Webometric Study on Selective Content Curation Websites: An Explorative Study

Ponerulappan.P[1],   Dr. N. Thilagavathy[2]

[1]*Research Scholar, Bharthiyar University, Coimbatore & Lead, Library And Information Dissemination Team, Cognizant Technology Solutions India Pvt. Ltd. Chennai, Tamilnadu, India - Pincode -603103*
[2]*Librarrian, Dr.M.G.R Janaki College Of Arts & Science For Women, Deshmuk Road, R.A Puram, Chennai, Pincode -600028*

Abstract- **The purpose of the study is to analyze the some of the content curation websites through webometrics indicators including URL analysis, Back link analysis and citation to the URL. The content curation is increasingly gaining prominence and popularity as more and more information professionals are turning to curated content to reach their targeted audience. This study is to explore the availability of content curation websites and its value for information dissemination support. The link quality is determined based on the number of clicks made at a particular link in the respective websites. The reliability of a particular website is determined by the quality ration which is measured through link volume, and link quality.**

INTRODUCTION

In the digital era, the users are increasingly want to be responsible for their own learning and every industry is transforming digitally to cope up with the demand of the consumers/users. Considering the growth of information, the information professionals are required to use world wide web to support information need of the library users. Search engines are great tool to find the required information. But it has its limitations. when you're tired of doing keyword searches to get content and to share with your audience, the content curation websites support to find and information sharing instantly. The content curation websites aggregate information from all the sites and filter the same and publish on topic wise which save our time to service the users.

There is a tremendous amount of curated content sites throughout the web, with more being added every day. These sites are fantastic to keep up with the best of the web, latest news, best videos, etc. In the ever changing digital world what is increasingly obvious is that while learning still needs to be continuous, content needs to be properly *curated* and experts like Information Professionals those who know precisely what the user need and where the information is available.

REVIEW OF LITERATURE

Webometrics was coined by Almind and Ingwerson to study the website quantitatively. The websites are formed by hyperlinks and the links are counted and analyzed to measure the usefulness of websites. Björneborn and Ingwersen (2004) carried out the first attempt to develop a consistent terminology on the webometric field. Some years later Thelwall, Vaughan and Björneborn (2005) reviewed the webometric literature and elaborated various types of link analysis. In their study, they realized, the quantitative methods that originally designed for bibliometric analysis of citation patterns of scientific journal articles, could be applied to Web by using various search engines to provide the raw data. Thelwall and Wilkinson (2008) proposed a generic lexical framework that, based on the previous work, intended to unify and extend existing methods through abstract notions of link lists and URL lists. McKiernan (1996) for "backlinks" coined the term "sitation" to make it different from the term "citation" because backlinks are originated from the Web. In another study, Thelwall (2002c) analyzed the backlink structure of the top 100 linked-to pages on UK university websites, and characterized the page types of these 100 pages. Barjak et al. (2007) also have studied this to know the factors affecting web impact of personal homepages of scientists, and found that personal and institutional characteristics of scientists play significant role in yielding higher impact.

PURPOSE OF THE STUDY

This paper is intended to show how webometric techniques could be applied to content curation websites This study is carried out to identify the content curation websites available freely to use for information services and to analyses the quality of some of the content curation websites. Webometric techniques are based on the manipulation of information contained in the hyperlinks that connect the different documents contained on the Web. So far studies in this field have focused on academic and scholarly web spaces; however, this methodology is equally applicable to commercial sites which are more predominant on the Web.

SCOPE OF THE STUDY

The scope of work is limited to selected curation websites that are open for curation. Though, there are a lot more curation websites available, the website that are free to register and curate information are considered and analyzed. There are 102 websites are identified initially which were filtered based on the availability to the public freely and 10 websites are finally considered for the study.

RESEARCH ANALYSIS AND DISCUSSION

The present study is designed to study URL analysis, Back link analysis, link quality (the number of clicks from a seed set of trusted sites to a given URL, or Domain) and the link volume (the number citations to a given URL, or Domain). The URL's of content curation websites are collected through search engines. In addition to that, Majestic SEO metrics tool are also used to compute the link quality and link volumes. The data has been collected in three rounds in three months during August, September and October randomly by the use of different search engines, and the following expressions were used for data collections.

Links to single URL/Webpage: To Identify the how many webpages are linked with the URL.
Example

Query: link: < Website URL>

Links to a Domain: To identify the how many links are pointing to domain/website.
Example
Query: link domain: < Website URL>

Link Volume: The number citations to a given URL, or Domain.

