Examination of work task and criteria choices for the relevance judgment process

Arthur Taylor
Information Systems and Supply Chain Management, Rider University, Lawrenceville, New Jersey, USA

Received 22 December 2011
Revised 31 August 2011
Accepted 2 September 2012

Abstract

Purpose – The purpose of this paper is to better understand the dynamic nature of the relevance judgment process and the influence of work task on that process.

Design/methodology/approach – The empirical study reported here examined the information seeking behavior of a group of undergraduate college students assigned a set of research assignments (work tasks). Subjects recorded their selection of documents used for an assignment and the criteria used to judge those documents relevant. Statistical analysis was used to associate relevance judgments and the criteria used to make those judgments with work tasks.

Findings – Findings indicate a strong statistical association between work task and criteria used to judge relevance. Findings also include identification of specific criteria used to judge relevance and the relative importance of those criteria based on frequency of selection of criteria for a work task.

Research limitations/implications – Findings provide additional insights into the dynamic nature of the relevance judgment process. Relevance judgment influences revealed in these findings in the form of criteria used to make relevance judgments further explicate the relevance judgment process and provide suggestions for the improvement of information retrieval systems and information literacy efforts.

Originality/value – Understanding the relevance judgment process is critical to understanding information behavior in general. Few studies have examined relevance criteria selections as part of the relevance judgment process and fewer still have studied these selections in relation to work tasks. A better understanding of this relationship is an essential part of understanding the dynamic nature of the relevance judgment process and its influences.

Keywords Relevance, Relevance criteria, Work task, Information behaviour, Students, Undergraduates, Information retrieval

Paper type Research paper

1. Introduction

Though relevance has often been treated as dichotomous and static in information retrieval (IR) research, studies have shown that a more complex framework is required to describe the cognitive processes involved in selecting relevant documents as part of an IR search. User relevance judgments are dynamic and multidimensional. As users progress through an information search process, their relevance judgments may change, a document that was once relevant may be considered irrelevant at a later point in time. Their relevance judgments also vary among users. A document which received a positive relevance judgment from one user may produce a negative relevance judgment from another.

Users base relevance judgments on criteria. Some research has attempted to discern the relative importance of these criteria to users but how these criteria may change in

Journal of Documentation
Vol. 69 No. 4, 2013
pp. 523-544

0022-0418
DOI 10.1108/JD-12-2011-0054

importance relative to task, or if they change in importance is undetermined. A deeper understanding of relevance judgments and the dynamic nature of those judgments requires a better understanding of the criteria used to make those judgments in the context of the task which generated the information search.

The longitudinal study reported here examined the selection of criteria used to judge relevance in relation to a work task. Subjects in this study were assigned a set of work tasks as part of a college class. These work tasks were individually graded research assignments with a related topic and included a project abstract, a detailed outline, a rough draft of presentation slides, and a set of presentation slides with a bibliography. Subjects recorded their selection of documents and the criteria used to select those documents. Statistical analysis was used to determine correlations between the subject's relevance criteria choices and work task.

Findings indicate a strong statistical relationship between work task and relevance criteria selections. Findings also include identification of specific criteria used to make relevance judgments while completing tasks and provide additional insights into information seeking behavior in general and the dynamics of the relevance judgment process. These findings have implications in the development and improvement of IR systems, providing guidance on possible criteria and statistical weighting which may prove useful. The findings further provide suggestions for information literacy (IL) efforts by identifying criteria which are important to undergraduate college students in evaluating documents and judging relevance. This information could be used by educators and library staff to identify weaknesses in document evaluation (criteria which were not used) and assisting information seekers in finding documents which are important to a specific work task (assignment).

2. Literature review

Relevance is an important, foundational concept in information science. While early research treated relevance as dichotomous and static, later information science research began to focus on a more cognitive, user-centered view of the relevance judgment process which treats relevance as both dynamic (changing over time) and multidimensional (varying among users) (Saracevic, 1975).

Saracevic (1996, 2007a, b) has provided several examinations of the progress of relevance research in information science and notes that relevance is a key measure for the retrieval of information objects and that users are the ultimate judge of that relevance. Saracevic views relevance in context and identifies a multifaceted system of relevances. He considers manifestations of relevance as attributes or dimensions of relevance which move beyond the commonly identified topical relevance. Saracevic indicates there are clues which identify criteria used to judge relevance and the importance of these criteria changes with task, and progress in task over time. The author emphasizes that searchers use the same criteria, but individually assign different weights to these criteria (Saracevic, 2007b).

The criteria used to judge relevance therefor represents an additional dimension of the relevance judgment process. Schamber *et al.* (1990) propose an examination of these criteria in relation to information behavior and specific evaluation of the consistency of these criteria choices.

Early research on relevance criteria focused on the identification and grouping of criteria into categories. Research by Barry (1994) identified 23 categories of criteria

used by subjects to determine the relevance of documents. The criteria identified were subjective aspects of the document's relevance and contextual factors. Schamber (1991) conducted similar research, and Barry and Schamber (1998) combined data collected from both studies and categorized the criteria into groupings (Table I). Despite the differences in situation and context and slight variations in methodology, a consistent set of criteria was identified in both studies. The relevance criteria identified by Barry and Schamber represent a cross-section of the attributes and manifestations of relevance later identified by Cosijn and Ingwersen (2001). The consistency of these relevance criteria have been strengthened by their identification in other studies (Crystal and Greenberg, 2006; Hirsh, 1999; Maglaughlin and Sonnenwald, 2002; Park, 1993; Tang and Solomon, 2001).

Recent research also recognizes current mobile computing trends and examines relevance in relation to the location of the searcher, examining what is termed geographic relevance from a variety of perspectives (Raper, 2007; Reichenbacher and De Sabbata, 2011; De Sabatta and Reichenbacher, 2012). Other contextual factors include the task that generated the search. An information need may be partially or wholly derived from a work task, a sequence of activities directed at fulfilling the information need (Hansen, 1999). Ingwersen and Jarvelin (2005) suggests there are classifications of information that must be considered and that information seeking behavior may be different for classifications such as the problem at hand, the knowledge domain, and problem solving tasks. Task complexity and other factors may also contribute to these differences (Vakkari, 1999). Some studies have demonstrated a relationship between relevance criteria selections and stages in an information search process but did not examine work task as a variable (Bateman, 1998; Tang and Solomon, 1998; Tang and Solomon, 2001; Taylor *et al.*, 2007, 2009; Vakkari, 1999; Vakkari and Hakala, 2000; Wang and White, 1999; Yuan *et al.*, 2002).

