Library and Information Science Research 39 (2017) 213–223

Contents lists available at ScienceDirect

Library and Information Science Research

journal homepage: www.elsevier.com/locate/lisres

A clickstream data analysis of Chinese academic library OPAC users' information behavior

Tingting Jiang[a,*], Yu Chi[b], Huiqin Gao[c]

[a] School of Information Management, Wuhan University, Luojiashan, Wuchang, Wuhan 430072, China
[b] School of Information Sciences, University of Pittsburgh, 135 N Bellefield Ave, Pittsburgh, PA 15260, USA
[c] Department of Computer Science, W & N, Vrije Universiteit Amsterdam, De Bolelaan 1081a, 1081HV, Amsterdam, Netherlands

ABSTRACT

Chinese academic libraries have been devoting great effort into introducing next-generation online public access catalogs (OPACs) in order to provide a better user experience. However, there is a lack of empirical research on their usage. In this study, a transaction log file from a typical next-generation OPAC, the Wuhan University Library OPAC, formed the basis for an investigation of users' information behavior, using a clickstream data analysis framework. The 26,732,368 clickstream records in the original log file were cleaned, parsed, coded, then analyzed at the footprint, movement, and pathway levels. The results showed that the users relied heavily on the single-box simple search interface, seldom involved themselves in an exploratory search process, and preferred page navigation over search refinement when interacting with search results. The OPAC was used as a lookup tool to locate known academic resources and as a personal information management tool to utilize basic library services, instead of as a discovery tool as intended.

1. Introduction

Next-generation online public access catalogs (OPACs), which have been on the rise since 2000, distinguish themselves from traditional ones by introducing discovery tools that feature rich sources and content, improved search functionality, and customizable search interfaces (Breeding, 2009; Liu, 2009; Yang & Hofmann, 2011). The scope of search is no longer limited to the physical collections of libraries, but also extends to digital collections, article databases, webpages, and other academic resources. Users are offered a Google-like search experience, including effective querying assistance, better relevance ranking, faster response, and higher consistency. More importantly, exploratory search and social search are enabled through faceted navigation and social networking features, which have dramatically enriched the possibilities of what OPACs can provide (Breeding, 2009; Tam, Cox, & Bussey, 2009; Barton & Mak, 2012).

There has been an obvious growth in next-generation OPACs among North American academic libraries in recent years. Correspondingly, researchers have become increasingly interested in how these OPACs are used, focusing on users' search activities and task performance and how their behavior changes over time, to assess the effectiveness of discovery tools (Niu, Zhang, & Chen, 2014). This body of research uses interviews, surveys, search logs, and experiments to collect users' behavioral data and the various studies have contributed useful empirical findings (Griffiths & Brophy, 2005; Lau & Goh, 2006; Mahoui & Cunningham, 2001; Moulaison, 2011).

Chinese academic libraries also attach great importance to the next-generation OPAC, as evidenced by the updates that a majority of top universities in China have been making to their library systems since 2009, including Peking University,[1] Tsinghua University,[2] Fudan University,[3] and Wuhan University.[4] Most universities adopted discovery tools from popular vendors such as ProQuest's Summon Service[5] Ex Libris Primo[6] and EBSCO Discovery Service[7]; very few libraries have designed and developed their own discovery tools.

* Corresponding author.
E-mail address: tij@whu.edu.cn (T. Jiang).

[1] http://pku.summon.serialssolutions.com/
[2] http://discovery.lib.tsinghua.edu.cn/
[3] http://discovery.fudan.edu.cn/
[4] http://whu.findplus.cn/
[5] http://www.proquest.com/products-services/The-Summon-Service.html
[6] http://www.exlibrisgroup.com/
[7] https://www.ebscohost.com/discovery

http://dx.doi.org/10.1016/j.lisr.2017.07.004
Received 5 June 2016; Received in revised form 29 March 2017; Accepted 6 July 2017
Available online 25 July 2017
0740-8188/ © 2017 Elsevier Inc. All rights reserved.

2. Problem statement

The popularity of next-generation OPACs in Chinese academic libraries has inspired domestic researchers to study them from a library management perspective, focusing on strategies for introducing, implementing, and promoting the use of discovery tools (Han & Liu, 2010; Hu, 2006; Liu, 2009; Si, Shi, & Chen, 2011; Yu, 2010). These researchers sought to provide guidelines for the future development of Chinese academic library OPACs, but their studies were largely based on evaluations of specific system functions according to either self-developed frameworks or on their knowledge of successful experiences with North American or European academic library OPACs.

Very few studies have been based on empirical evidence, and little effort has been devoted to user-centered studies for the purpose of understanding OPAC usage. This study fills this gap by analyzing real users' behavioral data in the setting of a representative Chinese academic library by probing Chinese academic library OPAC users' actual information behavior, in order to reveal the overall usage patterns of the next-generation OPACs. The study is guided by the following questions:

- How do users visit the OPAC as a whole? And how do they use specific functions of the OPAC?
- How do users search for resources of interest in the OPAC?
- What are the characteristics of the pathways along which users navigate through the OPAC?

The results of this study will have practical implications for systems design as well as library management, and will add to the burgeoning body of knowledge on user behavior when using next-generation academic library OPACs.

