

# AN ANALYSIS OF DIGITAL DIVIDE IN INDONESIAN ONLINE MASS MEDIA BASED ON WEB METRIC ANALYSIS

Widya Silfianti\*, Matrisya Hermita, Ruddy J. Suhatri and Budi Hermana

*Gunadarma University, Indonesia*

---

**Abstract:** The improvement of content society with the increasing internet literacy encourages the development of online media portal in Indonesia. Online news portal in Indonesia is at national or local scale in general, throughout the geographical area in Indonesia, which still shows a disparity in the infrastructure of access to the internet. The purpose of this research are evaluate content richness and rankings of web traffic using search engines. Research using a quantitative approach based on data from search engines, this study also uses a qualitative approach to analyse the evolution of the homepage design and navigation structure starting from the site launched. This research explores 140 online media in Indonesia using the webometrics analysis. The difference in the kind of mass media shows that mass media with online media only owns more content or content richness, compared with a hybrid mass media (media with online and printed versions). There is a significant difference in content richness and traffic between national and regional online news portals, outside Java in particular. Another finding from the webometrics analysis is that global ranking and content richness between Java and outside Java show a highly significant difference based on the statistical test. Discriminant analysis shows the result that content richness and website ranking are the factors to predict the performance of online mass media based on its location. The main implication of this research is the need for management and editor of online news portal to design a website which is friendlier to the search engine and to improve news productivity with the quality content.

**Keywords:** Webometrics, Content Richness, Web Popularity, Online Media, Digital Divide

---

## Introduction

Development of the newest communication technology allows the audience to interact with the news, while the journalists can monitor the behaviour of audience online (Tandoc, 2014). The exponential growth of content on world-wide-web (WWW) and public awareness to the internet as an important infrastructure in 1990s encouraged millions of people to search content on website every day (Sujithai & Jeysankar, 2013). Digital media causes some extensive changes in the society for the last one-quarter century (Schroeder, 2018). Nowadays, all people want to be on the internet due to its content richness (Thanuskodi, 2012). The Internet provides several resources and content services, hypertext document in particular which relates to each other from the website (Shafi & Bhat, 2014). The number of web pages on the internet has increased rapidly for the last two decades, but only some of them are indexed by several search engines (Arunachalam, Koumpis & Handschuh, 2018). One of the sectors or professions which make use of internet as communication media is mass media, including newspaper, radio, television, or other kinds of mass media.

The established media and entertainment companies, such as newspaper and broadcasting, face the challenges in terms of income and opportunity when consumers keep increasing and changing the use of traditional media to the digital experience (Berman et al., 2011). In 1990s, traditional news media started uploading content or news on website since it was afraid of losing its business in digital age (Alves, 2001). Migration from offline to online news has occurred (Ahlers, 2006). The potency in the use of internet by conventional media was stated twenty years ago by Williams & Nicholas (1999); they stated that the availability of capacity in electronic environment is the main factor in design and content of online newspaper. Graham & Greenhill, (2013) stated a concept of synergy between online and offline media in which the companies of news media must rebuild their

identity of brand and market position in the online market. More contents with better quality offered by this online media tend to be made for the Web (Noci, 2013). Chung et al. (2010) stated that one of the aspects which

need to be investigated further about online media is the used characteristic of technology. This characteristic of technology as the main topic of this research is an analysis of website with webometrics approach and the tracking of evolution in the web design in certain period.

The importance of internet as the main source of content encourages the research by scientists in the field of content (Shafi & Bhat, 2014). One of the approaches which can be used by the researchers in the field of content is webometrics. Since the use of its terminology in 1997, webometrics keeps growing rapidly as a method which can measure Web Impact Factor (WIF), which is a key metric to measure and analyze hyperlink of a website (Thelwall, 2012). In the analysis of link to social web analysis, including the research in the social science, the webometrics approach is used as a discipline which emerges from the growth of WWW to spread and publish content or content by individual and organization (Pandey, 2014). A research conducted using webometrics is devoted to analysis or measurement of the effect from a website (Kumar, 2018). This research aims to see the comparison between content richness and website ranking in Indonesian online media, based on the kind of media, business model, and location of office where the company manages the online media. These media companies are newspaper/magazine, radio, and television which operate in Indonesia. Migration of readers, audiences, and advertisers from newspaper and television broadcast to website is not limited in the developed countries, but it has been a real trend in the developing countries (Chukwu, 2014).

