Delivered at the 1st International Scientific Conference eRA: The Information Technology to Science, Economy, Society & Education (Cultural Center of Tripoli, Greece, 16-17 September 2006)

Information Behaviors of Academic Researchers in the Internet Era: An Interdisciplinary & Cross-cultural Study

Peiling Wang

School of Information Sciences, The University of Tennessee Knoxville, USA
Tel: (865) 974 3700; Fax: (865) 974-4657, E-mail:peilingw@utk.edu

This paper reports on part of a study of academic researchers' use of Internet information and communication technologies (IICTs) to support information-seeking activities. The goal of this research is to gain insight into disciplinary and cultural differences of information seeking in the Internet Era. The project is ongoing to include more participants from different cultures.

1. Introduction

This paper reports the preliminary results of an ongoing research project on the use of Internet information and communication technologies (IICTs) by academics in research. The purpose of the project is two-fold: (1) to identify interdisciplinary differences/gaps and factors; (2) to identify cross-cultural differences/gaps and factors. The Internet has transformed the world and affected every domain of human life by providing an information-rich environment. Meanwhile it also creates a "digital divide[1]" in that adoption of IICTs is affected by various factors, such as usefulness of the resources and usability of the system. Although researchers and scholars have adopted the Internet for information and communication as early as its inception, their uses of the Internet for research information needs are wide ranging from little to heavy. It is crucial for Information Science to understand the needs and factors underlying the uses of IICTs in order to design effective Internet-enabled information resources. This study will focus on research uses of IICTs and the factors affecting use and nonuse. The following research questions will be investigated:

1. Which Internet information and communication technologies (IICTs) are used in research? How important is each of the IICTs in supporting research?
2. How do researchers use the IICTs to facilitate specific information-seeking activities (such as monitoring a field's development, searching for information, accessing information objects, etc.)?
3. To which extent are information needs satisfied by digital resources?
4. What are the interdisciplinary differences in the use of IICTs for research?
5. What are the cross-cultural differences in the use of IICTs for research?

[1] Digital divide is defined as a "gap between those who can effectively use new information and communication tools, such as the Internet, and those who cannot." (Gunkel, 2003)

2. Literature

This literature reviews recent studies of research use of IICTs to support information seeking (IS) activities identified by Ellis (1989, 1993) and extended by Meho & Tibbo (2003). Ellis (1989) identified and categorized six IS activities associated to research projects based on interviews with social sciences researchers in academic environment: (1) *starting*, (2) *chaining*, (3) *browsing*, (4) *differentiating*, (5) *monitoring*, and (6) *extracting*. *Starting* refers to seeking information on a new topic and gathering initial relevant information. *Chaining* refers to following references in a work to its cited works (backward) and finding new citations to this work (forward). *Browsing* refers looking casually (semi-directed) for information in an area of interest. *Differentiating* refers to discriminating between information sources using specific criteria. *Monitoring* refers to keeping abreast of developments in areas of research interests. *Extracting* refers to working systematically through sources to identify relevant material of interest. These IS activities are interactive and iterative. Researchers use and prefer different information resources and tools to support different IS activities.

Ellis's model received wide attention because of its implications for designing information products and services. Over the last two decades, the Internet has changed our information world. Researchers adopted new Internet technologies for information and communication needs. The model has been revisited and adapted in several current studies.

Meho and Tibbo (2003) interviewed social science researchers working on similar topics and extended Ellis's behavior model with four additional IS activities: (1) *accessing*, (2) *verifying*, (3) *networking*, and (4) *information managing*. *Accessing* refers to obtaining the materials or the identified information objects; *verifying* refers to checking the accuracy of the found information; *networking* refers to communicating and maintaining a close relationship with people and organizations; *information managing* refers to filing, archiving, and organizing the information objects they use in research. They also proposed a four-stage IS model based on project lifecycle: (1) *searching*, (2) *accessing*, (3) *processing*, and (4) *ending*. The *searching* stage can be defined as a period of identifying and gathering relevant materials. The *accessing* stage can be defined as obtaining needed materials or gaining access to information sources; thus the bridge between the searching and the processing. The *processing* stage is where researchers analyze and synthesize the obtained information and write the final product. The *ending* stage marks the end of a research project cycle.

