

Title: Mediating Off-Site Electronic Reference Services: Human-Computer Interactions between Librarians and Web Mining Technology

May Y. Chau/ Assistant professor/Agricultural Sciences Librarian/Oregon State University, Oregon, USA.
121, The Valley Library, Corvallis, OR. 97331-4501. E-Mail: may.chau@orst.edu Fax:541-737-8224

Introduction:

The Internet enables global dissemination and retrieval of information at any time, as a result, it has a tremendous impact on how libraries provide reference services to this new web-based clientele. One impact is the increase in off-site library users who make in-person reference interviews impossible. To serve this new clientele, academic libraries are setting up electronic reference services to meet off-site information needs. This new approach is not without challenges. It requires effective human-machine connections to provide smooth reference transactions since librarian-patron contact happens solely through computers. The most challenging aspect of providing electronic reference services to off-site users is deciding when patrons need help. If their needs are not detected, many unsuccessful searches will go undiscovered and end-users may never find the information they were seeking. To meet this challenge, pertinent information technology is needed to track web traffic patterns and user behavior (e.g. web mining), addressing immediately the difficulties encountered by end-users. The machine's capability to track potentially unsuccessful searches depends on human experts' ability to identify trouble signs, and add value to the electronic transaction experience with their experiences and skills. This paper's emphasis is on human-computer connections relating to reference librarians' application of web mining technology to transfer their experiences and skills from the reference desk to the web environment. It examines web mining techniques and their potential to assist different library functions. A scenario is used to conceptualize a model that can effectively apply web mining technology to perform various functions of electronic reference services. It also proposes a methodology that librarian can adapt in offering mediation during the search at patrons' discretion. The main goal is to provide more seamless electronic reference transactions. Discussion of library professionals' roles in human-computer interactions in the electronic world follows.

Human-Computer Connections:

Machine as tools can only perform as well as human can apply them. In the realm of electronic reference services, the success of human-machine connections

depends on library professionals' abilities to transfer their experiences and skills from the reference desk to cyberspace. These abilities include: (a) identifying problems that exist in electronic references services (b) applying one or more technologies to solve the problems and, (c) adding value to the products of this technology [1] to better serve off-site end-users.

First, however, several problems need to be addressed.

Problems:

The challenge for electronic reference services is to know when end-users need help, since no direct contact between librarians and patrons occurs. The difficulty is to offer help unobtrusively, 24 hours a day, with existing staff.

Possible Solutions:

A logical approach is to apply web mining techniques to track user traffic so that navigation patterns and behavior can be identified. The results should be analyzed by librarians, and appropriate action determined.

Potentially Useful Web Mining Approaches:

Web mining is used for the "discovery of meaningful business correlations and trends by integrating and analyzing Internet/Intranet web traffic information and traditional business data" [2]. The evaluated traffic information extracted by this technology is used in generating reports for decision-making guidelines, such as when new products are created to meet customer needs [3]. It is an important tool for the commercial sector in providing better customer service. The potential for using web mining approaches in electronic reference services will be examined in this paper. There are many strategies incorporated into document retrieval and web traffic analysis. The most common ones are: (a) web agents, (b) pattern discovery, and (c) web personalization.

Web Agents: web agents retrieve, filter, and categorize web documents guided by user profiles (e.g., WebACE).[4] This approach is a form of

artificial intelligence that "acts autonomously or semi-autonomously on behalf of a particular user, to discover and organize web-based information" [5]. For example, ShopBot can "retrieve product information from a variety of vendor sites using only general information about the product domain" [6]. This approach is extremely valuable in collection development of the online resources provided by electronic reference services. Academic libraries have a mission to support teaching and research trends in their home institutions. In the pre-electronic era, librarians spent hours screening book or journal announcements, advertisements and reviews, to select and purchase pertinent sources. This was the standard procedure used to accumulate a collection that would provide appropriate support for degree programs offered by specific departments. Since the Internet has become a major tool for research and communication, libraries have become aware of the convenience and timeliness of online documents. Some commercial digital libraries exist, such as Westlaw, Lexis/Nexis, ISI, Elsevier and JSTOR, [7] etc. Academia is building digital libraries as well (e.g., UCSB Digital Library Project) [8]. Web agents are effective tools in assisting libraries in building online libraries. The agents automatically retrieve, filter and organize web resources, based on profiles set up in different disciplines and research areas.