Link Quality: The number of clicks from a seed set of trusted sites to a given URL, or Domain.
The main thing to remember is that the link volume is quantity of links available and the link quality is how close those links are to trustworthy, authoritative sources

CONTENT CURATION WEBSITES

AllTop: URL - http://alltop.com/ This site is great site to curate content. "The purpose of Alltop is to help you answer the question, 'What's happening?' in all the topics that interest you." They do that by rounding up content from the top blogs and publications across the web, then group those articles into "aggregations". Follow the aggregation that's most closely related to your subject to find content to share. AllTop is updated "approximately once an hour."

Feedly: URL - https://feedly.com It delivers fast, mobile-optimized content using RSS feeds.it helps to browse and share content from any favorite news sites, feeds, blogs and YouTube channels. it serves as an RSS feed to curate content around specific keywords, topics, domains, etc.

List.ly: URL - https://list.ly/ It Curate, crowdsource, and engage readers via live embedded list content inside blog posts. Listly lets user suggest items to be added to the list, helping continuously curate great content. Users can also vote on items and share individual items from the list

Medium: URL - http://medium.com/ This is a site where anyone can blog in public. Every post gets rated by readers and the best floats to the top of the "Top Stories on Medium" list. The result is some fantastic content – original thinking, fresh perspectives, and a chorus of cool new voices to discover. If you curate content, sprinkle at least a few posts from Medium into your mix. The blog posts

based on their popularity and editor choice get curated

Paper.li:URL-http://paper.li/ It Publish 'newspapers' comprised of content from anywhere on the web to treat your readers to fresh news daily.

Quora: URL - https://www.quora.com/ It helps to find answers niche questions around specific industry-related topics, It's all about finding the answers to specific questions with accuracy and efficiency. Quora also works as a keyword research tool and is a great place to find question-based keywords to answer with your content.

Scoop.it: URL - https://www.scoop.it. It helps individuals and businesses publish content in an efficient and impacting way. Uses big data semantic technology to help you quickly find relevant content. Scoop.it runs by crawling more than 35 million web pages every day and then organizing content into the most relevant topics. The tool makes it easy to filter, edit, and share out relevant content.

Storify: URL - https://storify.com It create 'stories' or 'timelines' by collecting and publishing content from social media channels such as Facebook, Twitter, and Instagram.

The Tweeted Times: URL - http://tweetedtimes.com/ This aggregates news in our Twitter stream and ranks them by popularity among friends. The Tweeted Times rebuilds newspaper hourly, and is always up-to-date. We can create a newspaper for any topic of our interest. Topical newspapers are based on streams produced by Twitter Lists or Twitter search

Triberr: URL - https://triberr.com/ is a great way to help connect with industry leaders and curate their content. The most valuable feature of this app is that you get access to groups based around a specific area of interest. In addition to engaging topic-specific content, you also get exposed to insider lingo, up-to-date info on what's hot and what's not in your niche, and interesting angles and opinions on your topic. You can find your area(s) and join the tribe to begin curating great content, and you can share that content when you find something useful.

Table 1: *Comparison of Back links, Link Volume and Link quality of Curation Websites.*

Content Curation Websites	External Back Links	Link Volume	Link Quality	Quality Ratio
http://alltop.com/	3,309,036	54	44	0.8
https://feedly.com	964,810	62	54	0.85
https://list.ly/	3,062	50	39	0.78
https://paper.li	4,406	56	76	1.35
http://medium.com/	437,266	51	46	0.90
http://quora.com	49,984	49	31	0.63
https://www.scoop.it	312,606	60	39	0.65
https://storify.com	144,675	61	52	0.85
https://triberr.com/	689	49	25	0.51
http://tweetedtimes.com	11,481	43	32	0.74

FINDINGS

In this study we find that majority of the websites are having better link capacity than the link quality except a site paper.li. The influential website might be based on how many sites link to it. It is clear that

the quantity of links available may not be an essential requirement for a curation website. The effectiveness of curation website is based on the number of users registered with the corresponding websites and using it. Unless the link is useful, there is a less obvious that user clicks the link. Link quality is measured in terms of the click made in respective websites. Hence, the quality of the site measured by the usage of the links that curated to the website. However, the availability of huge number of backlinks in a website show some impact on the link quality.

Value Ratio is calculated by = Link Quality/ Link Volume

Example:

Value Ratio of Paper.li = 76/56 = 1.35 (high value)

Value Ratio of Tribber = 25/49 = 0.51 (Low value)

The Value Ratio which contains above 1 are highly reliable websites. However, this check is only a quick overview and does not always apply to every site. There are many perfectly trustworthy sites out there that have a Quality lower than the volume.