It is well-documented that information seeking is situational and context-based. There are disciplinary differences in information and communication behaviors. Since Ellis's first study to observe IS behaviors, the information environment has changed significantly, mostly due to the inception of the Internet. How do IICTs affect IS behaviors in different disciplines? Within each discipline, what are the individual differences? What might be the factors for the differences?

In her thesis research based on Ellis's behavior approach, Ge (2005) conducted interviews with 30 academic researchers in the social sciences and humanities on their information-seeking activities and their use of Internet resources for relevant information. Her findings corroborate those of Meho and Tibbo (2003).

3. Conceptual Framework

General IS Behavior

- ***Monitoring*:** Keeping abreast of developments in areas of research interest
- ***Browsing*:** Looking casually for information in research areas
- ***Managing*:** Storing and organizing information

Task-based IS Behavior

- ***Starting*:** Gathering initial relevant information
- ***Searching*:** Searching through specific sources (focused)
- ***Accessing*:** Obtaining identified information objects
- ***Chaining*:** Following references (backward) or citations (forward)
- ***Ending*:** Stopping information gathering or writing

IICTs Type

- Web
- Email
- Online library catalog
- Database
- Digital library (Library portal)
- E-journal
- FTP (file transfer protocol)
- Listserv
- Blog
- BBS (bulletin board systems; or discussion/message board)
- Newsgroup
- Wiki
- Instant Messaging (MSN/Skype)

Despite the differences in research areas or methods, academic researchers follow similar research cycles, which affect their information seeking. Information seeking is time-consuming and tedious. The various IICTs can effect information seeking and save research time. In this study, we adopt a framework, which incorporates modification to the ten IS activities emerged in the studies by Ellis (1989) and Meho & Tibbo (2003). The framework includes eight IS activities classified into General IS Behavior and Task-based IS Behavior (left column above). The general IS activities satisfy long-term goals while task-based IS activities meet the current needs for information. The four activities, *differentiating*, *extracting*, *verifying*, and *networking*, are not treated individually. The first three activities are grouped together under a new label *searching* (corresponding to Meho & Tibbo's first stage), although *verifying* is closely related to the use of retrieved relevant information. *Networking* is an important communication behavior. Its complexity deserves an individual study. Thus, this study will not ask focused questions about it, but expect it to emerge naturally in the interviews. The *ending* is adopted to enclose the research lifecycle. It looks into how

researchers wrap up a project (whether they repeat any IS activities) or disseminate their research products using IICTs in addition to publishing.

The Internet is the most influential information technology innovation since computers. New IICTs are being rapidly developed and adopted. The framework (above) lists thirteen IICTs types ordered by the number of adopters in this study. The last two types have emerged from the interviews. But they are not yet widely used for research, although they may grow in importance as information channels in the future.

4. Research Design

This project takes a qualitative approach to gain understanding of researchers' perspectives and insights. In-depth face-to-face interviews with the participants allow interactions between the interviewer and the participant. Data collection started in summer 2005 in the US and China. The project is ongoing to including more participants from more disciplines and more countries.

4.1 Participants

Researchers in higher education from Computer Science, Engineering, Information Science, Journalism, and the humanities are the primary target populations for this project. The chosen disciplines represent the spectrum from computer-oriented to people-oriented in terms of research focus and methods. Productive and active researchers are identified from the WebPages of the departments or schools; they are contacted via email or phone. At the time of this writing (May 2006), we have interviewed 65 researchers in the United States (55 participants) and China (10 participants). The distribution of the academic status is as follows: 22 professors, 15 associate professors, 13 assistant professors, and 15 doctoral students.

4.2 Interview Guide

A four-part interview schedule was developed as a data collection instrument to ensure consistency among interviewers and to take notes during the interview. The questions are arranged to make the interview flow easily but not necessarily in the order of the research questions or the conceptual framework. The first part asks the interviewee to select from a list of 11 types of IICTs the ones which they use for research. The list includes the first 11 IICTs in Section 3 presented in alphabetical order. The interviewee is asked to add any types not on the list. The second part asks questions about the selected IICTs regarding the length and frequency of usage. The third part asks the interviewee to sort the IICTs in the order from the most important to the least important. The interviewee is also asked for comments on the nonuse IICTs types. The forth part asks how the selected IICTs types support his/her IS activities based on the framework in Section 3. The guide includes in each part an open question for additional comments or information. Alternative questions are used when needed. For example, the alternative to "*How long have you used* ..." is "*What year did you start to use* ...?" The former was preferred by some participants; the latter others. Ordering selected IICTs by importance is explained when needed: "*One way is*

to ask 'Can I do my research without it?" Or, "Will my research be more difficult to carry out without it?"