3. Literature review

3.1. Information behavior in academic library OPACs

Early studies on academic library OPACs have contributed to our understanding of information behavior in these settings, mainly in terms of the known-item search and subject search, which were the main types of search supported (Large & Beheshti, 1997). When bibliographic information about a specific item is known, users can search by title, author, and other fields. By contrast, subject search is more open-ended as users attempt to locate all items related to the subject of interest. Findings related to the proportions of both search types have varied greatly across studies (Hunter, 1991; Wildemuth & O'Neill, 1995). In fact, these search types cannot always be explicitly differentiated from each other. For example, “information retrieval” could be a search for a known book title or a broad subject. It is also possible for a known-item search to evolve into a subject search, which blurs the boundary between them (Hancock-Beaulieu, 1990). Subject search has often been found to have a low success rate because it demands greater expertise in expressing information needs with appropriate queries, and in reformulating queries after retrieving poor results (Connell, 1995; Hunter, 1991). The use of subject headings as a field of topical search has largely been replaced over time by keyword search, which in some cases has been found to produce more satisfactory results (Larson, 1991; Tillotson, 1995).

The introduction of newer discovery tools has evoked much research interest in next-generation OPACs. Academic library users have become increasingly dependent on advanced discovery tools which provide a Google-like search experience (Chickering & Yang, 2014). Recent findings suggest that users prefer a single search box that accepts any keywords (Gross & Sheridan, 2011; Lown, Sierra, & Boyer, 2013), and querying tools such as spell checking and query suggestions (Tam et al., 2009). In addition, some studies have found that users mainly examined only the first search results page (Williams & Foster, 2011), satisfied with the relevancy ranking and quality of the results (Comeaux, 2012; Way, 2010). More importantly, discovery tools are able to create a superior experience of information exploration or discovery for academic library users through faceted navigation and social networking features.

Faceted navigation is based on “a set of category hierarchies each of which corresponds to a different facet that is relevant to the collection to be navigated” (Hearst, 2006, p. 60). The facets for academic library resources usually include author, subject, year, region, language, and so on, and are extracted from the MARC record. One can examine any number of facets in any order and select the category of interest under each facet for further browsing. This approach favors recognition over recall to alleviate cognitive load (Jiang, 2014). Users were found in some studies to rely heavily on facets to distinguish between different types of resources, and considered faceted navigation an intuitive tool that helped them obtain a clear overview of the result space, and narrow down the results (Becher & Schmidt, 2011; Denton & Coysh, 2011; Tam et al., 2009; Williams & Foster, 2011). A series of empirical studies involving various discovery tools consistently reported favorable attitudes toward faceted navigation and provided strong evidence of improved search experience (Fagan, 2010; Walsh, 2012). Studies note that users might encounter difficulties in exploiting the facets, however, so the facets and associated categories need to be designed carefully (Emanuel, 2011; Wynne & Hanscom, 2011).

Social networking features, such as tags, reviews and ratings, and RSS feeds, also play a role in defining next-generation OPACs (Osborne & Cox, 2015). The phrase Library 2.0 (modeled on the now obsolete Web 2.0) was coined to highlight the value of individual users' distributed participation in reconstructing the OPAC (Özel & Çakmak, 2010), though this term has fallen out of use lately. Studies have found that users are not only willing to contribute their knowledge, but also ready to take advantage of such peer-generated content (Sadeh, 2008; Yang & Wagner, 2010). Nevertheless, researchers found through interviews and experiments that the idea of embedding social features into OPACs was somewhat controversial. Many users were actually uncertain about the usefulness of these features and reluctant to adopt them in information seeking. A major reason is that they were accustomed to the simple search interface and not familiar with the newer technologies (Bar-Ilan et al., 2012; Osborne & Cox, 2015; Tam et al., 2009).

3.2. Clickstream data analysis

Information behavior research has depended on a variety of methods to collect users' behavioral data (Case, 2012), and most have been employed to study OPAC users (Large & Beheshti, 1997). Interviews and questionnaires, experiments, and diaries are obtrusive methods featuring the intrusion of the researcher or research instrument. Such intrusion can be problematic because users may not behave in a natural manner if they know they are being studied and researchers may pay too much attention to the behavior they expect to see. In contrast, transaction logs are typical unobtrusive methods that avoid observer effect and bias. They provide trace data about users' interaction with the system in their real-world settings (Jansen, 2009).

The transaction logs stored on Web servers mainly fall into two types, search logs and clickstream data. Search logs originate from users' interactions with a Web search system, such as a general-purpose search engine and an internal search application within a website (Jansen, 2008). They normally contain data fields such as user ID, access date and time, user query, results page, and click-through. In contrast, clickstream data concentrates on clicks instead of queries. Between the time that a user enters and exits a website, every click on a link or a function button will be recorded. A clickstream, hence, is a click path that provides information about the sequence of page requests made by a user when navigating through a website (Montgomery, 2001). Common data fields include user ID, date and

Download English Version:

https://daneshyari.com/en/article/5123875

Download Persian Version:

https://daneshyari.com/article/5123875

Daneshyari.com