In the above study indicates that paper.li website only have exceeded the quality ratio required for a reliable website and the remaining are quite close to the same.

CONCLUSION

This study shows that the higher quality of links makes the website more reliable on their content. However, the effectiveness and usefulness of the website content can only be measured through number of page views and visitors visit to the page. The steady increase of subscribers to the site show that that the users are interested in their content and that defines the quality of the content. Content curation is increasingly gaining prominence and popularity as more and more information professionals are turning to curated content to reach their targeted audience and that save the time of LIS professionals to search for an authenticated content for the information dissemination services.

REFERENCES

[1] Björneborn l.,Ingwersen p. Perspectives of webometrics. Scientometrics. 2001;50(1): 65–82p.

[2] Björneborn L. Small–World Link Structures across an Academic Web Space: A Library and Information Science Approach (Doctoral Dissertation). Copenhagen, Denmark: Royal School of Library & Information Science; 2004.

[3] Thelwall M, Vaughan L, Bjorneborn L. Webometrics. Annual Review of Information Science and Technology.2006; 39(1): 81–135p.

[4] Thelwall M. Bibliometrics to webometrics. Journal of Information Science. 2007; 34(4): 1–18p.

[5] Thelwall M. Webometrics: Emergent or doomed? Information Research. 2010;15(4): colis713. Available at http://InformationR.net/ir/15–4/colis713.html

[6] Maharana RK, Panda KC, Sahoo J. Web impact factor (WIF) and link analysis of Indian Institute of Technologies (IITs): A webometric study. Library Philosophy and Practice (e–journal). 2012; 789:1–11p.

[7] Maharana RK, Panda KC, Sahoo J. Web impact factor (WIF) and link analysis of Indian Institute of Technologies (IITs): A webometric study. Library Philosophy and Practice (e–journal). 2012; 789:1–11p.

[8] Pechnikov AA, Nwohiri AM. Webometric analysis of Nigerian university websites.Webology. 2012; 9(1): Article 96.

[9] Chu H, He S, and Thelwall M. Library and Information Science schools in Canada and USA: A webometric perspective. Journal of Education for Library and Information Science. 2002;43(2): 110–25p.

[10] Shukla A, Tripathi A. Webometric analysis of institutes of national importance in India. IASLIC Bulletin.2009; 54(3): 165–80p.

[11] Tania Sheko (2013) Why the fuss about content curationhttp://www.slav.vic.edu.au/synergy/volume-11-number-2-2013/technophilia/326-why-the-fuss-about-content-curation-.html

[12] ClayShirky2010)http://curationchronicles.magnify.net/video/Clay-Shirky-6#c=1RMHLF18HLLBQ6WK&t=Talk about Curation

[13] Gold, A. (2007). Cyberinfrastructure, data, and libraries, Part 2. Libraries and the data challenge: Roles and actions for libraries. *D-Lib magazine, 13*(9/10). doi:10.1045/july20september-gold-pt2

[14] Hahn, K., Lowry, C., Lynch, C. & Shulenberger, D. (2009). *The university's role in the dissemination of research and scholarship – A*

call to action. Retrieved from http://www.aau.edu/WorkArea/showcontent.asp x?id=8924

[15] Newton, M.P., Miller, C.C. & Bracke, M.S. (2011). Librarian roles in institutional repository data set collecting: Outcomes of a research library task force. *Collection Management, 36*(1), 53-67. doi:10.1080/01462679.2011.530546

[16] Swan, A. & Brown, S. (2008). The skills, role and career structure of data scientists and curators: An assessment of current practice and future needs. A report to the Joint Information Systems Committee (JISC). Retrieved from http://ie-repository.jisc.ac.uk/245/1/DataSkillsReport.doc

[17] Walters, T. & Skinner, K. (2011). New roles for new times: Digital curation for preservation. Washington, DC: Association of Research Libraries. Retrieved from http://www.arl.org/rtl/plan/nrnt/

[18] Tibbo, H., Hank, C. & Lee, C.A. (2008). Challenges, curricula, and competencies: Researcher and practitioner perspectives for informing the development of a digital curation curriculum. Paper presented at the IS&T Archiving Conference. Bern, Switzerland.

[19] Jeff Jarvis "Content farms v. curating farmers" Buzzmachine blog. December 14, 2009 http://buzzmachine.com/2009/12/14/content-farms-v-curating-farmers/