4.3 Procedure

The interviews were conducted mostly in the participant's office and occasionally, preferred by the participant, in the researcher's office. The session began with a brief description of the project and procedure, which was followed with an explanation of voluntary participation and measures for data anonymity and confidentiality. Upon agreeing to be interviewed, the participant signed a consent form. The interview was audio recorded with only one exception when the participant did not want to be recorded (the interviewer took notes). A typical interview took approximately half hour.

5. Results and Discussion

The recorded interviews are transcribed in text. Quantitative data are coded and analyzed in SPSS; qualitative data are analyzed using QSR N6.

Due to the space limitation, this paper reports only the preliminary findings on the use of IICTs, the perception of their importance, the specific IS behaviors they support, and the extent of using digital or print information resources. The mean number of IICTs used by all participants (N=65) is 7 (SD=2). Eight IICTs are used by more than half of participants: (1) the *Web* by all (n=65), (2) *email* by 92% (n=60), (3) *online library catalog* by 91% (n=59), (4) *database* by 83% (n=54), (5) *E-journal* by 80% (n=52), (6) *digital library* by 78% (n=51), (7) *listserv* by 55% (n=36), and (8) *FTP* by 54% (n=35). The remaining 5 types are used by fewer researchers (<=17%): *Blog*, BBS (and various Internet forums), *newsgroups*, *wiki*, *instant messaging (MSN* and *Skype*). The last two types of IICTs are added to the list by a few participants.

The perceived importance of the IICTs is determined according to how the participants ordered them. Each participant places the selected IICTs in the order of importance. The order is mapped into numerical scores; i.e., the most important IICT is assigned 1. If several IICTs are placed in a tied rank, the scores are normalized. For example, when the *Web* and *email* share the same rank as the most important IICTs, each will get a normalized score of 1.5 (half below one and half above 2). The eight widely used IICTs are ranked according to mean importance scores: (1) *Web* (Mean=2.84; SD=1.71), (2) *email* (Mean=2.94; SD=1.84), (3) *database* (Mean=3.34; SD=2.17), (4) *e-journal* (Mean=3.90; SD=1.60), (5) *online library catalog* (Mean=3.83; SD=1.98), (6) *digital library* (Mean=4.12; SD=1.66), (7) *ftp* (Mean=5.69; SD=2.53), and (8) *listerserv* (Mean=6.02; SD=1.79).

Interdisciplinary and Cross-culture Comparison. To compare disciplinary differences, only faculty participants (N=51) are included from the selected disciplines and countries: Humanities, 4; Journalism, 8; Information Science, 9; Computer Science 14 (US), 10 (China); Engineering, 6. Doctoral participants are not included due to a lack of participants in certain disciplines. Table 1 shows the differences in the perception of importance of IICTs to research. The *Web* is the most important IICT in Computer Science, Engineering, and Information; the second most important in the

humanities. *Email* is the most important in the humanities in the US and Computer Science in China, and the second most important one in Engineering and Computer Science in the US. *Database* is the most important one in Journalism, and the second most important one in Information Science. *Online library catalog* (*OPAC*) is not the most important one for any discipline, but the second most important one in Journalism. To Chinese participants (CS), the *Web* is notably much less important than *email*, *digital library*, or *e-journal*.

The extent to which information needs were satisfied by either digital resources or print resources shows remarkable differences (Figure 1). Although there are only four participants in the humanities, the trend is in consistent with Ge's findings (2005). She found that academics in the humanities use about 40% e-resources to satisfy research information needs while those in social sciences use more than 63% e-resources.

Two important factors for nonuse are availability of the resources and usability of the IICTs. Other factors include time needed to learn a new IICT, stability of IICTs, and preference of information habits such as reading.