Li (2005) identified a relationship between work task facets and information behavior. The facets identified include time, goal, process, urgency, subjective task complexity, knowledge of task topic, and salience of tasks. These results provide some indication of a relationship between work tasks and information seeking behavior, but Li did not examine the relevance judgment process specifically.

Tombros *et al.* (2005) conducted a study in which subjects were assigned search tasks in a simulated work task environment. The document features identified in the study represent criteria used to judge relevance and the study provides some analysis using descriptive statistics. Findings include identification of a relationship between document feature choices (criteria for relevance judgment) and the search task (work

Grouping	Criteria category
Content of documents	Depth/scope, objective accuracy, tangibility, effectiveness, clarity, recency
User's experience and background	Background/experience, ability to understand, content novelty, source novelty, stimulus document novelty
User's beliefs and preferences	Subjective accuracy/validity, affectiveness
Sources of documents	Source quality, source reputation/visibility
Document as a physical entity	Obtainability/cost

Table I.
Relevance criteria groupings

task) assigned to the subjects. However, the researchers did not perform correlation analysis on their results so the study does not provide strong statistical support for this relationship.

Cosijn (2006) conducted a study in which subjects performing different work tasks answered questionnaires which determined the "reasons for usefulness" for the documents selected (relevant) to their work task. These "reasons for usefulness" can be related to criteria for relevance selection. The study established a statistical correlation between the selection of these criteria and the subject's work task but only examined a small set of criteria, collected data after the completion of the work task and did not examine the relative importance of these criteria for relevance judgments in relation to work task.

While these studies provide some indication of the dynamic nature of relevance in relation to task, additional work which examines criteria selection and a broader selection of tasks would help provide a clearer understanding of this relationship. The purpose of this empirical study is to extend this previous work and add depth by examining more relevance criteria, additional tasks and examining criteria selection in relation to task using statistical analysis.

3. Research goals, definitions and variables

The goal of this research was to examine the relevance decision in more detail, specifically examining the choice of relevance criteria over time in relation to work task as part of the information seeking process. The definitions used and the variables examined in this study are as follows:

- Relevance is defined as the utility of the document to the task for which it is being evaluated.
- The subjects' relevance criteria choice is operationalized as the subjects' choices of criteria which were critical in making their relevance judgment.
- The subject's relevance judgment is operationalized as their judgment (relevant, not relevant, or partially relevant/unsure about relevance) on whether the document will be useful in solving their information problem.
- The work task is operationalized as the specific research assignment due. Assignments were discrete graded assignments: a project abstract, a detailed outline, a rough draft, and a set of presentation slides with a bibliography.

4. Methods

In the study reported here, subjects were assigned a series of research assignments to complete within five weeks as part of a college course (Table II). Subjects were not experts in the subject area of their research topic and had to gather information to successfully complete their assignment. Subjects performed their work in an unmonitored environment (a naturalistic setting) used a web-based search engine accessible from the internet to perform searches for documents relevant to their research. Data was collected anonymously as subjects made their selections using instruments integrated into the web search engine interface. Participants were required to submit three assignments on a weekly basis and the final assignment approximately two weeks after the submission of the third assignment. Each assignment had specific

Assignment	Subject instructions
Project abstract	One or two paragraphs which describe your topic and explain how you will develop the topic in your report
Detailed outline	A detailed outline running two or three pages which provide the various subtopics and issues you will cover in your report
Rough draft	A fairly complete set of PowerPoint slides on your research topic. The slides do not need to be complete and will not be graded down for spelling or grammar. Notes and bibliography are optional
Presentation slides	The complete and final set of PowerPoint slides. The submission should be 15-25 slides with optional slide notes. A bibliography is required

Table II.
Assignments and subject instructions

instructions and grade rubrics and were graded separately, and thus represent discrete work tasks.

Relevance judgments were one of either relevant, not relevant, partially relevant/not sure about relevance (Greisdorf, 2003; Spink *et al.*, 1998). Subjects also indicated the criteria used to make that relevance judgment by selecting from a predetermined list of relevance criteria (Table III).

The web site used for data collection accepted search queries from the subjects and then directed the search at the Yahoo! commercial search engine (Appendix 1). The results returned from the Yahoo! search engine were then reformatted to include the data collection instruments used to collect the data for the study (Appendix 2). This data collection included inputs for search stage in the information search process (data not used in this study), relevance judgments and criteria used to make those relevance

Criteria term	Description
Topic	Document is on my topic
Instructions	Document contains basic advice and instructions (tips)
Understandability	Document is easy to understand; the technical information is easier to read and interpret
History	Document contains a history and/or background of the topic
Guidelines	Document contains basic guidelines and directions
Novelty	The content of the document adds new information to what I already have
Affectiveness	Document is enjoyable
Source	The document is from a source (web site, journal) which has a good reputation in this area
Authority	The author of the document is considered an expert in this field
Recency	Document is up to date and contains current information
Definitions	Document contains basic and/or advanced definitions
Descriptions	Document contains good descriptions and explanations
Breadth	Document covers many topics/subtopics in this subject area
Structure	The structure of the document makes it easier to read and understand
Time	Document is useful because of time constraints
Accuracy	Document seems to have accurate information about my topic
Bias	Document author takes a stand and has a specific opinion (bias); the author is not neutral
Depth	Document contains good depth on the topic

Table III.
Criteria for relevance decision descriptions displayed to subjects

judgments (Appendix 3). If the subjects required additional help on using the web site, web pages could be reviewed to provide that help. Participants were provided instruction on using the research web site and allowed to practice using the site in a tutorial session.

The relevance criteria list displayed to subjects is shown in Table III. This list is based on criteria identified by Barry (1994), Barry and Schamber (1998) and Cool *et al.* (1993). A number of studies have identified these criteria and have provided some confirmation as to their consistency across IR tasks (Park, 1993; Schamber, 1991; Schamber and Bateman, 1996; Xu and Chen, 2006). A subset of these criteria were presented to subjects not as specific criteria, but using the contents of the "description" column in Table III (Appendix 3). Subjects were allowed to choose one or more criteria which they felt contributed to their relevance decision. Not all criteria presented in previously cited studies were used. The reason for the selection of a subset are as follows. Criteria common to both Barry (1994) and Barry and Schamber (1998) were used since these criteria demonstrated consistency. To reduce the potential for confusion and survey exhaustion on the part of the subject, the number of relevance criteria presented to the subject was limited to 15. Excluded were those criteria which were specific only to Schamber's (1991) earlier study and relate to document qualities which were peculiar to her topic (weather reports) and did not apply to the computer-related technical topics used in this study. To address issues of order effects, the order of the list of criteria choices was randomized when displayed to the subjects.