### **Research objectives**

The purpose of research are evaluate the content richness and ranking of web traffic from online media sites in Indonesia. The approach used is a combination of quantitative and qualitative methods. The quantitative approach is applied by using web metrics that include the parameters of content richness and web traffic using search engines. A qualitative approach was used to analyse the evolution of the homepage design and navigation structure starting from the site launched until the research was conducted.

### **Theoretical Framework**

#### ***Online Journalism and Digital Media***

Berman et al. (2011) stated three main drivers in the service of media in digital age, namely value shift, substitution, and weaker digital revenue models. Value shift relates to the balance in power which has shifted due to manufacturer and distributor/aggregator of device which keep making innovation to give a better experience to the consumers. This substitution factor is shown with the adoption of digital device and content in all age groups and it keeps encouraging fragmentation and the decrease in traditional media consumption. However, the use of digital media does not yet always result in a profit which is higher than in traditional media; it is related to the third factor, in which the consumers have been familiarized to expect free content delivered to them, while the company currently struggles to get profit from a paid digital model. This factor becomes a consideration why a media company still implements its business model in a conventional way, or a paid online service for certain category or kind of content service. Alves (2001) stated that most of the newspapers have shifted to Web in a highly small and affordable operation with a reason to 'wait and see'.

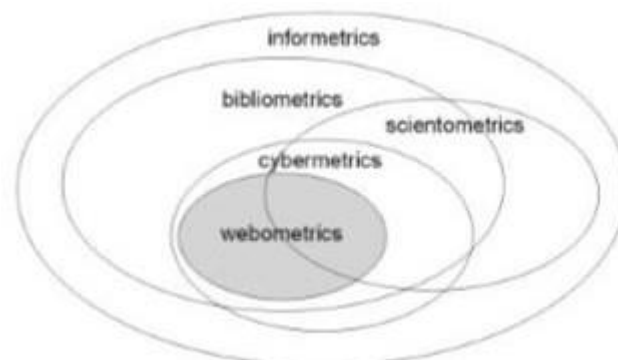
Williams & Nicholas (1999) stated that the delivery of content content on website has superiority over traditional media, namely proximity, hyperlinking, multimedia, customization, and interactivity. Hyperlink is a capability to jump from an electronic file to another as a serene feature from World Wide Web. Chung et al. (2010) stated three important factors in online newspaper, namely skill, belief, and attraction. In the perspective of economy, the kind of media does not change message and online newspaper as a platform in maintaining its profit despite the emergence of alternative media platform which is newer (Caruana, 2013). This economic consequence is in line with a statement from Ahlers (2006) in which the cost for distributing electronic edition is

almost zero and it is lower than a cost for shipping printed version in the case of newspaper and magazine, or through broadcast for television and radio. The cost consideration in shifting from traditional to online media is called the switching cost.

### ***Website Traffic Measurement***

Connectivity plays an important role in recognizing a website in virtual environment (Didegah & Erfanmanesh, 2010). Web traffic analysis and the accurate interpretation of its result are important aspects to optimize website and effective online communication, in which some companies or organizations consider them more important than offline communication (Hodakova & Nemethova, 2011). Website is a set of the related web pages, picture, video, or other digital statements with domain name or IP address which are similar in internet protocol-based network. These websites are hosted at least on a web server which can be accessed through internet or private local area network (Thanuskodi, 2012).