Pattern Discovery: Pattern discovery extracts relationships from knowledge collected by detecting association rules and sequential patterns in large web data repositories. The knowledge extracted can be used to "restructure a web site for increased effectiveness, for better management of workgroup communication in intranets, and analyzing user access patterns to dynamically present information tailored to specific groups of users" [9]. In academic research, it is important to be aware of professional trends. In recent years, obtaining grants or funding has become highly competitive. Proper preparation and excellent strategies are necessary to obtain research funding. Pattern discovery can be a tool in discovering what research trends exist in specific fields, for example, the discovery of which keywords are most frequently used to retrieve certain types of information in certain subject areas [10]. In addition, pattern discovery can help reveal user behavior (e.g. how often they visit a particular site.)

Web Personalization: Web personalization is "any action that makes the web experience of a user personalized to the user's taste" [11]. Elements required for personalization include modeling and categorizing web pages and users, matching between and across web pages and users, and determining

actions to be recommended for personalization [12]. A common approach is to rely on human participation in data collection to build a user profile [13] that meets individual needs. The other is automatic personalization, which does not rely on human participation, but preprocesses web usage and groups URL references to obtain user transaction patterns, and makes recommendations [14]. In electronic reference services, web personalization can customize individual research components, enabling searcher to build individual mini-libraries for long-term research. Knowledge extracted from pattern discovery tools can be used to establish user profiles, or in this context, establish a search history. A typical search history tracks search strategies by recording terms used and combinations of those terms. Search histories can be downloaded to the end-users' own computer for future use. Web agents can rely on these profiles to automatically retrieve, filter and categorize information continuously for specific topics. Web mining technology can discover user access patterns (or user habits), and automatically retrieve document according to profiles derived from those patterns. It can be used to address the needs of off-site patrons, to expand the collection of web sources, and as a monitoring device that connects librarians and end-users. To illustrate the potential of this technology, the following scenario envisions ways librarians can mediate between users and machines, and the effect of their mediation on customer service.

Scenario:

An Academic library plans to establish an electronic reference service center on the web. This is a joint project between two libraries located in opposite time zones (e.g., day time for Library A is night time for Library B). This center will provide access to pertinent information for the university community, residents of the entire state, and information seekers worldwide who are affiliated with the university. Its goal is to provide access to information 24 hours a day and seven days a week. Resources include commercially produced research databases, electronic journals, web sources and various library services.

The staff brainstormed the possibilities of providing timely help to off-site users and realized that unless they could accurately recognize signs of trouble during a search, it would be difficult to provide assistance promptly. Their concern was that if help were not available quickly, patrons would become frustrated and end their searches without finding the information they needed.

Web mining is the logical technology for detecting signs of trouble because of its ability to track user access and behavior. Two elements-- the number of retrieval and the time elapsed between searches -- should be used to determine the effectiveness of the search. A pattern discovery tool records interrelationships between these two elements. By comparing the results of combining numbers of retrievals and time elapsed with librarians' experiences, a mechanism for detecting trouble signs is drafted.

Trouble signs:

1. Too many records (hundreds or thousands) and foreshortened elapsed time between searches. For example, the web usage log shows a certain patron retrieved thousands of records on a specific topic, and made many other searches within a short time. The elapsed time between each search is less than one or two minutes. From observations at the reference desk, it has been determined that a patron who retrieves thousands of hits, performs a search that is not effective because too much information results and possibly many unwanted returns appear. If this kind of result persists in consecutive searches, it indicates that refinement of search strategies or narrowing the search's focus should be done, and patrons may need help to do so.

2. Too few documents retrieved and patrons spend a long time on the same topic. Elapsed time between each search can be long or short. It may mean that the end-user is having difficulty finding information on a specific topic. The patron may need help in broadening the search, as guidance from librarian in the library would have done.

3. Reasonable amount of retrieval, but very short elapsed time. This may mean that the search retrieves a reasonable amount of information but that it may not match the end-user's need exactly. It may require a more in-depth reference interview to help clarify research objectives.

These signs of trouble are derived from librarians' own experiences helping patrons use electronic research databases in the library. There are many other variables that may affect such derivatives. Therefore, the staff members thought it best to allow patrons to decide whether they want help or not. Librarians suggested the appearance of a button on the screen whenever trouble signs are detected. Then patrons could decide to turn it off or press it to connect to a librarian.

