1-1-2011

Internet Usage Data

Adam Knowlton

University of Nebraska - Lincoln, acknowlton@yahoo.com

VI

Internet Usage Data
Adam Knowlton

History of Internet Usage Statistics

Interest in quantifying the the amount of traffic directed to specific websites grew soon after the rise of the internet in the early-to-mid 1990s. Drawing from scholarly metrics such as citation analysis, Larry Page and Sergey Brin developed a ranking system for the internet that would apply numerical value to a website based on the number of hyperlinks contained within, and linked to, that same website. This measurement tool opened the door for academic scholars to learn more about how their work circulates online. However, personal websites are not the only way that scholars have been able to make public their work on the open web. Corresponding with the rise of internet, institutional repositories have begun to slowly grow in popularity. The first ever online repository arXiv was launched in 1991 and is associated with the Los Alamos National Laboratories.[1] The success of arXiv, has resulted in the launch of many other institutional and subject-based repositories around the world (see Table 1).

Finally, as the internet has continued to evolve, numerous additional sources

---

[1] Joan Cheverie, Jennifer Boettcher, & John Buschman. "Digital Scholarship in the University Tenure and Promotion Process: A Report on the Sixth Scholarly Communication Symposium and Georgetown University Library," *Journal of Scholarly Publishing* (2009). 199-230.

Table 1

Prominent Digital Repositories

Repository	Host	Location
DSpace	MIT	dspace.mit.edu
Eprints.org	University of South-ampton, UK	eprints.org
Digital Access to Scholarship at Har-vard (DASH)	Harvard	dash.harvard.edu
Focus on Access to In-stitutional Resources	Joint Information Sys-tems Committee, UK	jisc.ac.uk
Caltech Collection of Digital Archives (CODA)	Caltech	library.caltech.edu/digital
CARL Institutional Repository Project	Canadian Association of Research Libraries	carl-abrc.ca

have arisen giving scholars additional avenues for online publishing. In December of 2003 Google launched "Google Print" (predecessor of Google Books), and in October of 2004, Google launced "Google Scholar" which sought to provide "a free service for searching scholarly literature such as peer-reviewed papers, theses, books, preprints, abstracts and technical reports."[2]

Strengths and Weaknesses

Joan Cheverie, Jennifer Boettcher, and John Buschman argue that with the meteoric rise of the internet in becoming a viable publishing option for scholars, we are often left with more questions than answers about the viability of internet usage statistics within academic review processes.[3] They cite the 2006 Modern Language Association who states that we have reached a "threshold moment in digital scholarship and the promotion and tenure process," but has left the change itself up to individual departments and institutions. This section seeks to identify first the strengths of the internet usage metrics, before moving on to consider the weaknesses.

---

3   Cheverie, Boettcher & Buschman, "Digital Scholarship."

The strength of internet usage lies in the fact that despite being 90% text, the ability to incorporate design elements, imagery, and color allows scholars the unique opportunity to better explain their work.[4] Kevin Lomangino argues that it is this advantage of internet usage data that translates into higher citation rates than comparable material published in subscription-only journals. Additionally, these higher citation rates play a significant role within Google Scholar's ranking algorithm, allowing materials with both a high number of citations by other sources and a large number of citations within the article itself to be ranked highly. Outside of citation ranks, scholars may also use download rates to quantify the popularity of their work. Kevin Lomangino notes that as repositories grow in popularity they may become a serious rival for traditional publishing outlets. Lomangino points to the subject-based repository arXiv which on average has 23% more downloads than corresponding traditional publishing websites.

Despite these strengths internet usage metrics do have significant weaknesses. Cheverie, Bottcher, & Buschman argue that the usage and download statistics digital repositories offer are merely popularity of content statistics.[5] It is nearly impossible for evaluators of these statistics to determine whether or not an individual visiting the site found the information valuable and read through the entire article, or simply read the abstract or introduction and moved on.

Additionally, the complex issue of search terms points to a significant gap within usage statistic metrics. According to Beel, Gipp, and Elide, "none of the major academic search engines currently consider synonyms."[6] The impact of this claim illustrates that if one were searching for "scholarly internet usage metrics," all articles discussing "academic evaluation of web-based content" would be ignored. This could significantly alter the number of total visits, and in turn future citations, a piece of scholarly work could enjoy. Additionally, in these searches engines such as Google Scholar focus on length of titles and number of times that key-word terms are used in the title, abstract, and full-text4. This means that despite being a leader in the field, by using a variety of synonyms within their writing and not including the key-word term in the document title, an author can

---

4   Google Milestones: Corporate Information, http://www.google.com/corporate/history.html (Accessed June 13, 2010).

5   Cheverie, Boettcher & Buschman, "Digital Scholarship."

6   Jordan Beel, Bela Gipp, & Erik Eilde. "Academic Search Engine Optimization (ASEO): Optimizing Scholarly Literature For Google Scholar & Co.," *Journal of Scholarly Publishing* (2010): 177-190.

be ranked less relevant than an author who constantly repeats key-terms.

Judgment

Despite their ability to make academic work considerably more available to the public, and other scholars, than traditional publishing; internet usage statistics still fail to paint an accurate picture of relevance, impact, and popularity. While statistics such as the 23% higher download rate enjoyed by arXiv as opposed to traditional publishing outlets are significant; it is impossible to properly evaluate whether or not the material was found to be impactful and relevant to the reader. Additionally, the inability of complicated algorithms used by numerous academic search engines, Google's page-rank, and Google scholar to find what Michael Jensen, director of strategic Web communication for the National Academies calls the "nuanced perspective."[7] This nuanced perspective is currently impossible for modern search engines to accomplish since their design and intent is find facts and specific information, not to evaluate the countless factors that contribute to an author's ethos.

Field Relevance

In light of this judgment, I believe that Internet usage metrics should not be wholly avoided as a method of evaluating scholarship within the field of communication. However, it would be incredibly unwise to use Internet usage metrics as the sole determinant of an author's relevance and authority. Internet usage metrics should be used in conjunction with numerous other metrics that will allow evaluators to properly address the complexity of every author's work, and will allow them to reach the "nuanced perspective" advocated by Jensen. Therefore, I believe that the utilization of digital scholarship in the open web will bring countless advantages to readers, authors, and institutions alike; but this form of scholarship will require further evaluation and promotion before it can be considered a stand-alone form of academic evaluation.

---

7   Michael Jensen. "The New Metrics of Scholarly Authority," *The Chronicle of Higher Education,* June 15, 2007.