The School of Information Studies Faculty Scholarship

School of Information Studies (iSchool)

2-2005

A Framework for Internet Archeology: Discovering use Patterns in Digital Library and Web– based Information Resources

Scott Nicholson
Syracuse University, scott@scottnicholson.com

First Monday, Volume 10, Number 2 - 7 February 2005

Home > Volume 10, Number 2 - 7 February 2005 > **Nicholson**

ƒ i ® s † m ¤ ñ d @ ¥

PEER-REVIEWED JOURNAL ON THE INTERNET

A framework for Internet archeology:
Discovering use patterns in digital library
and Web-based information resources

by Scott Nicholson

Abstract

Archeologists use artifacts to make statements about occupants of a physical space. Users of information resources leave behind data–based artifacts when they interact with a digital library or other Web–based information space. One process for examining these patterns is bibliomining, or the combination of data warehousing, data mining and bibliometrics to understand connections and patterns between works. The purpose of this paper is to use a research framework from archeology to structure exploration of these data artifacts through bibliomining to aid managers of digital libraries and other Web–based information resources.

Contents

Introduction

For hundreds of years, scientists have worked to understand our history through the "recovery, systematic description, and study" [1] of artifacts left behind by a particular culture. This field of study, archeology, has several tasks, one of which is to reconstruct the "lifeways of the peoples responsible for the archeological remains" [2]. Archeologists may know little about the people themselves who occupied a physical space; instead, they examine artifacts left behind for patterns to understand the communities who lived there.

There are similarities between a physical place and the virtual space supported by the Internet. Communities of people form, grow, evolve, and dissipate in this virtual space. As users travel through this virtual space, they leave behind data–based artifacts. These shards of virtual pottery, such as searches and browsing behavior, and burial mounds of dead discussion groups, hold valuable information — if the pieces can be collected, cleaned, organized, and examined. This information can give "Internet archeologists" an idea of the communities and cultures that have existed in virtual spaces.

The purpose of this paper is to discuss one current framework for archeological thought and

explore how that framework can be applied for the systematic exploration of digital library and other Web–based information resource use. This structure is useful to those wishing to apply a culture of assessment to their organizations and to continuously seek ways to improve their information services.

Traditional archeology and digital library evaluation

Traditionally, archeologists gather large amounts of material from a site, describe it, and then present their findings. This focus on the collection itself is akin to many evaluation studies of digital library services. The usage data are collected, grouped, and described much in the same way that archeological artifacts were collected, tagged, and grouped through traditional archeology (Johnson, 1999). Many contemporary digital library evaluations published today fall into this "gather and describe" cycle [3].

New archeology

New archeology came out of the realization that the collection of more artifacts was not leading to higher–level understanding of phenomena. This caused a shift in thinking from "what what found" to "why it was there." Binford (1968) was a leader in moving the field toward using hypotheses to guide the collection of artifacts. Thinking about these explorations from the viewpoint of generalizable hypotheses can guide Internet archeologists to move from artifact collection and description to building knowledge.

This collecting of more and more artifacts and aggregations represents the current state of many digital library evaluation projects. Reports that look at data from only a single service, with no consideration of larger implications, may meet the needs of a local library but do not help advance the field of librarians. Librarians and researchers are collecting more and more artifacts, but lack the structure to pose larger questions that could be explored through data.

> **As users travel through this virtual space, they leave behind data–based artifacts. These shards of virtual pottery ... hold valuable information — if the pieces can be collected, cleaned, organized, and examined.**

This need to move from a practical gather–and–describe evaluation toward a hypothesis–based exploration to improve the science of librarianship has been voiced explicitly and implicitly by other library scientists. As McClure (1989) states, "library and information science fosters little research that is intended to produce 'knowledge for the sake of knowledge'" [4] and instead focuses on the gap between the generalizable research of library scientists and the applied action research desired by librarians. He argues for the need for ways to increase the impact of research on libraries. Tenopir (2003) examines many significant large–scale library evaluations and find that most of them draw conclusions only about individuals or specific groups of users. She also develops numerous generalizations based on the collection of studies; this is an example of the next step in generalizing results from traditional library evaluations and could serve as the inspiration for many testable hypotheses.

Combining the frameworks: The Hypothetico–Deductive–Inductive (HDI) scientific cycle

This cycle, first outlined by Kenemy (1959), was applied to archeology by South (1977) and provided the bridge between traditional and new archeology. Traditional archeology focused on describing and finding patterns within the data, and new archeology started with a problem and sought data to support or refute hypotheses.