Two Facades of Design and the Linking Interface:

The design of the library electronic reference center is divided into two facades. One derives from the library, the other from the clients. The design determines how the electronic reference service center should be presented to the patrons and how effectively it functions.

The Library: An accessible electronic reference center requires a strong, high-quality web resource collection, commercial databases and electronic journals. Web agents' ability to retrieve, filter and organize web documents is best applied to online resources' collection development. This activity relies on profiles, therefore, subject area reference librarians need to compile subject profiles that instruct the machine how to do the work. These profiles are hierarchical arrangements of topics related to a specific subject discipline. They are dynamic and need constant updating.

Pattern discovery tools (Web Miner, Web Trends, etc.) can be used to track usage activities with respect to retrieval numbers, elapsed times and keywords used. These two elements— numbers of retrievals and elapsed time — are used to detect signs of trouble, as determined above. In addition, statistics on keywords used may suggest a demand for information on a particular topic. These statistics can also be used to establish a user profile analogous to the "search history" provided by subject research databases. Search history tracks terms (e.g., keywords, subject terms etc.) and their combinations during searches, providing records of individual search strategies used by a certain patron. This search history will not be stored in the computer but will be erased every time the connection ends, so that no permanent storage space is required. End-users are welcome to download the "search history."

The Library also negotiates the collaboration of libraries in different time zones, extending the hours for consultations. The electronic reference center is open from 8 a.m. to 10 p.m., leaving ten hours from 10 p.m. to 8 a.m., without staff. Coverage of these hours will be supplied by the partner library, so that 24-hour service remains in place.

End-Users: End-users, who log into the electronic reference service center, can search any resources provided by the library. They can search the commercial subject databases, e-journals and web sources at the same time. They can use keyword,

author, title or subject searches. There is also a "Help" button in the upper corner of the screen, and end-users can click it to activate the phone line at the electronic reference center. The ideal situation would be librarians and end-users communicating via speaker-phone. End-users would wear headsets. In addition, if the web mining tool were to detect trouble signs during a search, a message would appear, prompting end-users to click the "Help" button to contact a librarian any time, day or night, seven days a week.

End-users can download their profiles to their computers for future use. These profiles are deleted automatically after logout or every 12 hours. Every time the end-users decide to resume a previously conducted search, they can use the downloaded profiles to initiate it. The web agent tool embedded in the computer will automatically search the former subject and web personalization tools, updating the user's mini libraries.

Interface: The interface between the two facades is directly linked through the "Help" button, which is activated whenever the web-traffic-tracking device detects trouble signs during a search. This interface is important in forming the human-machine connection, resulting in end-users' contact with a live person instead of an unsympathetic machine. The interface involves the knowledge and expertise of subject reference librarians and technical staff, and is crucial in this high-tech world.

Reality:

For implementation of the electronic reference center to proceed as planned, many issues need to be dealt with. First are the fears that end-users' feel about off-site searching, due to intensified competition for resources and constant introduction of new technologies [15]. It is a continuing battle to assure that all searchers, even those off-site, get personal service. The second issue is to determine which mining tool is suitable and cost-effective. This requires much vendor research, which results in negotiations with appropriate vendors. Thirdly, it will be difficult, for both political and practical reasons, to establish partnerships with other libraries from correlating time zones. This approach is essential, nevertheless, to combat limited budgets and the shortage of staff. In addition, there will be copyright and licensing issues to deal with. Most importantly, the effectiveness of using the two elements (numbers of retrievals and elapsed times) to detect trouble signs will need validation from end-users. This scenario, therefore, offers a conceptual model that potentially

can provide an effective electronic reference service, but further investigation is needed.

Conclusion:

Ideal patrons might be described as information literates who will have little difficulty in adapting to a well-designed system. These model clients do not exist, and therefore, end-user education remains essential [16]. The expectation of relying on technology alone to address all information needs is not realistic at this stage, especially when dealing with the complex nature of academic research.