Webometrics is the implementation of webmetrics in World Wide Web (Almind & Ingwersen, 1997). Noruzi (2006) stated that webometrics is an expansion of theory and practice in the bibliometrics technique on Web. According to Thelwall (2012), the webometrics approach learns website, including source of content, structure, and technology. Webometrics is defined as: “The study of the quantitative aspects of the construction and use of content resources, structures and technologies on the Web drawing on bibliometric and informetric approaches” (Björneborn, 2004), which covers four main areas, namely web page content analysis, web link structure analysis, web usage analysis, and web technology analysis. Webometrics is a part of cybermetrics and in the group of a wider science as part of informetrics, as shown in Figure 1.



*Figure 1: Relationships between webometrics and other webmetric measures*

*(Björneborn & Ingwersen, 2004)*

One of the important concepts in webometrics is Web Impact Factor (WIF), defined “as the sum of the number of links to that entity divided by the number of pages found in that entity at a given point in time” (Ingwersen, 1998). WIF is determined by the number of link to the website in which the data can be obtained using some search engines or commercial web service. Link analysis is a term widely used to evaluate a website based on the calculation or analysis of hyperlink to or between a set of websites (Thelwall, 2011).

### **Method**

Considering that webometrics research method has a potency to collect a large number of data (Shari, Haddow, & Genoni, 2012), this research is limited to website on online mass media in Indonesia; there are only 140 websites and two attributes, namely number of webpages and website ranking. The category of online mass media consists of printed media, radio, and TV broadcast. Two kinds of business models are (1) hybrid with conventional/offline service and online service, and (2) online service. Analysis of digital divide was conducted by considering the result of website analysis based on location, in which the companies are headquartered in Java and outside Java.

Sources of data used in this research were two search engines. Online search engine is widely used for all kinds of webometrics study (Kumar, 2016). The research used an online search engine which can be accessed for free, namely google.com, and a paid service from Alexa.com. Search engine has a role in making World Wide Web accessible by end user (Arunachalam et al., 2018). Google was used to calculate the content richness in a website, measured from the number of its web pages, while Alexa.com was used to see international and national rankings, based on the web traffic analysis, particularly the number of visit in certain period. Measurement period for the number of web pages and website ranking was in November 2018. Description of webmetrics and data source is stated in Table 1.

Table 1. Webmetrics attribute, query syntax, and online search engine used

Webmetrics Attribute	Query Syntax	Data source
Total number of web pages extracted for the given website	Site: domain URL Example: site: kompas.com	google.com
Global rank of website	Input URL address of website	alexa.com
National Rank of website	Input URL address of website	alexa.com

The applied statistical analysis was discriminant analysis to see the prediction rate of web attribute on the grouping of website based on its location, kind of media, and business model. Discriminant analysis is based on discriminant function, namely variation from independent variable as a distinguishing power in predicting the group membership (Hair, 2014). In this research, what is meant by group is the group of location, in Java and outside Java. This research is also completed with investigation of change in web design in a chronological way for website which occupies the first rank in Indonesia. The evolution of web design from its launching to 2018 can be analyzed with a site of wayback machine which can be accessed at <https://archive.org/web>.

## Result And Discussion

### Descriptive Statistics

The analysis was conducted on 140 media websites, including 130 e-newspapers/magazines, two radio stations, and 8 television channels. Viewed from its business model, 94 companies (67.1%) run its business in a conventional way besides owning website; they are called hybrid model. There are 71 companies (50.7%) headquartered in Java. Description of content richness measured from the number of web pages is shown in Table 2.

Table 2. Web page number based on business model, location, and type of media

	N	Minimum	Maximum	Mean	S.D.	First Rank
Total	140	1	16,100,000	274,097	1,457,216	tribunnews.com
By type of media						
Newspaper	130	1	16,100,000	293,301	1,516,986	tribunnews.com
Radio	2	157,000	249,000	203,000	-	elshinta.com
Television	8	2	52,400	11,010	19,217	Sctv.co.id
By business model						
Hybrid	94	1	1,770,000	106,782	280,452	bisnis.com
Online only	46	25	16,100,000	625,049	2,517,019	tribunnews.com
By location						
Java	71	1	16,100,000	459,026	2,030,498	tribunnews.com
Outside Java	69	1	1,770,000	86,231	234,253	Bisnispapua.net