4.1 Data collection process

A convenience sample of subjects was drawn from a sample of junior and senior business students at an American university. Subjects were students in a business class on information technology, and were assigned a research project as part of a class assignment. While the assignment was required, participation in the data collection required for this research was not; all class members chose to participate. Subjects were allowed to choose a research topic from a list of predetermined research topics (Appendix 4). Subjects received a short IL course taught by library staff at the start of the research assignment.

Research topics were all related to information technology (the course was on business management information systems) and were of the same level of difficulty for subjects. The structure and rubric for grading the assignment were the same for all topics. Since there are an abundance of good information technology articles available on the web, subjects were able to use a web browser to conduct their research and, using the programs developed for this research, provide data for this study.

A total of 82 subjects examined and reported on a total of 758 distinct documents found on the web. 64 of the 82 subjects provided background information in the form of a survey. These results indicated that approximately 36 percent of the subjects reported searching for information at least once a day, and approximately 41 percent reported searching for information several times a day. Approximately 64 percent of the subjects were male, and 35 percent were female. Most subjects (77 percent) spoke English as their primary language. Those who were not speaking English as their primary language had an adequate degree of proficiency in the English language to take an English language course at an American university.

Each subject performed searches at their own convenience and at their own pace within the parameters of the due dates for the assignments. Specifically, the process of searching for information and reporting on results of the search as part of this research study was as follows:

- Subjects were asked to sign an informed consent form which explained the purpose of the research and that the information they provided would be treated anonymously.
- Subjects logged into the research web site to conduct their search using a login ID previously assigned, and a personal password they had chosen.
- Subjects entered search query terms as if they were using a commercial search engine such as Yahoo! and executed a search.
- The research web site generated a search results page with a list of results returned by a commercial search engine for the search query the user entered. For each result returned, the search results page included an explanation of the result page/document (as returned by the commercial search engine), links to the results page, and links which the user used to enter relevance information about the page/document (Appendix 2).
- Subjects were asked to enter a relevance judgment for the document, a search stage which identified where the subject was in their search process when they made the judgment (not used in this study), and criteria which were considered by the subject in making that judgment. Relevance assessments were one of either relevant, not relevant, or partially relevant/ not sure (Greisdorf, 2003; Spink *et al.*, 1998). To address issues of order effects, these choices were randomized in the list of relevance criteria choices displayed to the subjects.
- When the subjects finished providing information for the documents reviewed on the results page, they submitted the information they entered to the data collection program on the research web site which stored the results and the date the entries were made for later analysis.
- Subjects repeated the process outlined above as often as they felt necessary and whenever they wished in order to gather the information they needed to complete their assignments.

5. Results and analysis

A total of 82 subjects examined and reported on a total of 758 documents over the five week duration of this study. The criteria used to judge relevance were recorded in a database along with the date and time the judgment was made. These criteria selections were then related to assignments based on the assignment due during the time period when the judgment was made.

Considering that the frequency of selection for relevance criteria for a task is a representation of the interest of subjects in using that criteria, analyzing the frequency of selection of these criteria across tasks provides statistical indication of subject interest in relation to task. Descriptive statistics in the form of tabular counts of criteria selections and percentages within groups of assignments provide some indication of relationships. A more rigorous χ^2 statistical analysis of variance was also used to analyze the frequency of selection for relevance criteria selections and correlate those selections with

the assigned tasks. The variance examined was the frequency of selection for a criteria in moving from one work task (assignment) to another. Statistically significant changes would indicate an association between the criteria and the work task. The commonly used significance standards of 0.1, 0.05, and 0.001 were used.

Subjects were not required to examine a specific number of documents within a given time frame for a specific assignment. As shown in Table IV, subjects reported results for a varying number of documents for each of the assignments. The results in this table indicate that the number of documents evaluated for the detailed outline, rough draft, and final presentation assignments were roughly equal and is therefore a good sample for this study. The selection of criteria for the Project Abstract assignment were not used in this analysis due to the skew which would be produced by the small number of relevance judgments reported during the preparation of that assignment.

Most relevance judgments reported by subjects were for relevant documents, fewer for partially relevant documents, and very few for "not relevant" documents. Since statistical analysis of variance is sensitive to low frequency counts, only partially relevant and relevant document judgments are reported here. The statistical Yates' correction for continuity was applied where needed for low cell (frequency) counts.

5.1 Relevance criteria selections for partially relevant documents

Table V lists the frequency counts for partial relevance selections for the preparation of the "detailed outline" and the "rough draft" assignments (subjects selected a choice of "partially relevant/not sure about relevance"). Subjects selected fewer partial relevant documents for the "presentation slides" assignment. Since this was the last assignment the subjects completed, knowledge gained from research for previous assignments may have had an impact on the subjects' relevance judgments leading to more certainty about judgments of relevance and the selection of more relevant documents (Tang and Solomon, 1998).

Applying a χ^2 test for variance for the data in Table V using a Yates continuity correction for low cell counts, a statistical relationship is found between criteria selection and tasks for the two tasks identified ($\chi^2 = 53.79$, $p < 0.10$). (Analysis of other tasks did not reveal a statistically significant association.) Analyzing the data using a dichotomous χ^2 test to analyze the relationship between each criteria and the two tasks indicates statistically significant relationships for the criteria of "recency" ($\chi^2 = 9.78$, $p < 0.005$) and "understandability" ($\chi^2 = 3.27$, $p < 0.10$). Since these criteria choice counts decrease in value over the two tasks, this association is an indication that these criteria were considered less important for the preparation of the "presentation slides" assignment for judgments of partial relevance than for the "detailed outline" or "rough draft" assignments.