South presented the stages of the HDI cycle as:

- Induction (Pattern Recognition),
- Theory (Lawlike Generalty),
- Deduction (Logical Analysis),
- Prediction (Hypothesis), and
- Verification (Testing) [5].

In an archeological site, just as in the logs of a digital library, there is a large amount of data available. The first step is to collect samples of data from around a site and explore those data for patterns. This is where bibliomining is important. Bibliomining is the application of tools from data mining and bibliometrics to discover non–trivial and useful patterns in large amounts of data (Nicholson and Stanton, 2003). Inspired by these patterns, the researcher creates basic generalizations about the data. Research questions are created to explore the hypotheses, and then additional data are gathered to test those hypotheses. These data may come from the same source or may require different sources. This method may support or refute the hypotheses, which has the effect of building the knowledge base for the field.

Postmodern archeology

Postmodern archeology, one form of postprocessual archeology, eschews the process that produces generalizations about a culture and instead focuses on the individuals within that culture (Praetzellis, 2000). Any culture is made up of individuals, each of whom makes his or her own life decisions. In order to understand a culture, a researcher must understand the individuals who make up that culture.

Due to the focus on users by information science researchers such as Dervin and Nilan (1986), there is considerable research on the importance of users in artifact–based evaluation in the library evaluation process. Saracevic and Kantor (1997) were clear in the importance of involving users to learn about both the relevance and usefulness of the resources retrieved from a system. The holistic framework for library measurement in Table 1 (Nicholson, 2004) shows the relationship between bibliomining, or the internal view of the use of the system through data–based artifacts, to other areas of measurement, which include:

- The internal view of the system, which focuses on what information and services are offered;
- The external view of the system, which involves users to learn if the information provided matches the request by a user; and
- The external view of the use, which involves users to learn if the information was useful to the user in meeting the information need.

	Topic	
Perspective	System	Use
Internal (Library System)	Procedures Standards	Recorded interactions with interface and materials, Bibliomining
External (User)	Aboutness Usability	Knowledge states Value of works

Table 1: : Measurement matrix from Nicholson's holistic measurement framework.

The important focus for postmodern Internet archeology is the bottom row (the external/user view). One advantage in examining current information resources is that users of the system are still available to be involved in the evaluation process. There are several scenarios for involving users in the evaluation:

- If a login procedure is used, the actual users may be available to discuss how well the results matched the query and how useful the information was in meeting their need. In order to capture these views, a survey immediately after interaction with the system will prevent relevance decisions from changing over time.
- If there is no login procedure, building in elicitation methods during the process of information delivery may help to capture the relevancy decisions of a user. Asking for an e–mail address at this time and following up with a survey about the usefulness of the information can capture the needed data.
- Another approach is to locate research subjects that are typical users of the system. These users can either be given queries or can search on their own needs, and then measures can be collected through this process. This is the least desirable scenario, but may be the only one available.

While bibliomining can aid in the understanding of what patterns of usage exist, it cannot aid evaluators in understanding what documents people find relevant (as compared to what document they used). In addition, the results from bibliomining are constrained by the system used; if a feature is not offered by a system, no information will be gathered about that feature. This is why bibliomining, like archeology, cannot produce the truth, as the truth about the information–seeking situation lies within the individual doing the seeking. Bibliomining can only recover the facts about what interactions an individual has had with the system and cannot recover anything about the user's mental state. Therefore, gathering these other measures allows the researcher to bring in qualitative elements to the quantitative bibliomining process.

Completing the cycle

Adding postmodern archeology to the existing hypothetico–deductive–inductive cycle allows researchers to move beyond the data in the system to gain a better understanding of users. Processes have been separated from products in South's cycle in order to better fit the needs of a digital library. By employing the full cycle, researchers can move from describing to improving the knowledge base about library users. The significant processes of this cycle are:

- Collection: Gathering artifacts about library services, users, and resources,
- Induction: Pattern recognition through bibliomining and visualization,
- Deduction: Logical analysis of patterns to produce generalizations,
- Prediction: Creation of hypotheses from generalities, and
- Testing: Research developed to test hypotheses involving both data and users.