Most of online media portals in Indonesia are not yet indexed in search engine and they do not have global ranking in which only 54% of them are ranked in Alexa.com. Content richness in online media portal in general shows a high diversity and it is averagely low. Low productivity of content in Indonesian online media shows that most of the administrators in website of online mass media do not yet make use of the potency of online portal in content age or they do not yet consider search engine optimization. The difference in the kind of mass media shows that mass media with online media only owns more content or content richness, compared with hybrid mass media with its online and printed versions. Viewed from the kind of its media, newspaper has more web pages than radio and television. Media companies in Java also have more content than other companies outside Java. This content richness relates to an intensive update process by every administrator of website.

The next attribute to be analyzed is web ranking based on the result of observation using Alexa.com. Website ranking was calculated based on the result of analysis in web traffic by considering the number of visitors to the related sites. Comparison of web ranking based on media type, location, and business model is shown in Table 3.

*Table 3: Comparison of website ranking based on location and business model*

Group of website	Rank average	
	Global Rank	National Rank
<b>By Location</b>		
Java	4,251,251	1,293
Outside java	4,978,469	14,130
<b>By Business model</b>		
Hybrid	6,418,692	3,046
Online only	2,189,436	6,511

Smaller number in website ranking in Table 4 above shows better ranking and more visitors. Averagely, web ranking on online media website in Indonesia is highly low in which the average global ranking is 4,695,661, while national ranking is 4,737. The ranking is relatively low compared with all websites indexed by search engine, Alexa.com. A website ranking for online-only model is better for global ranking, compared with online model which is 2,189,436 : 6,418,692, but it is relatively good for national ranking which is 6,511 : 3,046. The difference in global and national rankings is caused by different number of site at national and international scales as well as percentage in the number of sites in which the international ranking is not yet tracked since most of them are not yet detected by Alexa.com.

Top three rankings of online news portals in Indonesia are occupied by Tribunnews.com, Detik.com, and Kompas.com. Top two websites only have online news portal, while kompas.com have its printed version. Tribunnews has the largest content since it has subdomains in which the content management is decentralized in some regions or cities in Indonesia. A strategy by making subdomain which contains local content or news in an integrated domain is effective in increasing the number of content in its main portal which is managed in a centralized way. The continuous update of website and important content delivery to the stakeholders through website can improve the ranking of a website (Jalal, Sutradhar and Mukhopadhyay, 2016), and the date of update should be explicitly stated or written on website (Thanuskodi, 2012). One of the efforts to update online news portal is the change in design and feature of service along with time, starting from its launching. The example of evolution in the homepage design is discussed below.

### ***Homepage Design Evolution***

Home Page evolution based on qualitative approach using wayback machine. The development in homepage design takes an example from Detik.com which occupies second ranking in Indonesia as well as the first online

news portal in Indonesia. The development in homepage design was taken from 6 points in time, from November 11, 1998 to December 29, 2018. It is shown in details in Figure 2.