Table IV.
Documents assessed by deliverable due

Deliverable	Count	Percent
Abstract	81	10.69
Detailed outline	225	29.68
Rough draft	187	24.67
Final presentation	265	34.96
Total	758	100.00

Criteria	Detailed outline	Presentation slides
Accuracy	14	9
Affectiveness	1	6
Amount of information	10	7
Authority	1	1
Bias	1	4
Breadth	6	5
Depth	9	4
Novelty	3	0
Recency[b]	19	4
Source quality	2	4
Structure	10	5
Time constraints	0	2
Understandability[b]	24	13
Total	100	64

Notes: [a]Yates' correction for continuity applied for low cell counts; [b]statistically significant correlation with task

Table V.
Comparison of criteria code selections for detailed outline and final presentation – partial relevance[a]

5.2 Relevance criteria selections for relevant documents

Figure 1 lists the frequency counts for the criteria used to judge relevance selections made by subjects for the preparation of assignments (work tasks) for all documents judged to be relevant (Appendix 5). The graph in Figure 1 removes some criteria for clarity. This presentation suggests some variability in the selection of criteria in relation to assignment and provides some hints of a correlation between the relevance criteria choices and work tasks, but additional analysis is required to determine statistical significance.

The results of a χ^2 analysis across the relevance criteria selections for the "detailed outline," "rough draft," and "final presentation" assignments found several associations at various levels of significance as shown in Table VI. These criteria demonstrated a statistically significant correlation to task for this sample. These results will be examined and analyzed with descriptive statistics for each assignment.

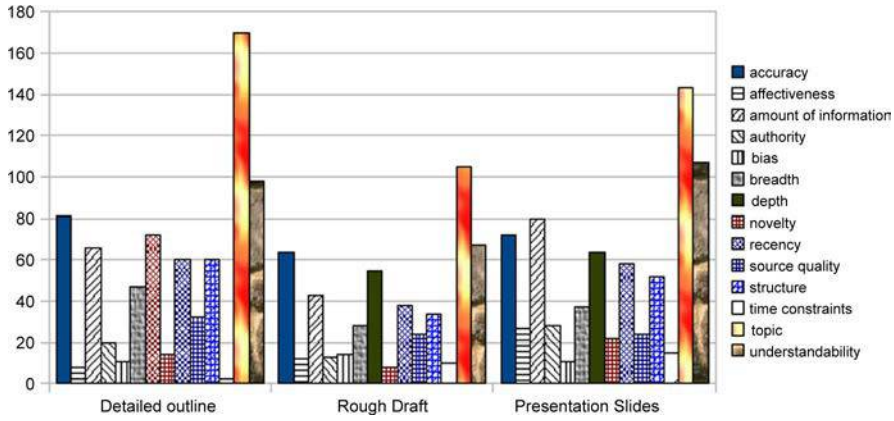


Figure 1.
Frequency counts of select relevance criteria choices for assignments

In Tables VII-IX the frequency counts for relevance criteria count selections is examined as a percentage of all selections for the time period in which subjects were directed to complete the specific assignment. The results in these tables provide some indication of the preference for criteria for a specific assignment. However, these percentages are not used to examine correlation between relevance criteria selection and task; a more rigorous and sensitive analysis of variance test is used to analyze correlation.

The criteria of "topic," "understandability" (ability to understand), "accuracy," "depth" and "amount of information" were consistently selected by subjects with the highest frequency for each assignment. Given that subjects for this study were unfamiliar with the technical topic assigned, these choices are logical. Subjects sought documents with depth and a good quantity of information, which were also accurate. Since the assigned topics were all within the field of computer technology, very specific and often confusing technical language was avoided and subjects instead sought documents which they could understand (had "understandability"). Based on the χ^2 analysis reported in Table VI, there was a statistical correlation between the "understandability," "depth," and the "amount of information" criteria and work task, indicating that for this sample, work task may have influenced the selection of these criteria.

Table VI.
Statistically significant associations across the detailed outline, rough draft, and presentation slides

Criteria	χ^2	p
Affectiveness	12.81	<0.05
Amount of information	11.08	<0.05
Authority	5.54	<0.10
Breadth	4.84	<0.10
Novelty	6.73	<0.05
Recency	5.69	<0.10
Structure	7.29	<0.05
Understandability	9.71	<0.05

Table VII.
Selection of criteria for detailed outline assignment

Criteria	Selected (%)
Topic	22.94
Understandability[a]	13.23
Accuracy	10.93
Depth[a]	9.72
Amount of information[a]	8.91
Structure[a]	8.10
Recency[a]	8.10
Breadth[a]	6.34
Source quality	4.32
Authority[a]	2.70
Novelty[a]	1.89
Bias	1.48
Affectiveness[a]	1.08
Time constraints	0.27

Note: [a]Statistically significant correlation with task

Criteria	Selected (%)
Topic	20.39
Understandability[a]	13.01
Accuracy	12.43
Depth	10.68
Amount of information[a]	8.35
Recency[a]	7.38
Structure	6.60
Breadth[a]	5.44
Source quality	4.66
Bias	2.72
Authority[a]	2.52
Affectiveness[a]	2.33
Time constraints	1.94
Novelty[a]	1.55

Note: [a]Statistically significant correlation with task

Table VIII.
Selection of criteria for rough draft assignment

In Table VII the frequency counts for relevance criteria count selections is examined as a percentage of all selections for the time period in which subjects were directed to complete the "detailed outline" assignment. The criteria of "structure" and "recency" were selected with higher in-group percentages for this task than the other tasks, possibly indicating that subjects were more conscious of document structure in gathering information for an outline assignment than for a more information intensive assignment such as writing presentation slides. The criteria of "recency" may have been more important to subjects for this early assignment (the second in the series) where subjects were gathering information and learning about the topic. Recent information is important in computer technology, a field where changes occur rapidly and information that was current last year may not be significant this year.

In Table VIII the frequency counts for relevance criteria count selections is examined as a percentage of all selections for the time period in which subjects were directed to complete the "rough draft" assignment. The criterion of "depth" was selected with more frequency for this assignment than the other two assignments, and the criterion of "accuracy" was also selected more, though the analysis of variance test did not indicate the criterion "accuracy" demonstrated a statistically significant variance across all three assignments. In contrast to the "detailed outline" assignment, subjects were required to complete presentation slides in this assignment and needed to gather more information than would be required for a detailed outline. The need for additional information may have led to an increased interest in the documents with depth and an increased interest in the accuracy of the information in those documents.

