Academic Information Retrieval Behavior in the Web 2.0 Environment

Authors: Gu Liping

Date: 2016-06-11T00:00:00+00:00

Abstract

Users search for information in nonlinear, non-unitary, and non-mass ways in the Web 2.0 interactive environment. Based on LISAAbstract Users search for information in nonlinear, non-unitary, and non-mass ways in the Web 2.0 interactive environment. Based on LISA literature and C. C. Kuhlthau' s information retrieval behavior model, this paper designed a questionnaire survey on academic information retrieval by graduate students from the Graduate University of Chinese Academy of Sciences in the Web 2.0 environment. This empirical study using the information retrieval behavior model produced recommendations for 16 optimizations and 4 application pathways.

Full Text

Preamble

Total Issue No. 126
November 2008

Academic Information Seeking Behavior in Web 2.0 Environments

Gu Liping[1],[2]
(1. Graduate University of Chinese Academy of Sciences, Beijing 100190;
2. National Science Library, Chinese Academy of Sciences, Beijing 100190)

Abstract In the interactive Web 2.0 environment, users search for information in non-linear, non-uniform, and non-mass ways. Based on a review of LISA literature and C.C. Kuhlthau' s Information Seeking Behavior Model (ISBM), this study designed a questionnaire survey on academic information retrieval among graduate students at the Graduate School of Chinese Academy of Sciences in Web 2.0 environments. This empirical study using ISBM generated 16 optimization recommendations and 4 application approaches.

[**Keywords**] Information retrieval; Information seeking behavior; User behavior; Web 2.0; Social software
[**CLC Number**] G354.2
[**Document Code**] A

1 Introduction

Traditional library science research on information retrieval has been based on the assumption that when users interact with libraries or librarians, a sequential phenomenon occurs: entry → media contact → technology and computers → information → borrowing → returning → isolation → pricing systems, suggestion boxes, library architecture → librarian quality → interior design → catalogs, indexes, classification, and regulatory order [1]. However, given the rapid development of Web 2.0 environments, models established for physical libraries and online information retrieval face new challenges. As network society evolves, new models continue to emerge. Between the old and new, it is necessary to revisit and analyze the applicability of different models to understand current issues in information retrieval, research progress, and unresolved problems, thereby proposing possible solutions.

This paper establishes a research framework based on C.C. Kuhlthau' s Information Seeking Behavior Model, conducts empirical research through questionnaire surveys, and proposes 16 model differences in Web 2.0 environments and 4 ways to optimize the model.

2 Literature Review

2.1 Background Literature: Information Retrieval Behavior in LISA Database

From the LISA database, 1,871 records were downloaded using the keyword "information seeking." A JAVA program converted LISA' s .txt format into a .txt document readable by Derwent Analyst software. After name conversion and data cleaning, statistical results were generated using Excel charts. A total of 1,813 articles from 1970 to March 27, 2007 were analyzed.

Figure: Information Retrieval Research from 1970-2006 (excluding 1971, 1973, 1981)

The analysis revealed: (1) The academic peak began in 1987, climbing steadily from 1997 and reaching its maximum during 2003-2005 (150, 160, and 161 articles respectively), before declining in 2006; (2) Major contributing authors were: Spink, Amanda (33 articles), Savolainen, Reijo (29), Nicolas, David (29), Wilson (29), Ellis, David (17), Kuhlthau, Carol (16), Marchionini, Gary (15), Shenton, Andrew (14), Vakkari, P (14), Ford, N (11), and Large, A (11); (3)

Journal articles dominated (1,801); (4) English was the predominant language (1,661); (5) Top subject terms were Information Work (383), Searching (312), and User Surveys (275).

2.2 Related Literature: Information Seeking Behavior Models

From this background review, four main research directions in information retrieval can be identified:

(1) Actor Theory and Human-Computer Interaction. In 1982, Wilson [2] proposed the Model of Information Behavior, where information retrieval serves as the primary component within a cyclical framework. In 1988, Norman [3] proposed the Theory of Action, assuming human behavior is purposeful and consists of two parts: execution (doing something) and evaluation (comparing what actually happens with what we want to happen).