Different IICTs are used to support different IS activities. For general IS activities, participants mainly use the Web for *monitoring* and *browsing*. Folders and subfolders are the most mentioned methods of *information managing*. Most participants felt it a big challenge to organize their personal digital collections. The primary method is saving copies in different folders and subfolders. A few Computer Science faculty researchers manage the retrieved information objects quite differently. One noted: "I change computers often. There is no point to save them permanently. If I need them again, I will search the Web." Regarding ephemerality of the Web objects, they are not concerned. Here is a comment: "Important stuff will be there or in publications one way or another." An Engineering faculty member commented: "I keep everything and periodically archive the digital files (pdf, etc.). ... I prefer electronic copies, although there will never be 100% information coming in digital format. I read online too. For example, I read and comment my doctoral dissertations online. ... I don't have problems finding the files I saved in the folders."

For task-based IS activities, *starting* is not a critical activity for most researchers because they usually build new projects on previous research. Focused *searching* is carried out in databases and library catalog. *Accessing* varies from e-journals to authors' WebPages (or emailing the author). Most researchers follow the references (backward *chaining*) and few follow citations (forward *chaining*) using CiteSeer or Google Scholar. Projects never really end, as many believe. The *ending* can be seen as finishing the writing. Most rely on traditional publication channels (conferences or journals) to disseminate results or put a copy on personal or project WebPages. No one uses listserv or email to publicize their results. One person commented: "... I stopped mailing out pre-prints. I found far fewer instances where people would publish things almost identical to what I had without a reference to my work [rib me off], but I'm less visible to the community ... You know, good and bad."

Table 1. Comparison of Perceived Importance Ranking of IICTs (N=51)

Rank	Humanities	Journalism	IS[1]	Engr[1]	CS[1] (US)	CS[1] (CN)
(Mean, SD)	N=4	N=7	N=9	N=6	N=14	N=10
Web	*2nd*	*3rd*	*1st*	*1st*	*4th*	*1st*
(2.64, 1.63)	(2.25, 2.00)	(2.69, 1.53)	(2.56, 1.51)	(1.75, 0.80)	(4.25, 2.10)	(2.33, 1.21)
Email	*1st*	*4th*	*4th*	*2nd*	*1st*	*2nd*
(2.65, 1.67)	(1.75, 1.00)	(3.14, 0.90)	(4.13, 1.13)	(1.96, 1.12)	(2.00, 2.49)	(2.40, 1.52)
Dig Lib[2]	*5th (tie)*	*5th*	*6th*	*3rd*	*2nd*	*5th*
(3.80, 1.66)	(4.50, 5.00)	(3.63, 0.75)	(4.69, 1.51)	(4.15, 1.68)	(2.25, 0.71)	(3.78, 1.48)
Database	*4th*	*1st*	*2nd*	*7th*	*5th*	*3rd*
(3.83, 2.19)	(3.33, 4.00)	(2.21, 0.99)	(2.89, 2.26)	(6.17, 1.58)	(4.50, 2.00)	(2.82, 1.66)
OPAC[2]	*3rd*	*2nd*	*5th*	*5th*	*6th*	*6th*
(4.20, 2.06)	(2.75, 3.00)	(2.44, 1.92)	(4.44, 2.24)	(4.92, 1.24)	(5.79, 2.23)	(3.84, 1.28)
E-journal	*7th*	*7th*	*3rd*	*4th*	*3rd*	*4th*
(4.26, 1.65)	(6.00, --)	(5.29, 1.11)	(4.06, 1.90)	(4.20, 1.32)	(3.75, 2.05)	(3.58, 1.71)
FTP	*8th*	*8th*	*7th*	*6th*	*8th*	*7th*
(5.55, 2.53)	(7.00, --)	(7.50, 2.12)	(5.00, 2.16)	(4.96, 2.78)	(6.44, 2.59)	(4.74, 2.37)
Listserv	*5th (tie)*	*6th*	*8th*	*8th*	*7th*	*8th*
(5.99, 1.92)	(4.50, 5.00)	(5.17, 1.72)	(5.93, 1.97)	(6.33, 2.16)	(6.38, 2.87)	(6.96, 1.20)

Note: Rank is based on Mean score (1 is the highest score for the most important one). The first column is ordered by the total mean scores. N = 51 faculty participants in the selected disciplinary groups.