In Table IX the frequency counts for relevance criteria selections is examined as a percentage of all selections for the time period in which subjects were directed to complete the "presentation slides" assignment. For this assignment, the final assignment in the series, subjects sought "understandable" documents with a good "amount of information" and which they perceived as having accurate information about their topic (had "accuracy"). These were criteria which were selected consistently across all assignments but were selected at a slightly higher rate for this assignment. Since this was the final assignment in the series, subjects may have had a stronger interest in these

Criteria	Selected (%)
Topic	19.32
Understandability[a]	14.46
Amount of information[a]	10.81
Accuracy	9.73
Depth[a]	8.65
Recency[a]	7.84
Structure[a]	7.03
Breadth	5.00
Authority[a]	3.78
Affectiveness[a]	3.65
Source quality	3.24
Novelty[a]	2.97
Time constraints	2.03
Bias	1.49

Note: [a]Statistically significant correlation with task

Table IX.
Selection of criteria for presentation slides assignment

core document qualities as they worked on an assignment where the "depth" and "breadth" of the content (the "amount of information") and the "accuracy" of the content would have an impact on their grade. Subjects also sought documents that were new ("novelty"), that provided information that had not been found previously.

6. Discussion

Examining relevance criteria selections for documents judged to be partially relevant in relation to work task, subjects appeared to have an interest in selecting documents based on criteria of accuracy, understandability and "amount of information" for the "detailed outline" assignment, but they show preference for different criteria when selecting partially relevant documents for the "presentation slides" assignment. For the "presentation slides" assignment, "accuracy" and "understandability" were more commonly selected criteria and the criteria of "recency" was selected less than with the "detailed outline" assignment. Since the "presentation slides" assignment was graded more heavily on content and quality, it is logical that the accuracy of sources (and the information in those sources) should have been of greater importance to subjects for this assignment.

Examining the tables of the percentage of criteria selections for the various assignments provides some insights into the selection of criteria for documents judged to be relevant. The understandability (the ability to understand) a document was important for all assignments. This is reasonable given the technical nature of the topic (computer technology) and the subjects' lack of knowledge in this area. The "authority" of the document was selected at a lower rate for the "detailed outline" assignment, but was used at a higher rate for the "rough draft" and "presentation slides" assignments. The "rough draft" and "presentation slides" assignments required a more detailed presentation of facts from the subjects, and were graded more based on content, and would therefore require the subjects to be more concerned with the correctness and authority of a source than they would for the preparation of an outline as was required for the "detailed outline" assignment.

The "amount of information" in a document appeared to be important to subjects throughout all of the assignments, but appeared to be more important for the "presentation slides" assignment. The preparation of slides required more detailed information than the outline (detailed outline) assignment, so it is logical that the subjects were more concerned about the amount of information available for this assignment.

The novelty of the source was not as important for the "detailed outline" assignment as it was for the "presentation slides" assignments. Since novelty infers the information is new, the selectivity of this criteria may be attributed to the previous gathering of information. For this study, the sequence of the assignments (work tasks) is part of the context of the work task and may have been relevant to the selectivity of this criteria. The sequence of the assignments was "detailed outline," "rough draft," and "presentation slides." It is possible that subjects had retrieved documents for the "detailed outline" assignment and were then looking for new (novel) documents for the "rough draft" and "presentation slides" assignments.

The breadth of the information in the document was more important for the detailed outline assignment. The breadth of the document applies to the span of topics/subtopics covered in the document. Since the "detailed outline" assignment was the first in the series of assignments and required subjects to learn about the topic, it is reasonable for the subjects to seek documents which included a broad discussion of the topic. A different task, such the development of an outline or presentation slides, would require subjects to find more detailed, in-depth documents.

The accuracy of the document was more important for the "rough draft" assignment than the "detailed outline" assignment. The task of creating a rough draft requires more accurate information than developing an outline and is reflected in the subject's use of this criteria to judge documents relevant for this assignment.

The structure of the document was important for the "detailed outline" assignment, but less important for the "rough draft" and "presentation slides" assignment. Subjects may have been more concerned with other document criteria for these assignments and less concerned about the structure of the document. Since subjects were using the web to search for documents, the "structure" of documents can vary greatly and could also relate to the navigability of the web site. A more succinct document or a web site with easy navigability may have been more commonly selected for this assignment.

Overall, the subjects were instructed to treat the "project abstract" and the "detailed outline" assignments as a broad information gathering exercise to learn about their topic. The learning exercise was completed during the preparation of the "rough draft" and "presentation slides' assignments. The "rough draft" and "presentation slides" assignments were graded more heavily based on content and accuracy and appeared to prompt the subjects to examine documents more critically and thus use appropriate criteria such as "authority" and "accuracy" for selection.

7. Limitations

A number of contextual factors represent intervening variables in the relevance judgement process and must be acknowledged in assessing these results. The subjects in the study were undergraduate students at an American university who had received limited IL training, so they were slightly more aware of the document qualities that common IL training would recommend be evaluated during the relevance judgement process (identified as "criteria" in this study). Members of the general population are

less conscious of this aspect of the relevance judgement process and might be less inclined to identify these document criteria.

The research experience of an individual also impacts information behavior. The subjects in this study were undergraduate students in a business school with limited experience in conducting research. IL is a well-known issue in undergraduate education (Taylor, 2012; Gross and Latham, 2009) and the purpose of the IL training the subjects took for this study was an effort to address this issue and improve the quality of the data sample analyzed. Longer and more rigorous IL training might lead to more discerning searches and evaluation of specific document characteristics (criteria) by subjects.

Web documents were used for this study. These documents were all related to information technology topics (a requirement of the subject's assignment) and were a combination of computer trade journal articles, a few academic journal articles, general information web sites, and industry "white papers" (documents prepared by technology vendors which are part informational, part advertisement). It is reasonable to expect that the document type has some influence over the relevance decision but the specifics of that relationship were not examined in this study.

Assignments used in this study were all related and were due in time sequence. Such connected and closely timed assignment are common in academic and professional information seeking. Time represents an intervening variable in most information behavior activities and is certainly a factor in this study, though efforts have been made to manage its impact with the methods employed. However, more discrete assignments (work tasks) with longer time gaps in between assignments might provide different results.

8. Conclusion

The findings in this study are as follows:

- There is a strong statistical relationship between the selection of the criteria used to determine relevance for partially relevant documents and work task.
- There is a strong statistical relationship between the selection of the criteria used to determine relevance for relevant documents and work task.
- Subjects show a preference for specific criteria based on work task and other contextual factors as demonstrated using descriptive statistics (Table AI).

The findings of criteria for relevance judgment and work task correlations extend Tombros *et al.* (2005) who saw a relationship between task and criteria selection. The researchers in that study note that subjects in their study identify scope and depth in relation to a "decision task," the task most similar to the work tasks used in this study. The criteria for relevance judgment and work task correlations also extend Cosijn (2006) who noted a relationship between criteria which can be mapped to a subset of the criteria used in this study and work task.