(2) Behavioral Process and Linear Models. In 1989, Ellis [4] proposed the Behavioral Model of Information Seeking, describing six linear steps: starting, chaining, browsing, differentiating, monitoring, and extracting. In 1991, Kuhlthau [5] proposed the Information Search Process Model, which examines user perceptions beyond linear processes and describes six stages of information searching along with users' behaviors, feelings, and thoughts. In 1995, Marchionini [6] proposed the Process Model of Information Seeking, discussing eight steps in electronic environments and arguing that information retrieval is both systematic and opportunistic.

(3) Dynamic Interaction and Non-linear Models. In 1996, Belkin [7] proposed the Information Seeking Episode Model, viewing user-information interaction as the main process, with other activities (comparing, representing, presenting, visualizing, and navigating) as functional activities supporting effective interaction. In 1997, Saracevic [8] proposed the IR Interaction Model, defining information retrieval as a dialogue between two participants (user and computer) through an interface with primary purposes. In 1998, Sutcliffe and Ennis [9] proposed the Process of Information Searching Activities, summarizing four main activities: problem definition, need expression, problem formation and planning, and result evaluation. In 1999, Choo [10] proposed the Information Seeking Process Model, a triangular cyclic model of information needs, retrieval, and use. In 2000, Sutcliffe et al. [11] proposed Model Mismatch Analysis, a five-stage retrieval process based on Human-Computer Interaction theory and task-oriented approaches. In 2002, Kim [12] proposed the Digital Library Information Seeking Process, designed for digital library environments to measure usability.

(4) Information Seeking Behavior in Network Environments. In 2003, Loeber and Cristea [13] proposed the WWW Information Seeking Process Model, integrating information retrieval with website design (information intermediation), arguing that user motivation, ability, and opportunity dominate information seeking, with no clear beginning but termination points

online. In 2004, Foster [14] proposed the Nonlinear Model of Information Seeking, developing three core processes (internal, external, cognitive) and three interaction contexts (opening, orientation, consolidation) as dynamic combinations. In the same year, Pharo [15] proposed the Search Situation and Transition Model, analyzing influences among work tasks, researchers, social/organizational environments, search tasks, and search processes. In 2006, Savolainen [16] proposed the Schematic Model of Information Seeking Process, arguing that: (1) time is a basic variable in the context of information seeking; (2) time is a necessary condition for obtaining information; (3) time is an indicator of the information search process.

2.3 Most Relevant Literature: C.C. Kuhlthau' s Model

Among the key authors identified in Section 2.1, Kuhlthau stands out with over 30 years of research experience, contributing to all four development directions in Section 2.2. Her model, developed over six years, has been consistently cited for the past 25 years.

(1) Kuhlthau' s Model. Beginning with her doctoral dissertation, Kuhlthau developed the ISP (Information Search Process) model through continuous work grounded in human cognition, psychology, education, and library science. Drawing from Dewey' s philosophy on "how we think" to design "information retrieval steps," Kuhlthau discussed Dewey' s book *How We Think* and its description of reflective thinking. According to Kuhlthau, Dewey described the dynamic rules of individual action in the process of using information for learning, from which she conceptualized six stages of user information searching [17]. She subsequently proposed [18,19] the famous six stages of information seeking: Task Initiation, Topic Selection, Prefocus Exploration, Focus Formulation, Information Gathering, and Search Closure. Later, she added two concepts to her research [20]: uncertainty and intervention.

(2) Research on Kuhlthau' s Model. Due to its clear concepts and operational nature, many information seeking theories and empirical studies have built upon Kuhlthau' s model. For example: Burdick [21] used Kuhlthau' s model to find that women may be clearer and more focused during the focus formulation stage than men, while men are more ambiguous during mid-search stages, though gender differences are not significant at search completion. Doris [22] found in his doctoral dissertation that PhD students experience the most anxiety during stages one and three of Kuhlthau' s model. Josette [23] directly applied Kuhlthau' s model to design in-depth interviews studying nurses' online searching behavior for medical education materials, finding equivalent steps to Kuhlthau' s model but not all nurses experienced feelings, thoughts, and actions matching the model. Esmaeel [24] used Kuhlthau' s model to design questionnaire sections in his doctoral dissertation, finding 41% of student respondents considered themselves in stage five, while 72% reported experiencing more than one stage simultaneously (non-linear information search process). Kyunghye Kim [25] detailed information seeking models including Kuhlthau' s in his doc-

toral dissertation, designing a Digital Library Information Seeking Process (DL ISP) Model to guide usability interviews. Additionally, Susan [26], Naseer [27], and Abdulmohsin [28] cited Kuhlthau' s model in their respective studies on information needs and geoscience education, higher-order thinking in secondary school internet searching, and university student library use.

