The Journal of Academic Librarianship xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

The Journal of Academic Librarianship

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

Evaluating the Impact of Web-scale Discovery Services on Scholarly Content Seeking

Xianwen Wang[a,*], Yunxue Cui[a], Shenmeng Xu[b]

[a] WISE Lab, Faculty of Humanities and Social Sciences, Dalian University of Technology, Dalian 116085, China
[b] School of Information and Library Science, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3360, United States

ARTICLE INFO

Keywords:
Crossref
Web-scale discovery service
Library discovery service
Referral
ProQuest
Time serial analysis

ABSTRACT

Web-scale discovery service is one of the primary trends in academic libraries. In this research, using referral data, we explore the main referring domains of DOIs, and evaluate the impact of web-scale discovery services on scholarly content seeking. The results show that web-scale discovery services play an important role in directing users to scholarly contents. ProQuest is found to be ranked the first among all the major referring domains. In addition, the time series analysis shows a significant growing trend of ProQuest as a referring service from 2011 to 2016. The monthly trends demonstrate seasonal fluctuations and daily periodicity of referral data. Specifically, there are two low ebbs throughout a year, corresponding to the Summer and Winter vacations. The lowest point in a week is found to be on Saturdays, while the peak normally occurs on Mondays or Tuesdays.

Introduction

The comprehensive transition from print to electronic academic publishing, along with the development of scholarly indexing databases/search engines have changed the information seeking behavior of researchers. Fewer researchers go to the libraries. Instead, they rely more on web-based resources and almost entirely electronic communication of information (Hemminger, Lu, Vaughan, & Adams, 2007).

As gateways to knowledge, libraries have always served as the access points for information and play a fundamental role in supporting education and inspiring innovation (Anunobi & Okoye, 2008; Virkus & Metsar, 2004; White, 2012). Digital technology has revolutionized not only the way information is packaged, processed, stored, and disseminated, but also how users seek and access information. Whether through a consortium or by independent subscription, academic libraries nowadays acquire and disseminate electronic portals and databases (Anunobi & Okoye, 2008).

Discovery systems are important tools that libraries use to connect users to the multitude of resources and subscription databases (Bonner & Williams, 2016). As one of the important content providers for libraries, library discovery services have been helping libraries to facilitate their services to users (Galyani & Moballeghi, 2007). Web-scale discovery services (WSDS) provide users with a similar one-stop shop version of a library catalog, allowing users to conduct single keyword searches over the virtual entirety of a library's accessible resources (Richardson, 2013). These services offer user-friendly search interfaces, relevance ranking, and large, centralized indexes, allowing rapid, simultaneous searching of multiple library resources (Hanneke & O'Brien, 2016).

Nowadays, there are numerous ways for users to seek and retrieve academic resources. Examples include search engines like Google and Bing, indexing databases like Web of Knowledge and Scopus, websites of publishers like Elsevier and Springer Nature, or library discovery services. Previous studies show that faculty and graduate students prefer Google and Google Scholar as search tools due to their speed and ease of use (Cothran, 2011; De Groote, Shultz, & Blecic, 2014). Academic researchers frequently choose these tools over traditional bibliographic databases because of their efficiency and relevance ranking (Ollé & Borrego, 2010). According to the report released by the Emerald publishing group, which describes the findings of a survey among 1000 authors, most researchers access literature through institutional resources. Google Scholar is found to dominate as the key content search tool for over 90% of researchers using Google Scholar to search for scholarly content. In contrast, 9% of researchers use institution's discovery services, and 2% use ProQuest (Flenley, 2016).

In the digital era, users' information seeking behavior may leave digital traces behind, which we can learn from referral data (Wang, Cui, Li, & Guo, 2017; Wang, Fang, & Guo, 2016). A “referral” is like a recommendation from one website to another. When a reader clicks a DOI (Digital Object Identifier) hyperlink of a scholarly article, the user is redirected to the article landing page. As an official DOI Registration Agency of the International DOI Foundation, the Crossref server

* Corresponding author.
E-mail address: xianwenwang@dlut.edu.cn (X. Wang).

https://doi.org/10.1016/j.acalib.2018.05.010
Received 1 December 2017; Received in revised form 20 April 2018; Accepted 30 May 2018

(https://doi.org/) keeps a resolution log of articles with DOIs. This means that we can measure the traffic generated by users clicking DOIs. Analyzing the DOI resolutions provides a novel method to track the digital trace left behind when users are searching scholarly content on the web. Relying on this data, we would like to answer the following questions,

(1) What are the main referring domains of DOIs? What is the role of library discovery service in directing users to scholarly content? What are the main referring library discovery services of DOIs?
(2) What is the referring temporal trend of library discovery services in the past years?