The dynamic nature of relevance includes the use of criteria to make relevance judgments in relation to work task. The work task is part of the broader context of the relevance judgment and as demonstrated in this study with this sample, impacts the relevance judgment process. Based on the results in this study, subjects alter the use of criteria in relation to work task. While other contextual factors are part of this decision process, and the statistical methods used here do not indicate the directionality of

influence between variables, the sum total of these results provide strong support for the case that work task impacts the criteria used to make relevance judgments.

These results inform the study of the relevance judgment process and provide practical suggestions for the development and improvement of IR systems. The finding that novel (new) documents are more important for later work tasks in a sequence of work tasks suggest that filtering previously selected documents would be a useful feature in IR systems. The finding that specific criteria for relevance have added importance for IR users for specific tasks suggests that criteria weighting algorithms could improve selections by applying varying weights for search selection criteria in relation to work tasks.

Understanding the specific criteria used to judge relevance also provides suggestions for IL efforts. Further understanding of how these criteria relate to work tasks such as research assignments would provide further guidance to educators and library personnel. The broad criteria identified in this paper provide some guidance; further research which identifies more detailed criteria could provide additional guidance.

Additional research should continue to clarify this process and examine criteria selection with different subject pools with deeper subject area knowledge. Additional studies should examine the influence of work task in more detail, using a broader set of diverse work tasks. Criteria should also be examined in relation to subject area, with more specific criteria for specific topics or additional criteria categories.

References

Barry, C.L. (1994), "User-defined relevance criteria: an exploratory study", *Journal of the American Society for Information Science*, Vol. 45 No. 3, pp. 149-159.

Barry, C.L. and Schamber, L. (1998), "Users' criteria for relevance evaluation: a cross-situational comparison", *Information Processing & Management*, Vol. 34 Nos 2/3, pp. 219-236.

Bateman, J. (1998), "Changes in relevance criteria: a longitudinal study", *Proceedings of the ASIS Annual Meeting, Pittsburgh, PA, October 25-29*, Vol. 35, pp. 23-32.

Cool, C., Belkin, N.J., Kantor, P. and Frieder, O. (1993), "Characteristics of texts affecting relevance judgments", in Williams, M.E. (Ed.), *Proceedings of the Fourteenth National Online Meeting, Medford, NJ*, pp. 77-83.

Cosijn, E. (2006), "Relevance judgements within the context of work tasks", *Proceedings of the 1st International Conference on Information Interaction in Context, New York, NY*, pp. 20-29.

Cosijn, E. and Ingwersen, P. (2001), "Dimensions of relevance", *Information Processing & Management*, Vol. 36 No. 4, pp. 533-550.

Crystal, A. and Greenberg, J. (2006), "Relevance criteria identified by health information users during web searches", *Journal of the American Society for Information Science and Technology*, Vol. 57 No. 10, pp. 1368-1382.

De Sabatta, S. and Reichenbacher, T. (2012), "Criteria of geographic relevance: an experimental study", *International Journal of Geographical Information Science*, Vol. 26 No. 8, pp. 1495-1520.

Greisdorf, H. (2003), "Relevance thresholds: a multi-stage predictive model of how users evaluate information", *Information Processing & Management*, Vol. 39 No. 3, pp. 403-423.

Gross, M. and Latham, D. (2009), "Undergraduate perceptions of information literacy: defining, attaining and self-assessing skills", *College & Research Libraries*, Vol. 70 No. 4, pp. 336-350.

Hansen, P. (1999), "User interface design for IR interaction: a task-oriented approach", in Aparac, T., Saracevic, T., Ingwersen, P. and Vakkari, P. (Eds), *CoLIS 3*, pp. 191-205.

Hirsh, S.G. (1999), "Children's relevance criteria and information seeking on electronic resources", *Journal of the American Society for Information Science*, Vol. 50 No. 14, pp. 1265-1283.

Ingwersen, P. and Jarvelin, K. (2005), "Information retrieval in context: IRiX", *ACMSIGIR Forum*, Vol. 39, ACM Press, New York, NY, pp. 31-39.

Li, Y. (2005), "Relationships among work tasks, search tasks, and interactive information searching behavior", *Journal of the American Society of Information Science and Technology*, Vol. 60 No. 5, pp. 327-342.

Maglaughlin, K.L. and Sonnenwald, D.H. (2002), "User perspectives on relevance criteria: a comparison among relevant, partially relevant, and not-relevant judgments", *Journal of the American Society of Information Science and Technology*, Vol. 53 No. 5, pp. 327-342.

Park, T.K. (1993), "The nature of relevance in information retrieval: an empirical study", *Library Quarterly*, Vol. 63 No. 3, pp. 318-351.

Raper, J. (2007), "Geographic relevance", *Journal of Documentation*, Vol. 63 No. 6, pp. 836-852.

Reichenbacher, T. and De Sabbata, S. (2011), "Geographic relevance – different notions on geographies and relevancies", *Geographic Information Retrieval*, Vol. 3 No. 2, (SIGSPATIAL Special Issue).

Saracevic, T. (1975), "Relevance: a review of and a framework for the thinking on the notion in information science", *Journal of the American Society for Information Science*, Vol. 26, pp. 321-343.

Saracevic, T. (1996), "Relevance reconsidered", *Proceedings of the Second Conference on Conceptions of Library and Information Science (CoLIS 2)*, pp. 201-218.

Saracevic, T. (2007a), "Relevance: a review of the literature and a framework for thinking on the notion in information science. Part II: nature and manifestations of relevance", *Journal of the American Society for Information Science and Technology*, Vol. 58 No. 13, pp. 1915-1933.

Saracevic, T. (2007b), "Relevance: a review of the literature and a framework for thinking on the notion in information science. Part III: behavior and effects of relevance", *Journal of the American Society for Information Science and Technology*, Vol. 58 No. 13, pp. 2126-2144.

Schamber, L. (1991), "User's criteria for evaluation in a multimedia environment", in Griffiths, J. (Ed.), *ASIS '91: Proceedings of the American Society for Information Science (ASIS) 54th Annual Meeting, Washington, DC, October 30-November 2*, Vol. 28, pp. 126-133.

Schamber, L. and Bateman, J. (1996), "User criteria in relevance evaluation: toward development of a measurement scale", *Proceedings of the ASIS Annual Meeting*, Vol. 33, pp. 218-225.