---

3 Survey Design

Due to Kuhlthau's model's strong extensibility, this study designed an academic information seeking behavior model based on it, decomposing the process into six steps: initiating search, judging information, exploring, using, obtaining, and learning information acquisition.

(1) Research Method. The survey method was employed with simple random sampling, using SPSS for data analysis.

(2) Research Participants. The study targeted doctoral and master' s students at the Chinese Academy of Sciences in Beijing' s Haidian District, all with at least four years of computer experience since undergraduate studies.

(3) Questionnaire Design. Open-ended interviews were first conducted with Chinese Academy of Sciences graduate students in libraries and graduate schools to design and evaluate questionnaire items. After literature review, a draft questionnaire was designed and pilot-tested with 30 participants, revised based on expert feedback, and finalized.

(4) Data Collection and Entry. From January 14 to 27, 2007 (14 days), questionnaires were delivered and collected door-to-door in dormitories every evening from 9-11 PM using a systematic sampling approach (3 out of 7 rooms). A total of 450 questionnaires were distributed, 380 were returned, and 358 were valid. After data entry, questionnaires with prime numbers were selected for verification.

(5) Data Verification and Analysis. Simple descriptive statistics were conducted to obtain preliminary exploratory intelligence for future research.

---

4 Survey Results

4.1 Basic Demographic Distribution of Respondents

**** Age, Internet Experience, and Degree Level of Respondents

Age distribution: 20-25 years (64.5%), 26-30 years (31%), 30-35 years (3.9%), over 35 years (0.6%). Internet experience: 3-5 years (35.5%), 6-9 years (57.2%), 10+ years (3.6%). Current degree: Master' s (55.6%), Doctoral (44.4%).

The age distribution shows normal distribution characteristics, with 95.5% of respondents aged 20-30. Internet experience of 3-9 years accounts for 92.7% of the sample, matching age characteristics. The doctoral-to-master' s ratio of 5.5:4.5 aligns with sample age characteristics and education levels. This indicates a focused and concentrated target group with suitable analytical conditions for this study' s purposes.

4.2 Basic Internet Usage Survey

**** Internet Usage Frequency and Costs

Usage frequency: Daily (91.1%), 2-3 days (6.7%), 4-7 days (2%), two weeks (0.3%), monthly or less (0%). Cost coverage: Fully public-funded (23.5%), fully self-funded (10.9%), both (65.6%).

Among respondents, 91.1% access the internet daily, demonstrating the necessity of studying information behavior. For internet costs, fully self-funded accounts for only slightly over 10%. Since Chinese Academy of Sciences institutes and the National Science Library provide network and computer equipment, and subsidize dormitory internet access for first-year students, external factors like equipment or economics do not significantly influence internet access.

4.3 Academic Information Behavior on the Internet

**** Online Academic Behavior Patterns

Starting search: Electronic journals (35.5%), search engines (33%), professional institution websites (14.5%), Web 2.0 libraries (13.5%), Web 2.0 bookstores (3.6%). Judging information: Personal knowledge (28.2%), cross-referencing (27.3%), never doubting (4%), asking experts (21.5%), asking librarians (0.6%). Facing search difficulties: Change search terms (70.4%), change retrieval system (18.7%), ask others (7%), original method (1.1%), give up (1.1%). Facing usage difficulties: Change method/tool (57.3%), ask friends (19.3%), terminate (2%), ask librarians (8.7%), other (3.6%). Facing access difficulties: Use abstracts (40.8%), give up (29.9%), other (24.9%), skip full text (4.6%). Learning information acquisition: Self-experimentation (63.1%), told by others (33%), other (including not learning) (2.6%), training courses (1.4%).