[1] IS: Information Science; Engr: Engineering; CS: Computer Science; CN: China
[2] *OPAC*: Online Public Access Catalog or online library catalog; Dig. Lib.: Digital Library

6. Conclusions

This project is ongoing and the tentative conclusions are based on the preliminary results of 65 participants. The most used IICTs for research include the *Web*, *email*, *database*, *e-journal*, *online library catalog (OPAC)*, and *digital library*. New IICTs such as *wiki*, *instant messaging* are not yet widely adopted for research. Older IICTs such as *FTP* are being replaced by newer IICT such as *Web*. There are differences across disciplines and cultures in terms of importance ranking and the amount of use of digital resources. It is found that academic researchers in Computer Science and Engineering are early adopters of the Internet for research and heavy users of digital sources, which results in increased availability of information resources on the Internet. In contrast, researchers in the humanities tend to use less digital resources for various reasons, including reliability and availability of information resources, as well as trustworthiness and usability of IICTs. Chinese academic researchers in

Computer Science rank the *Web* much less important than *email*, *digital library*, and *e-journal*. They also use slightly less digital resources than their US counterparts do, which is mainly due to the availability of the digital resources. Although the Internet in China had a later start, it is diffusing and developing rapidly. All participants experience difficulties in organizing digital information objects they collected. Moreover, few use personal information management tools. The "digital divide" between hard sciences and humanities is more obvious than that between the two selected cultures of the same discipline. The results will be updated as more data are collected to include more participants from different countries.

Acknowledgements

Sincere gratitude goes to all the participants and research assistants Yan Zhang, Steve Bales, Nadine Hawke, Eyasu Gutta, Rachel Kirk, and Sheri Edwards for research assistances in data collection, analysis, and literature.

References

[1] Gunkel, David. "Second Thoughts: Towards a Critique of the Digital Divide," New Media and Society 5, no. 4, pp. (December 2003): 499–522.

[2] Ellis, David. "A Behavioral Approach to Information Retrieval Design." Journal of Documentation 45, no.3 (1989): 171-212.

[3] Ellis, David. "Modeling the Information-Seeking Patterns of Academic Researchers: A Grounded Theory Approach." The Library Quarterly 63, no.4 (1993): 469-486

[4] Ge, Xuemei. "Information-seeking Behavior of Social Sciences and Humanities Researchers in the Internet Age" Masters Thesis. University of Tennessee Knoxville. (2005). Available at http://etd.utk.edu/2005/GeXuemei.pdf

[5] Meho, Lokman I., and Helen R. Tibbo. "Modeling the Information-Seeking Behavior and Use of Social Scientists: Ellis's Study Revisited." *Journal of the American Society for Information Science and Technology* 54, no.6 (2003): 570-87

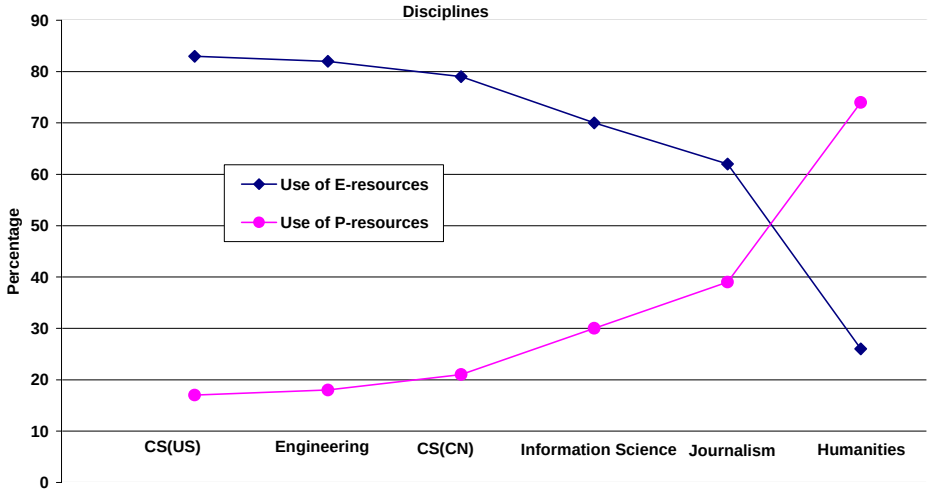

Figure 1. Satisfaction of Information Needs with Electronic Resources vs. Print Resources