment and factors influencing such a judgment. It is argued that these two processes need to be differentiated: aspects or facets of judgment are related to values that people perceive; factors and characteristics that lead people to make a such a judgment provide the bases for the judgment, and therefore they are criteria for the judgment.

It is hard and almost impossible to suggest one clear definition for quality. Multiple types of quality related to a product, service, and information objects, can coexist and further be integrated. Taylor’s (1986) definition and dimensions of quality in his value-added model can be used to conceptualize information quality. In his model, users choose particular information objects over others because they give value to the former, but not to the latter. Following this notion, one can argue that users make judgments of quality because they give more value of quality to one object than another. The kinds of values users can apply are identified by Taylor as: accuracy, comprehensiveness, currency, reliability, and validity. These five values were defined as “criteria” in the previous relevance criteria studies, but they will be characterized as “dimensions of judgment” in this study as they are related to the values users could perceive rather than to the factors influencing judgments.

Authority, too, is a difficult concept to give one single definition. In fact, there are several different kinds of authority whose differences seem to be important to note. Cognitive authority is a term coined by Wilson (1983). Wilson’s conceptualizations of
quality, credibility, and cognitive authority indicate the following sequences of
information seeking behaviors. When people look for information, they are interacting
with texts, books, magazines, or systems. Each information medium has own quality
control mechanisms. As a result, there are a number of aspects of quality control which
people can see: a text is clearly written; it is current; it is accurate. The aspects of quality
are not always consistent. For example, sometimes a text can show high quality in
currency, but low quality in accuracy. In this case, people rely on the chief aspect of
quality, which is credibility. Therefore, when people look for information, they tend to
look for the credibility of the source. If they are not familiar with the source, they
conduct several indirect tests to recognize the authority of source in terms of personal
authority, institutional authority, textual type authority, and intrinsic plausibility. These
four tests are the bases or criteria for cognitive authority.

The three empirical studies reviewed in Section 2.4 have a common limitation. In
their studies, the authors have taken a methodological approach in which: (a) the authors
had pre-identified potential factors of quality and authority (in Olaisen’s study only); (b)
they asked users to respond with a numeric scale or to rank the factors. In this approach,
users were perceived as passive recipients of objective information rather than active
information seekers. Since information quality and cognitive authority are related to
judgments that users make, an alternative approach in which users’ interaction with
information is reflected is needed.

Section 2.5 identifies criteria of quality from three different perspectives:
information professionals, users, and systems. There has been a fair amount of literature
on evaluation criteria for Web sites from the perspective of information professionals and
little literature from the perspective of users. It was found that criteria suggested by
information professionals are presented as a guideline or checklist, but none of them was
tested with users. The project conducted by Abels, White, and Hahn (1997, 1998) is a
good example that shows a way of integrating user-based criteria with system
development and implementation. The approach taken from a system perspective views
quality and authority as external, objective, and static. The researchers in this approach,
mostly from computer science, have conducted studies which attempted to measure and
rank the documents according to the attributes of quality and authority.