Based on survey results, students' online academic information behavior primarily involves using electronic journals and search engines to find materials. After locating materials, they judge online information reliability based on personal knowledge or cross-referencing. If desired academic information cannot be found, they change search terms or retrieval systems for another attempt. When facing information without full text, half choose alternative methods or tools. Even without full text, the largest proportion chooses to utilize abstracts. Over half learn information acquisition skills independently. These phenomena show that students' academic information behavior involves facing a multi-source information environment, using familiar methods to search, access, use, and learn,

while simultaneously adapting to the environment.

5 Discussion

Within this student population aged 20-30, with 3-9 years of internet experience and daily internet access, information needs clearly prioritize electronic journals supplemented by other resources. Information searching primarily begins with electronic journals and search engine websites. If satisfied, they continue using these methods; if not, they first attempt changing vocabulary or tools within the same environment before finally moving to other Web 2.0 spaces. Libraries play the role of information resource providers rather than first-access objects, while also serving as assistants helping users find, access, use, and learn information.

First, in Web 2.0 environments, researchers cannot ignore the impact of virtual communities on information retrieval. This creates significant differences and research challenges compared to past studies conducted in libraries or information consultation agencies regarding user group definitions, user identities, and information retrieval purposes.

Second, in Web 2.0 environments, influenced by other Web 2.0 information, information retrieval models may resemble Wilson' s cyclical model, but could also be random or jump-based models.

Third, in Web 2.0 environments, Norman' s 1988 description of two activities in Human-Computer Interaction (HCI)—execution and evaluation—still applies. However, since Web 2.0 provides non-single information sources that users can continuously select, evaluation and execution become alternately interactive.

Fourth, in Web 2.0 environments, information retrieval may consist of numerous short processes, with different linear processes running parallel across various information sources and task lists.

Fifth, in Web 2.0 environments, users' behaviors, feelings, and thoughts remain important, but rather than being influenced by retrieval results, users' behaviors, feelings, and thoughts dominate the information retrieval process.

Sixth, in Web 2.0 environments, the information retrieval process may not involve a sophisticated systematic formula but rather be assembled from extremely simplified steps through repeated retrieval trials.

Seventh, in Web 2.0 environments, beyond considering users' psychological and social factors, we must consider core causes of information composition: individual learning and working states, culture and society, information formats, technological changes, organizational culture and atmosphere, as well as other factors such as language ability, professional knowledge, critical thinking, associative ability, ambition, and creativity.

Eighth, in Web 2.0 environments, it is not merely one-on-one dialogue between

user and computer, but also dialogue and knowledge exchange between users, completed through computers, implementing a non-formalized co-creation.

Ninth, in Web 2.0 environments, throughout the process of problem identification and resolution, each stage requires information retrieval to obtain relevant evidence and latest information. Therefore, information retrieval is not a part but embedded throughout the entire process.

Tenth, in Web 2.0 environments, information needs, retrieval, and use not only influence each other, but also influence each other between users. Individual needs may expand to group needs, or group information retrieval may simultaneously influence individual users' information needs and use, such as increased desire to understand and full utilization of massive information.

Eleventh, in Web 2.0 environments, task-oriented information activities and purposeful information seeking behavior remain absolutely dominant, but we must also consider non-task activities and purposeless information seeking in Web 2.0 environments, and the process from purposeless to purposeful.

Twelfth, in Web 2.0 environments, users likely do not enter a single digital library but obtain digital library resources within Web 2.0 work environments.

Thirteenth, in Web 2.0 environments, surfing, browsing, searching, and bookmarking are all forms of information retrieval, transforming information retrieval from one-way closed to two-way open.

Fourteenth, in Web 2.0 environments, non-linear, dynamic models (that can start and end at any step) become an unavoidable research focus.

Fifteenth, in Web 2.0 environments, with expanded online scope and degree, besides user psychology, social and organizational factors become increasingly important influences between work tasks and retrieval activities.

Sixteenth, in Web 2.0 environments, time measurement in seconds becomes the most important metric, and any information retrieval model must consider time variables.

Researching information retrieval in Web 2.0 environments requires considering many more factors than before, with "environment" becoming the primary consideration. Web 2.0' s impact does not mean abandoning original information retrieval models but reminds researchers to optimize models based on existing foundations.

