Qualitative and Quantitative Methods in Libraries (QQML) 4: 947-–953, 2015

Introducing the "Getting Found" Web Analytics Cookbook for Monitoring Search Engine Optimization of Digital Repositories

Kenning Arlitsch[1] and Patrick OBrien[2]

[1]Dean of the Library, Montana State University, Bozeman, MT, USA
[2]Semantic Web Director, Montana State University, Bozeman, MT, USA

Abstract. A new toolkit that helps libraries establish baseline measurements and continuous monitoring of the search engine optimization performance of their digital repositories is one of the products of research funded by the Institute of Museum and Library Services. The "Getting Found" Cookbook includes everything necessary for implementing a Google Analytics dashboard that continuously monitors SEO performance metrics relevant to digital repositories. While the Cookbook has been created for use with Google Analytics, the principles and practices described can be applied to any page tagging analytics software.
Keywords. Search engine optimization, Web analytics, Google Analytics, digital repositories, toolkits

1. Introduction

Basic search engine optimization (SEO) is a prerequisite to advanced digital library and Semantic Web applications. Most efforts with linked data will be rendered ineffective if search engines cannot first harvest and index the contents of library websites and repositories. Most libraries pay little attention to SEO and the use and impact of their repositories is diminished. Even libraries that do employ some level of SEO tend to have difficulty monitoring, managing and reporting results because they lack the knowledge or resources to effectively implement or use Web Analytics tools.

In 2011 the authors were awarded a three-year grant from the Institute of Museum and Library Services (IMLS) to research the problem of poor visibility of digital library content in Internet search engines. Titled "Getting Found: Search Engine Optimization for Digital Repositories," the grant was proposed following the realization that ten years of digital library development at the University of Utah had resulted in a large corpus of material that was barely indexed by common search engines, and thus not easily discoverable to the

Received: 15.4.2015 / Accepted: 9.11.2015 ISSN 2241-1925
© ISAST

majority of Internet users. Visitation to the digital library was low as a result, and surveys conducted in 2010 had revealed that SEO problems are endemic in libraries. While millions of dollars are poured into digitization programs SEO tends to be an afterthought, if it is applied at all. The authors' research revealed several themes about SEO in libraries:

1. **Administration and Strategy**: SEO is rarely driven from the top of the organization; it is usually considered a technical issue and is left to IT with little consideration of strategy, goals or reporting.
2. **Communication**: administrators don't communicate the reasons for an SEO program and its impact to the rest of the organization. Communication among the stakeholders of an SEO program can also be poor.
3. **Ineffective Analytics Reporting:** Web Analytics software is often incorrectly configured, diminishing the ability to report use of a digital library or monitor the effects of change to the repository, both planned (e.g. upgrades, software patches) and unplanned (e.g., excluding content due to mistakes in server robots.txt files).
4. **Metadata:** can be inadequate or inaccurate, and the schema are sometimes inappropriate for Internet search engines.
5. **Website design:** graphics and labyrinthine paths to content can create barriers to search engine crawlers.
6. **Software:** content management software is often not tuned for SEO and includes a variety of problems, including lack of canonical links to digital objects.
7. **Hardware:** server hardware is sometimes inappropriately tuned, and changes in digital collections are often not communicated to search engines.

2. Does SEO Work for Libraries?

Indexing ratio is the number of digital objects that appear in a search engine's index compared to the overall number that were submitted, usually through a sitemap. The authors use indexing ratio as a metric for evaluating the accessibility of digital repository content: higher indexing ratios indicate more discoverable and accessible content. Over a period of two years the authors implemented SEO techniques at the University of Utah that improved the indexing ratio of the digital collections in Google from 12% to approximately 98%. The increase in indexing ratio in turn increased referrals from Google domains by 500%, which resulted in a 135% increase in the number of visitors to Utah's digital library.

The authors also addressed low indexing ratios of the USpace institutional repository (IR) at Utah in both Google and Google Scholar (GS). In 2010 the GS indexing ratio for USpace scholarly content was less than 1%. SEO efforts increased the number of full text USpace scholarly papers indexed by GS from approximately 400 to over 6,720. Harvesting and Indexing challenges unique to

GS included the Dublin Core metadata schema, which could not represent the individual elements of citations; i.e. there was not a separate field for each part of the citation (Arlitsch & O'Brien, 2012). Montana State University launched its first IR in 2013 and lessons learned at Utah helped create a current indexing ratio in GS of approximately 80%.