An important note is that one can never know what the user was thinking and experiencing from only the artifacts in the system; even talking to the user after the process will not provide an accurate portrayal of the thought process that the user went through. Just as an archeologist does not know the truth about the people who lived in a culture, we cannot know the truth about our users from only their artifacts. We must work directly with users in order to test our hypotheses about their behavior.

While bibliomining can aid in the understanding of what patterns of usage exist, it cannot aid evaluators in understanding what documents people find relevant.

This resulting archeological framework of explanation is fundamentally an application of the

traditional scientific process. Librarians and others providing Web–based information resources do not regularly use hypothesis–based research projects when evaluating their services. The result is that the science of librarianship has not grown and developed as other sciences have grown. The benefit of using the archeological approach is that it may aid those not accustomed to the scientific method to conceptualize how this method applies to their own information service.

Conclusion

Archeologists recognize that their craft is as much art as science, as giving meaning to a collection of artifacts requires a number of assumptions and guesses and does not provide the truth about the mental state of the users. We can draw an essential point from this for exploration of digital libraries and other Web–based information resources. Collections of data–based artifacts tell us an important part of the story, but the discovery of these patterns through bibliomining is just the beginning. By developing generalizations, creating hypotheses about library use, and testing those hypotheses through research involving data and users, researchers can move beyond descriptions and advance our understanding of users of digital information. 🔴Ⓜ

About the author

Scott Nicholson is an Assistant Professor in the School of Information Studies at Syracuse University (N.Y.) and a Research Scientist for the Information Institute of Syracuse. His research supports library measurement and evaluation through the bibliomining process. Other works by Dr. Nicholson can be found at http://bibliomining.com/nicholson.
E–mail: scott [at] scottnicholson [dot] com

Notes

1. Clark, 1978, p. 12.

2. Binford, 1968, p. 6.

3. Bollen and Luce, 2002; Jones and Sumner, n.d.; many examples in the literature review of Tenopir, 2003.

4. McClure, 1989, p. 283.

5. South, 1977, p. 15.

References

L. Binford, 1968. "Archeological perspectives," In: S. Binford and L. Binford (editors). *New perspectives in archeology.* Chicago: Aldine. pp. 5–32.

J. Bollen and R. Luce, 2002. "Evaluation of digital library impact and user communities by analysis of usage patterns," *D–Lib Magazine,* volume 8, number 6, at http://www.dlib.org/dlib/june02/bollen/06bollen.html, accessed 11 February 2004.

D. Clark, 1978. *Analytical archaeology.* Second edition. New York: Columbia University Press.

B. Dervin and M. Nilan, 1986. "Information needs and uses," *Annual Review of Information Science and Technology,* volume 21, pp. 3–33, and at http://communication.sbs.ohio-state.edu/sense-making/art/artabsdervinnilan86arist.html, accessed 23 January 2005.

M. Johnson, 1999. *Archeological theory: An introduction.* Oxford: Blackwell.

C. Jones and T. Sumner, n.d. "Evaluation of the National Science Digital Library," at http://www.uclic.ucl.ac.uk/annb/DLUsability/JonesSumner5.pdf, accessed 11 February 2004.

J. Kenemy, 1959. *A philosopher looks at science.* New York: Van Nostrand.

C. McClure, 1989. "Increasing the usefulness of research for library managers: Propositions, issues, and strategies," *Library Trends,* volume 38, number 2, pp. 280–294.

S. Nicholson, 2004. "A conceptual framework for the holistic measurement and cumulative evaluation of library services," *Journal of Documentation,* volume 60, number 2, pp. 164–182.

S. Nicholson and J. Stanton, 2003. "Gaining strategic advantage through bibliomining: Data mining for management decisions in corporate, special, digital, and traditional libraries," In: H. Nemati and C. Barko (editors). *Organizational data mining: Leveraging enterprise data resources for optimal performance.* Hershey, Pa.: Idea Group Publishing, pp. 247–262.

A. Praetzellis, 2000. *Death by theory: A tale of mystery and archaeological theory.* Walnut Creek, Calif.: Altamira Press.

T. Saracevic and P. Kantor, 1997. "Studying the value of library and information services. Part I. Establishing a theoretical framework," *Journal of the American Society for Information Science,* volume 48, number 6, pp. 527–542.

S. South, 1977. *Method and theory in historical archeology.* New York: Academic Press.

C. Tenopir, 2003. *Use and users of electronic library resources: An overview and analysis of recent research studies.* Washington, D.C.: Council on Library and Information Resources.

Editorial history