In summary, the greatest difference between Web 2.0 information retrieval models and other information retrieval models lies in diversified information sources, user community formation, changed retrieval patterns, and information classification shifting from institutional access to user cognition. Optimizing information seeking behavior models for Web 2.0 environments is shown in ****:

**** Application Approaches for Information Retrieval Models

Research Object	Single User
Information Sources	Databases; Search engines, instant messaging, Wikipedia, blogs, aggregation, social bookmarking, P2P transmission, Web 2.0 navigation, professional website portals
Information Classification	News, knowledge, data presentation, knowing, understanding, thinking
Retrieval Patterns	Linear, block (non-linear), spider web, fishbone, branch, multi-line intersection

6 Conclusion

Information retrieval has always been a crucial research direction in library and information science. It represents the culmination of information systems, information services, intelligence work, and user behavior, while relatively providing them with partial guiding principles. This was true in the bibliographic index era, the card catalog era, the retrieval system era, the online catalog era, the network era, and remains so in new-generation network technology service environments such as e-science or Web 2.0.

The Information Seeking Behavior Model (ISBM) initially faced single-machine operation behavior in physical libraries, explainable by a "start-to-end" linear model. However, since the 1990s, researchers realized that pure linear models could not fully explain behavioral phenomena. On one hand, developments in psychology, sociology, organization, and education drove ISBM to consider more internal psychological and external social aspects of actors. On the other hand, usage contexts became increasingly complex, and the arrival of internet services after online catalogs changed environmental factors in behavioral processes.

Thus, non-linear models and linear models under limited conditions became two mainstream explanations for information retrieval behavior before 2003, with the former mostly being exploratory research and the latter mostly continuation research. The recent wave of change comes from the maturation of weblog research methods, more rigorous validation techniques in social science research methods, and environmental changes in Web 2.0 concepts and related network service applications. These compel researchers to conduct a third wave of exploratory research on behavioral phenomena that differ greatly from the past.

First, information retrieval models applied in different practical environments

(hospitals, industrial factories, commercial institutions, schools, research centers) have generated many revised models. Similar parts of these models are summarized into one or two classic models, which are then applied to different practical environments. Over time, this has accumulated considerable achievements, representing its unique evolutionary generation and being one reason for the proliferation of information retrieval models.

Second, information retrieval models are both the summary and guiding principle of information systems, information services, intelligence work, and user behavior. Therefore, technological progress, demand changes, function adjustments, and social changes in the same environment also cause model revisions, another reason for its enduring research vitality.

Third, due to practice in different environments and situations, information retrieval models continuously absorb new disciplinary theories to improve themselves and adapt to objects, resulting in different models under different research perspectives.

Fourth, due to temporal shifts, environmental changes, domain integration, and empirical induction, continuous model derivation requires multidisciplinary perspectives and comprehensive research methods to conduct such model research.

Finally, different practical experiences caused by different environments, times, and perspectives have produced numerous different information retrieval models. Therefore, there is no permanent model, only problem-solving models. The currently urgent and unresolved problem is that Web 2.0 information seeking behavior has not been fully described by any model, warranting researchers' collective efforts.

Combining these accumulated linear, non-linear, and multi-task models, dynamic characteristics of time factors must now be added. Even in Web 2.0 environments, interpersonal interaction rather than simple human-computer interaction must be considered, indicating the magnitude and difficulty of the undertaking. This paper proposes optimization and application of "academic information seeking behavior in Web 2.0 environments" from C.C. Kuhlthau' s model as one reference approach for improving ISBM.