Technologies are most effective when queries are limited to those that can be simply described in machine-understandable ways, and remain almost static [17]. Library professionals are key players in facilitating human-machine interaction issues. For centuries, librarians have selected and organized information to provide access to information seekers. This was the mission of librarianship in the pre-electronic age, and this same mission continues in this new millennium, with some adaptations. The role of the librarian in the Internet environment is suggested as intermediary, guide facilitator, educator and website builder [18]. Librarians provide pointers that guide patrons to search and evaluate relevant resources. They also facilitate the arrangement of the information infrastructure, which includes network access and software licensing. Librarians educate end-users to employ various tools and they build websites that help patrons locate information and library services [19]. In addition, since information seekers sometimes do not ask direct questions about their research, it becomes more difficult for machine alone to handle all these mysterious queries. An important component of the research task for librarians is ascertaining what information is really necessary for patrons, in spite of the confusion of languages. A skillful librarian can extract useful information from patrons during reference interviews, and clarify their search directions. Another aspect of reference interviews is dealing with patrons' emotional stress at critical times [20]. Many library patrons come to the reference desk in an intense emotional state. For example, a faculty member has a publication deadline and needs specific information at the last minute. Information technology cannot assist in helping patrons regain their composure so that they can find the information necessary to complete their project. It becomes essential for librarians to transfer their valuable skills and experiences from the reference desk to electronic reference services, facilitating human-computer interactions that produce satisfactory results. A device or process that can connect off-site users to a live librarian is a step

toward successful human-computer interactions and deserves more attention. The device suggested by this paper used is a "Help button," available to patrons on demand, which offers assistance unobtrusively, and is used in conjunction with automatic detection of trouble signs. This is an example of applying librarians' expertise to an electronic world. To provide seamless information services, library professionals need to get involved with the information industry, stay abreast of new technological developments and to be aware of opportunities for offering the wealth of their experiences and skills. Industries can benefit from the abundance of knowledge library professionals have accumulated in decades of service. Therefore, library professionals should be regarded as assets to the emerging trend of human-computer interactions and frequently be included as resources in this kind of research.

References:

1. May Y. Chau, 1999. "Web Mining Technology and Academic Librarianship: Human-Machine Connections for the Twenty-First Century," *FirstMonday*, issue 4, number 6 at http://www.firstmonday.org/issues/issue4_6/chau/index.html
- 2.3. Marketwave Corporation. "Web Mining: A New Generation of Web Traffic Analysis," Marketwave Corporation Seattle: WA.
- 4.5. Bamshad Mobasher, 1999. "WebACE: A Web Agent for Document Categorization and Exploration," Bamshad Mobasher Research Overview (July), at <http://maya.cs.depaul.edu/~mobasher/Research-02.html>
6. Bamshad Mobasher, 1997. "Agent-Based Approach," at <http://maya.cs.depaul.edu/~mobasher/webminer/survey/node4.html#SECTION0002110000000000>
7. Roxanne Missingham, 1999. "Digital Libraries 99: Fourth ACM Conference on Digital Libraries," *Australian Academic & Libraries*, volume 30 issue 3 (September), p219-221.
8. King Communication Group, 1999. "UCSB Digital Library Project Receives \$5.4 Million," *High Performance Computing & Communications Week*, (August), pNA. <http://maya.cs.depaul.edu/~mobasher/Research-01.html>
9. 10. Bamshad Mobasher, 1999. "WEBMINER: A System for Pattern Discovery from World Wide Web Transactions," Bamshad Mobasher Research Overview (July), at <http://maya.cs.depaul.edu/~mobasher/Research-01.html>
- 11.12.13.14. Bamshad Mobasher, Cooley Robert, 1999. "Automatic Personalization Based on Web Usage Mining," Bamshad Mobasher Research Overview (July), at <http://maya.cs.depaul.edu/~mobasher/personalization/index.html>
15. Payne, David J., William A. Minneman, 1999. "Apply a Human Solution to Electronic Fears," *HR FOCUS*, volume 74, issue 4, (April), p51.
16. Jennifer Burrell, Mylee Joseph, 1999. "Now for the Hard Part: End User Education," *Australian Public Libraries and Information Services*, volume 12, issue 3 (September), p105.
17. Jakob Nielson, 1998. "Personalization is Overrated," Jakob Nielson's Alertbox, (October) at <http://www.useit.com/alertbox/981004.html>
- 18.19. Alastair Smith, 1999. "Beyond the Basic: Role of Librarian in the Internet Environment," A continuing education course offered by Victoria University of Wellington, New Zealand at <http://www.vuw.ac.nz/~agsmith/beyond/2role.htm>
20. Moody Janette and Elizabeth Carter, 1999. "Application of the Cognitive Interview by the Reference Librarian," *Reference & User Services Quarterly*, volume 38, number 4, (Summer), p389-393.