Data and methods

Web-scale discovery service

Web-scale discovery services are considered a primary trend in academic libraries (Lee & Chung, 2016). The number of discovery services has increased greatly since the advent of WSDS with WorldCat local in 2007 (Richardson, 2013). As of December 2014, there are 7881 academic libraries with discovery tools in the United States (Breeding, 2014). The website of Library Technology Guides (https://librarytechnology.org) provides comprehensive and objective information surrounding the different types of technology products and services used by libraries. We conducted a search for libraries using discovery services. The results are shown in Table 1. There are 1669 libraries using Primo Central of Ex Libris as the discovery index and 426 libraries using Summon of Serials Solutions, both of which belong to ProQuest currently. Meanwhile, EBSCO Discovery Service is adopted by 1355 libraries. Worldcat is used in 612 libraries. It seems that ProQuest has the largest market share. Therefore, in this study, we choose ProQuest as our research object.

DOI referrals

Digital object identifier (DOI) is a unique alphanumeric string assigned by a registration agency (the International DOI Foundation) to identify content and provide a persistent link to its digital location on the Internet, such as a URL. A DOI is assigned by the publisher when an article is published and made available electronically.

Launched in 2000, Crossref is an official DOI Registration Agency of the International DOI Foundation. As of 2016, the registered content of Crossref has reached over 82 million, including 6.82 million journal publications, 11.39 million books and book chapters, and 4.61 million conference proceedings, etc. Moreover, the number of DOI resolutions has reached 2000 million times (https://www.crossref.org/pdfs/annual-report-2015-16.pdf).

As a Crossref labs project, Chronograph is a platform that records and analyzes the websites users come from when they click on a DOI. Based on the log data every time a DOI link is clicked, Chronograph can track both how many times each DOI is resolved and the websites users are on when they click the DOI links.

When a user clicks a hyperlink in a web browser, the browser sends a request to the web server holding the destination webpage. This request includes the referrer field, which is usually in the form of the URL of the previous page the user was on. Websites log referrers as part of their attempt to track their users, for promotional or statistical purposes. The data used in Chronograph is the 'Referrer header' sent by the web browsers, which indicates the website the reader was on when the DOI links was clicked.

On the website of Chronograph, there are two search options on the interface (see Fig. 1 below). In the upper search box, users could enter a domain to retrieve the referrals from the input domain to DOI resolutions. In the lower search box, entering a DOI will return the DOI resolutions over time by month or by day.

In this study, we mainly focus on the referrals from ProQuest related domains to DOI resolutions. ProQuest provides solutions, applications, and products for libraries. Its resources and tools support research and learning, publishing and dissemination, and the acquisition, management, and discovery of library collections. In 2011, ProQuest acquired Serials Solutions. Serials Solutions provided e-resource access and management services to libraries, enabling librarians to manage electronic resources more easily that serve the needs of their users. In October 2015, ProQuest acquired Ex Libris. It was to merge the Workflow Solutions division of ProQuest, which included the former Serials Solutions, into Ex Libris.

Services such as Summon of Serials Solutions, Primo of Ex Libris and Alma of Ex Libris are used by libraries all over the world to deliver scholarly resources to users. For example, the Summon service of Serials Solutions provides a single search box for finding all items in a library's collection including books, electronic resources, academic databases, multimedia items and other holdings.

Therefore, in this study, three ProQuest related domains are selected as our research objects, which are proquest.com, exlibrisgroup.com, and serialssolutions.com. We collect the data from Chronograph by searching these domains in the search box. Although there may be various related subdomains for one main domain, the data of all subdomains are combined by Chronograph. As a result, we can collect all the data of one domain together with its subdomains by searching the name of the main domain. For example, the domain of serialssolutions.com includes thousands of subdomains with summon.serialssolutions.com or search.serialssolutions.com at the end, such as yale.summon.serialssolutions.com for Yale University Libraries, xx4ay4fv5x.search.serialssolutions.com for Carnegie Mellon University Libraries, etc. Similarly, the domain exlibrisgroup.com includes thousands of subdomains with hosted.exlibrisgroup.com or alma.exlibrisgroup.com at the end, such as ucl-primo.hosted.exlibrisgroup.com for University College London Libraries, harvard.hosted.exlibrisgroup.com for Harvard University libraries, etc. When searching the domains, we can get not only all the data of the domains but also the referrals data of thousands of libraries using Summon Service or Primo Service.

Some libraries using Summon or Primo services put the name and logo of Summon or Primo on the front page of their websites. Even when they do not have the names or logos, we can tell which service they are using. Specifically, in the returned URLs of the search results, the service that the library is using is indicated. For example, when we search in Princeton University Library, the returned URLs of the search results would begin with princeton.summon.serialssolutions.com, which indicates that Princeton University Library uses the Summon service.

Fig. 2 displays the original data from January 1st 2011 to January 1st 2016 collected from Chronograph.

Time series analysis

Considering that data points taken over time may have an internal structure (such as autocorrelation, trend or seasonal variation), time series analysis is employed to identify and extract the patterns and characteristics hidden in the daily data over the period of 2011–2016. We are interested in identifying increasing trend, seasonal variation, and periodicity, etc., which could not be told from the data with the

Table 1
Major library discovery services.

Company	Discovery service	Libraries
ProQuest	Primo Central	1669
ProQuest	Summon	426
EBSCO	EBSCO Discovery Service	1355
OCLC	WorldCat Discovery Service/WorldCat Local	612