References

1. Brigitte Kuehne. *Informationssuche Oder Informationskompetenz?* www.ib.hu-berlin.de/~libreas/libreas_neu/ausgabe3/pdf/003kue.pdf (2005)
2. Wilson, T. D. Revisiting User Studies and Information Needs. *Journal of Documentation*, 2006(6): 680
3. Norman, D. A. *User Centered System Design: New Perspectives on Human-Computer Interaction.* Hillsdale, New York, Lawrence Erlbaum

Associates. 1986: 31

4. Ellis, D. A Behavioral Approach to Information Retrieval System Design. *The Journal of Documentation*, 1989(3):
5. Kuhlthau, C. *Seeking Meaning: A Process Approach to Library and Information Services*. Ablex Publishing Company. Norwood: NJ. 1993: 18
6. Gary Marchionini. *Information Seeking in Electronic Environments*. Cambridge University Press, 1995: 27
7. Lin, Shin-jeng; Belkin, Nicholas J. Modeling Multiple Information Seeking Episodes. *Proceedings of the ASIS Annual Meeting*, 37, 2000: 133
8. Saracevic, T. Interactive Models in Information Retrieval (IR): Progress, Problems, Proposal. *Proceedings of the ASIS*, 1996: 33
9. A. G. Sutcliffe, M. Ennis, S. J. Watkinson. Empirical Studies of End User Information Searching. *Journal of Information Science*, 1999(3): 1
10. Choo, C. W., Detlor, B., & Turnbull, D. Information Seeking on the Web: An Integrated Model of Browsing and Searching. http://www.firstmonday.dk/issues/issue5_2/choo/ (2001)
11. Sutcliffe, A. G. and Ryan, M. Model Mismatch Analysis: Towards a Deeper Explanation of Users' Usability Problems. *Behav. Inf. Tech.*, 2000(1): 43
12. Kyunghye Kim. A Model of Digital Library Information Seeking Process (DL ISP Model) as a Frame for Classifying Usability Problems. (Doctoral Dissertation, New Brunswick, 2002) UMI number 3055068.
13. Susanne Loeber & Alexandra Cristea. A WWW Information Seeking Process Model. *Educational Technology & Society*, 2003(3): 43
14. Allen Foster. A Nonlinear Model of Information Seeking Behavior. *Journal of the American Society for Information Science and Technology*, 2004(3): 228
15. Pharo, N. A New Model of Information Behaviour Based on the Search Situation Transition Schema. *Information Research*, 2004(1): 203
16. Reijo Savolainen. Time as a Context of Information Seeking. *Library & Information Science Research*, 28, 2006: 110
17. Kuhlthau, C. *Seeking Meaning: A Process Approach to Library and Information Services* Second ed., Westport, CT, 2003.
18. Kuhlthau, C., Turock, B., George, M., Belvin, R. Validating a Model of the Search Process: A Comparison of Academic, Public and School Library Users. *Library and Information Science Research*, 1990(1): 5

19. Kuhlthau, C. Inside the Search Process: Information Seeking from the User' s Perspective. *Journal of the American Society for Information Science*, 1991(5): 361

20. Kuhlthau, C. Accommodating the User' s Information Search Process: Challenges for Information Retrieval System Designers. *Bulletin of the American Society for Information Science*, 2000(13):

21. Burdick, T. Gender in the Information Search Process: An Exploratory Study of Student Experience. (Doctoral Dissertation). The State Florida University, 1995.

22. Doris Judy van Kampen. Library Anxiety: The Information Search Process and Doctoral Use of the Library. (Doctoral dissertation). The University of Central Florida, 2003.

23. Josette F. Jones. Searching for Patient Educational Material Using Electronic Information Resources: An Exploration of Nurses' Search Behavior. (Doctoral Dissertation). University of Wisconsin-Madison, 2002.

24. Esmaeel Shamo. University Students and the Internet: Information Seeking Study. (Doctoral Dissertation). University of North Texas, 2001.

25. Kyunghye Kim. A Model of Digital Library Information Seeking Process (DL ISP Model) as a Frame for Classifying Usability Problems. (Doctoral Dissertation). New Brunswick, 2002.

26. Susan Ward Aber. Information Needs and Behaviors of Geoscience Educators: A Grounded Theory Study. (Doctoral Dissertation). Emporia State University, 2005.

27. Naseer Mohammed Aomarl. Implementation of Higher Order Thinking in Internet Searching in Secondary School Students. (Doctoral Dissertation). Fordham University, New York. 2001.

28. Abdulmohsin Hasan Al-Harbi. Internet Use by Graduate Students in the Communication Department of Florida State University and Its Impact on the Use of FSU Academic Libraries. (Doctoral Dissertation). Florida State University, 2002.

(Received: 2008)

Note: Figure translations are in progress. See original paper for figures.