Schamber, L., Eisenberg, M.B. and Nilan, M.S. (1990), "A re-examination of relevance: toward a dynamic, situational definition", *Information Processing & Management*, Vol. 26 No. 6, pp. 755-776.

Spink, A., Greisdorf, H. and Bateman, J. (1998), "From highly relevant to not relevant: examining different regions of relevance", *Information Processing & Management*, Vol. 34 No. 5, pp. 599-621.

Tang, R. and Solomon, P. (1998), "Toward an understanding of the dynamics of relevance judgment: an analysis of one person's search behavior", *Information Processing & Management*, Vol. 34 Nos 2/3, pp. 237-256.

Tang, R. and Solomon, P. (2001), "Use of relevance criteria across stages of document evaluation: on the complementarity of experimental and naturalistic studies", *Journal of the American Society for Information Science and Technology*, Vol. 52 No. 8, pp. 676-685.

Taylor, A.R. (2012), "A study of the information search behaviour of the millennial generation", *Information Research*, Vol. 17 No. 1.

Taylor, A.R., Xiangmin, Z. and Amadio, W.J. (2009), "Examination of relevance criteria choices and the information", *Journal of Documentation*, Vol. 65 No. 5, pp. 719-744.

Taylor, A.R., Cool, C., Belkin, N.J. and Amadio, W.J. (2007), "Relationships between categories of relevance criteria and stage in task completion", *Information Processing & Management*, Vol. 43 No. 4, pp. 1071-1084.

Tombros, A., Ruthven, I. and Jose, J.M. (2005), "How users assess web pages for information seeking", *Journal of the American Society for Information Science and Technology*, Vol. 56 No. 4, pp. 327-344.

Vakkari, P. (1999), "Task complexity, problem structure and information actions: integrating studies on information seeking and retrieval", *Information Processing & Management*, Vol. 35, pp. 819-837.

Vakkari, P. and Hakala, N. (2000), "Changes in relevance criteria and problem stages in task performance", *Journal of Documentation*, Vol. 56 No. 5, pp. 540-562.

Wang, P. and White, M.D. (1999), "A cognitive model of document use during a research project. Study II: decisions at the reading and writing stages", *Journal of the American Society for Information Science*, Vol. 50 No. 2, pp. 98-114.

Xu, Y. and Chen, Z. (2006), "Relevance judgment: what do information consumers consider beyond topicality?", *Journal of the American Society for Information Science and Technology*, Vol. 57 No. 7, pp. 961-973.

Yuan, X., Belkin, N.J. and Kim, J. (2002), "The relationship between ASK and relevance criteria", *SIGIR '02: Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, New York, NY*, pp. 359-360.

Appendix 1. Search engine interface

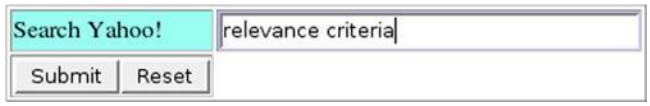
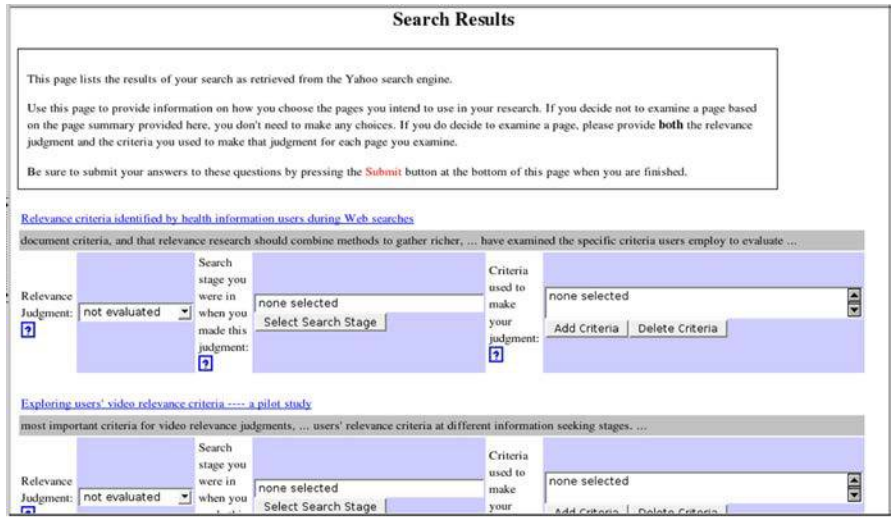


Figure A1.

Appendix 2. Search results page



Note: *Relevance judgment choices are presented in a drop-down list which presents a mutually exclusive choice of *relevant*, *not relevant*, and *partially relevant/unsure about relevance*

Figure A2.

Appendix 3. Criteria for relevance criteria selection

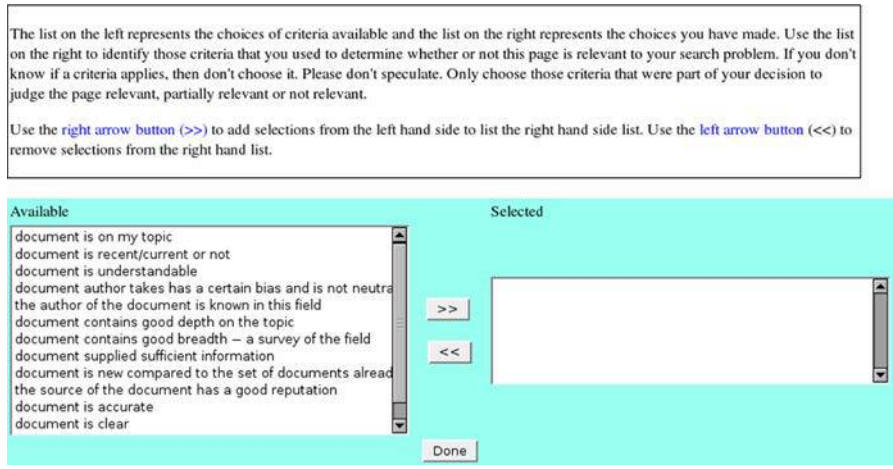


Figure A3.