Much of the research conducted during the course of the grant was published in articles, presentations, and a book (Arlitsch & OBrien, 2013), but the grant also called for the delivery of the Getting Found (GF) Cookbook to help other libraries measure and monitor the SEO health of their digital collections. The GF Cookbook offers a process to inventory and coordinate hardware systems, online services such as institutional social media accounts, and web properties such as various digital repositories or collections. It helps identify stakeholders as well as the various roles people play and the skills required to implement and maintain a library SEO effort.

This paper provides an overview of the GF Cookbook, which will become available for free online distribution in the Spring of 2015 through the website and Vimeo channels of the Council of Library and Information Resources (CLIR), a partner on the GF grant:

- Videos are organized on a new SEO Channel on the CLIR Vimeo site - https://vimeo.com/clir/channels
- Additional resources and tips are provided via CLIR's website
    - Publication Reports - http://www.clir.org/pubs/reports
    - Collaborations - http://www.clir.org/dlf/collaborations/seo

3. Measuring Results

"*We cannot call a digital-library or electronic-publishing system a success if we cannot measure and interpret its use"* (Peterson Bishop, 1998).

Libraries have long measured their holdings and services, and they frequently compare themselves to other libraries in the aggregate assessments of their professional organizations; see, for example (Kyrillidou & Bland, 2009). In recent years, assessment has become a common theme in the literature as libraries have felt pressure, along with many other educational and non-profit organizations, to justify the funding they receive for their products and services.

University administrators, funding agencies, donors, and professional organizations are all potential audiences, and tools are needed to help produce data-supported evidence to objectively compare and evaluate an organization's value and improve budgeting decisions.

Web Analytics, implemented correctly, are invaluable tools for helping administrators and librarians make informed decisions about optimal resource allocations and for reporting to these audiences. Analytics provide objective

feedback concerning the value of digital repository services being delivered to patrons.

4. The Getting Found SEO Cookbook

The Getting Found Cookbook guides an implementation of Web Analytics tools for accurate overview and long-term monitoring of digital repository performance. Google Analytics (GA) and Google Webmaster Tools (GWT) are powerful and freely available software tools that are most commonly used by libraries in the United States. While the GF Cookbook has been created for use with GA, the principles and practices used in the Cookbook can be applied to any page tagging analytics software.

GA and GWT are typically used in conjunction. GWT provides feedback from search engine crawlers as they navigate the website and harvest indexable data. Messages about barriers encountered by the crawlers are displayed in GWT, allowing SEO administrators to troubleshoot and address problems that may prevent harvesting.

GA provides information about users who visit the website. Challenges exist because GA has hundreds of features and capabilities that require time, resources and a level of technical marketing expertise that staff in most libraries do not have. The GF Cookbook provides tools and guidance to help quickly implement a pre-configured GA Dashboard to evaluate access and visibility of digital repositories.

The GA Dashboard design was based on analysis of the goals of the strategic plans and mission statements of various USA funding agencies to identify potential Web Analytics key performance indicators (KPI) useful for evaluating the progress toward the goals communicated by each agency. See the following sources for examples: (Institute of Museum and Library Services, 2011), National Endowment for the Humanities (NEH), n.d.), (National Science Foundation, n.d.).

Table 1: Example relationship of funding agency goal to KPI

NEH Goal	Related KPI	KPI Description
Preserve and provide access to cultural and educational resources.	Visitor Traffic	Tracks the amount of originating visitor traffic (search, direct, referral) over time. An increasing number supports growing access to cultural/educational resources.
	Page Views	Tracks the change in number of pages viewed over a period of time. An increasing number supports growing access to cultural/educational resources.
	Visitors by Geography	Tracks visitors by geographic location. Shows evidence of increasing access to resources by location.

The authors also conducted extensive interviews with digital repository managers to identify common objectives that would benefit from web analytics reporting. These included:

- Increase the number of visitors, as well as the number of items each visitor views.
- Increase the number of returning visitors
- Reduce user departures by improving navigation
- Increase funding and content donations

Based upon our analyses of funding goals and repository manager objectives we identified the most relevant KPI's and incorporated them into a reusable GA dashboard to help answer the following questions.