Appendix 4. Research topics assigned to subjects

- Computer Security: Making Computer Technology Accessible and Secure
- Computer Security: Making Desktop Systems Secure
- Computer Security: Preventing Computer Fraud
- E-Commerce: After the Internet Bubble

- E-Commerce: How to Put Your Company on the Web
- Internet Business Models
- ERP Systems: The Future
- Customer Resource Management (CRM) Systems: Current Status
- Does IT Matter: What Role Will IT Take in the Future?
- New Technologies: Can Linux be Mainstream ?
- New Technologies: the Future of WiFi
- Microsoft: Dealing with the 500 Pound Gorilla
- Ethics and the Information Age: Is It Really Stealing if It's Digital?
- Distributed Computing
- Grid Computing
- Group Collaboration with Computers
- Computer Aided Design (CAD) Systems
- Supply Chain Management with Computers
- Privacy and Computers
- Decision Support Systems
- Implementing Enterprise Resource Planning (ERP) Systems
- Alternatives to ERP Systems
- The Current State of Artificial Intelligence and Expert Systems
- Systems Design and Development
- Enterprise Portals and Application Integration
- Open Source Software on the Desktop: Current Status
- ERP: Implementation Issues

Appendix 5. Frequency counts of relevance criteria choices for assignments – relevant documents

Criteria	Detailed outline	Rough draft	Final presentation
Accuracy	81	64	72
Affectiveness	8	12	27
Amount of information	66	43	80
Authority	20	13	28
Bias	11	14	11
Breadth	47	28	37
Depth	72	55	64
Novelty	14	8	22
Recency	60	38	58
Source quality	32	24	24
Structure	60	34	52
Time constraints	2	10	15
Topic	170	105	143
Understandability	98	67	107
Total	741	515	740

Table AI.

Appendix 6. Criteria terms, descriptions, and sources

Work task	Relevance judgment	Important criteria	Discussion summary
Detailed outline	Partially relevant	Recency, understandability	Current information is important for the assigned task; difficult complex material must be understandable for undergraduate students unfamiliar with the topic
Detailed outline	Relevant	Structure, recency	Document structure (perhaps outline form with short paragraphs) is a consideration in the preparation of an outline; documents should be recent to be relevant for topic (computer systems)
Rough draft	Relevant	Depth, accuracy, authority	Information depth and authority of the information source is important for an assignment graded on content; accuracy is important for assignment grading
Final assignment – presentation slides	Relevant	Understandable, amount of information, accuracy, novelty, authority	Information depth, authority, and accuracy of information source is important for an assignment graded on content; novel (new) sources are needed (portion of grade for work task was based on the quality and quantity of sources used)

Table AII.
Criteria for relevance judgments – summary of findings

Criteria	Type	Source[a]	Used	Description
Depth/scope/specificity	Document	Barry, Cool	Yes	Document contains good depth on the topic
Accuracy/validity	Document	Barry, Cool	Yes	Document appears to be accurate
Currency	Document	Barry, Cool	Yes	Information is current, recent, up-to-date
Tangibility	Document	Barry, Cool	Yes	Information relates to real, tangible issues; not esoteric or theoretical
Quality of sources	Document	Barry, Cool	Yes	Source is reputable, trusted, considered expert
Accessibility	Situation	Barry, Cool	Yes	The effort required to access the information; assumes some cost or effort is involved
Availability of information	Situation	Barry, Cool	Yes	The extent to which the information is available
Verification	Document	Barry, Cool	Yes	The information is consistent with the body of knowledge the field; the information supports the user's point of view
Affectiveness	Document	Barry, Cool	Yes	The user's emotional response to the information; pleasure, enjoyment, entertainment
Amount of information	Document	Cool	Yes	Document provides sufficient information
Depth	Document	Cool	Yes	Document covers the topic in sufficient detail (similar to depth/scope)
Effectiveness of proposed approach	Document	Barry	Yes	How effective is the approach proposed
Consensus within the field	Document	Barry	Yes	How much consensus there is in the field for what is proposed in the document
Time constraints	Situation	Barry	Yes	How much time is allowed for the task to be completed
Background/ experience/ability to understand	Situation	Barry	Yes	Expression of concern over the ability to understand a document (same as "understandability")
Novelty/content novelty/source novelty	Document	Barry	Yes	The source or content of the document is new to the subject
Geographic proximity	Document	Schamber	No	Refers to weather information in a geographic location
Dynamism	Document	Schamber	No	Refers to the ability to dynamic manipulate the information in a document
Presentation quality	Document	Schamber	No	Indication that the source of the information could be manipulated in some way
Structure	Document	Cool	Yes	The structure of the document; how the information is presented/organized

(*continued*)

Table AIII.
Relevance criteria identified in previous studies

Criteria	Type	Source[a]	Used	Description
Timeliness (age of document)	Document	Cool	Yes	Is the time frame of the document appropriate; (current where recent information is required; written in a certain time period for historical significance)
Understandability	Document	Cool	Yes	The document is understandable by the subject (ability to understand)
Guidelines	Document	Cool	Yes	Provides basic direction and structure
Ideas	Document	Cool	Yes	Provides basic ideas and thoughts
Tips	Document	Cool	Yes	Provides basic advice and instructions
Definitions	Document	Cool	Yes	Provides basic and/or advanced definitions
Connections	Document	Cool	Yes	Provides links for related topics and subtopics
Survey	Document	Cool	Yes	Provides a good high level overview
History	Document	Cool	Yes	Provides a good history and background
Level of detail	Document	Cool	No	Provides good depth (similar to scope/depth)
Descriptions	Document	Cool	Yes	Provides explanations and adds clarity
Precision	Document	Cool	No	The document is written with precision (similar to clarity)
Bias	Document	Cool	Yes	The document is written with a particular viewpoint
Specificity (to topic)	Document	Cool	Yes	Specific to the topic (topicality, on topic)
Authority	Document	Cool	Yes	The author or publication has a good reputation in this field

Table AIII.

Source: [a]From Barry (1994, p. 154), Barry and Schamber (1998, p. 226) and Cool *et al.* (1993, p. 3)

About the author
Arthur Taylor is a Professor of Computer Information Systems at Rider University. Prior to joining Rider University, he worked for over 17 years in the field of information systems. His research interests include relevance, the information search process, information literacy, and computer supported education. Arthur Taylor can be contacted at: ataylor@rider.edu

This article has been cited by:

1. Robert Detmering, Anna Marie Johnson, Claudene Sproles, Samantha McClellan, Rosalinda Hernandez Linares. 2014. Library instruction and information literacy 2013. *Reference Services Review* **42**:4, 603-715. [Abstract] [Full Text] [PDF]
2. Ton Mooij. 2014. Exploring a prototype framework of web-based and peer-reviewed "European Educational Research Quality Indicators" (EERQI). *Scientometrics* . [CrossRef]