1. How are web properties performing when users visit?
2. What are users looking for?
3. Who is referring users?
4. What are the most common entry pages?
5. When/why are users exiting the website?

Baseline Measurements

The GF cookbook offers a process and tools designed for a single person who is responsible for establishing and maintaining GA for their library organization. This includes a survey to inventory and coordinate hardware systems, online services such as institutional social media accounts and web properties including digital repositories or collections. It also helps identify stakeholders as well as the roles people play and the skills required to implement and maintain a library SEO effort.

Dashboard

Based on GA, the dashboard allows library staff and administrators to monitor common baseline performance metrics relevant to library repositories. It also empowers administrators to ask pertinent questions when changes are observed. For instance, a sudden drop in the number of referrals from a social media source could prompt an administrator to question the appropriate manager or staff person responsible for outreach about the collection.

Video Tutorials

The GF Cookbook provides a series of basic video tutorials to integrate the necessary analytics for consistent evaluation and monitoring of library objectives. The videos describe the steps and information necessary to implement baseline SEO metrics relevant to library administrators and collection managers. The tutorials culminate in a reusable dashboard that is hosted in the *Google Analytics Solutions Gallery* where it is titled "Getting Found SEO Dashboard." (OBrien, 2015).

Justifying the Need
The first video describes the reasons for an institutionalized SEO program and provides graphic data demonstrating the authors' achievements in improving use of digital libraries.
Strategic Planning
Integrating SEO into the library strategic plan is discussed in this video, as well as the reasons for failure or ineffectiveness of many SEO programs.
Prepping for Inventory
This video begins the planning process for inventorying existing web properties. It describes the roles of various stakeholders in the organization and their roles in an SEO program.
Creating a Master Google Account
Part of the difficulty in establishing a comprehensive and integrated SEO program is that various people in the organization tend to manage different accounts. This video demonstrates the process for establishing a single master account to maintain organizational continuity and to administer all Google Web Service accounts owned by the library organization.
Serving the GF SEO Dashboard
The final video describes how to install the preconfigured Getting Found Google Analytics dashboard and its features.

5. Conclusions

The Getting Found Cookbook is designed to help libraries implement Web Analytics tools so that they may monitor and measure the effectiveness of their SEO programs. In a time when Internet search engines have become the default discovery layer for the vast majority of users it is crucial that libraries are able to report that their websites and digital repositories are discoverable through those search engines as well. Assessment begins with accurate baseline measurements and the ability to track performance and react to changes. The GF Cookbook will help empower libraries to tell stories with the support of data.

References

Arlitsch, K., & O'Brien, P. S. (2012). Invisible institutional repositories: Addressing the low indexing ratios of IRs in Google Scholar. *Library Hi Tech*, *30*(1), 60–81. http://doi.org/10.1108/07378831211213210

Arlitsch, K., & OBrien, P. S. (2013). *Improving the visibility and use of digital repositories through SEO*. Chicago: ALA TechSource, an imprint of the American Library Association. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=578551

Institute of Museum and Library Services. (2011, December). *Creating a Nation of Learners; IMLS Five-Year Strategic Plan 2012–2016*. Retrieved from http://www.imls.gov/assets/1/AssetManager/StrategicPlan2012-16_Presentation.pdf

Kyrillidou, M., & Bland, L. (2009). *ARL Supplementary Statistics 2007-2008*. Washington D.C.: Association of Research Libraries.

National Endowment for the Humanities (NEH). (n.d.). About the National Endowment for the Humanities. Retrieved April 12, 2012, from http://www.neh.gov/about

National Science Foundation. (n.d.). US NSF - Merit Review. Retrieved April 23, 2012, from http://www.nsf.gov/bfa/dias/policy/meritreview/index.jsp

OBrien, P. (2015). *Getting Found SEO Dashboard* [Google Analytics]. Retrieved from https://www.google.com/analytics/gallery/#posts/search/%3F_.viewId%3DtsoHt4DTTqewgeR5WrwNWQ/

Peterson Bishop, A. (1998). Logins and Bailouts: Measuring Access, Use, and Success in Digital Libraries. *Journal of Electronic Publishing*, *4*(2). http://doi.org/http://dx.doi.org/10.3998/3336451.0004.207