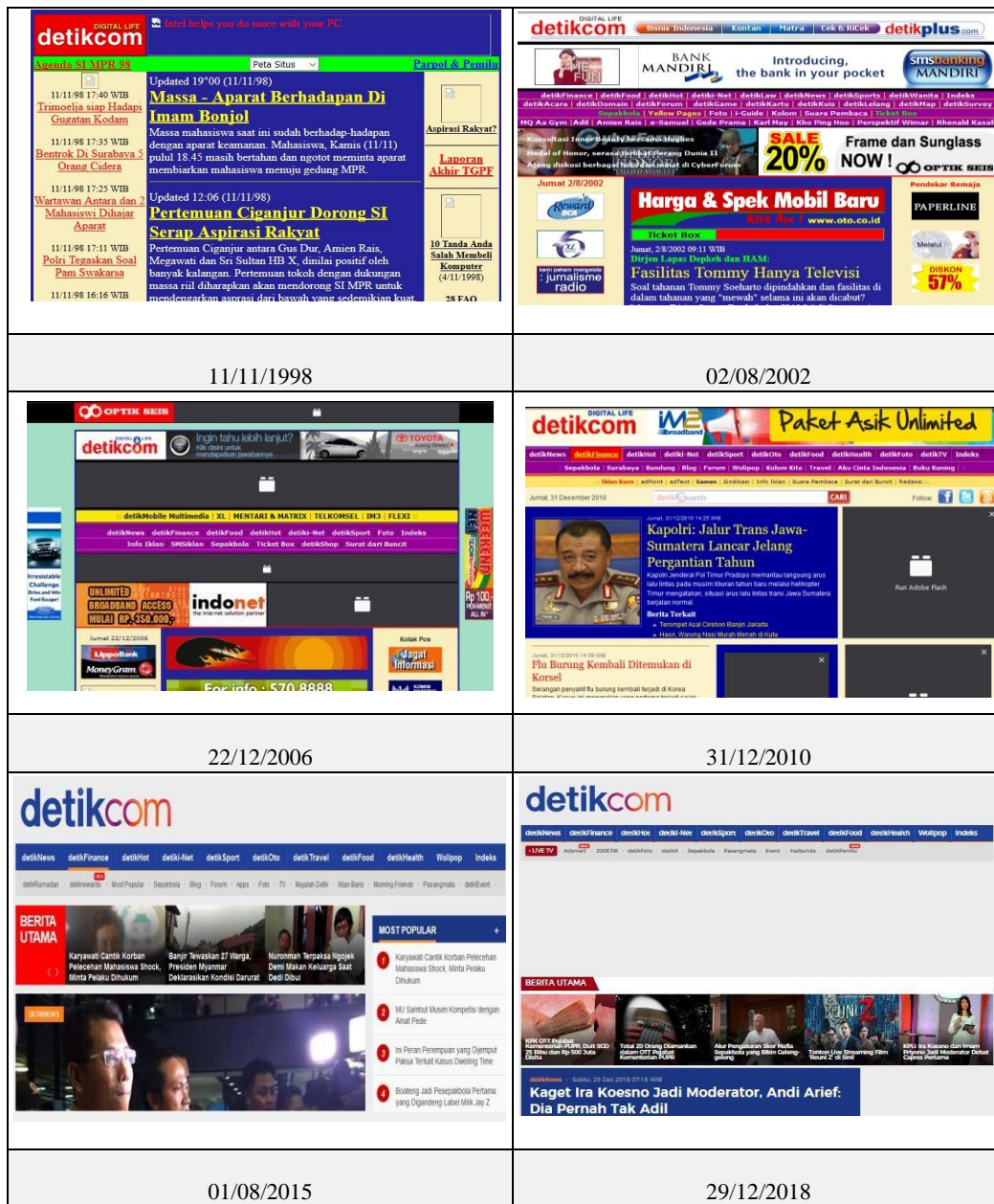


Figure 2. Case of homepage design evolution (Source: wayback machine)

The development of homepage design shows the unchanged feature and some changes in homepage design on detik.com which occupies the first ranking in Indonesia in terms of content richness and website ranking. Observation in general on 6 points in time in ten years shows that URL address, brand name, and corporate colour are relatively unchanged. Detik.com belongs to a news portal which is published for the first time in Indonesia in 1998. The changes are in menu structure, navigation, hyperlink, and layout of advertisement, including the content completeness, including multimedia and exclusive service which requires authentication from readers.

## Discriminant Analysis

Digital divide viewed from the location is statistically proven with discriminant analysis, in which the result of prediction is shown in Table 4.

Table 4. Classification Results of discriminant analysis<sup>a</sup>

		Location	Predicted Group Membership		Total
			Java	Outside Java	
Original	Count	Java	9	0	9
		Outside Java	4	2	6
	%	Java	100.0	.0	100.0
		Outside Java	66.7	33.3	100.0

a. 73,3% of original grouped cases correctly classified.

Discriminant analysis shows the result that content richness and website ranking are the factors to predict the performance of online mass media based on its location in which prediction rate of 73.3% and Global Ranking have the largest discriminating power. The difference in content richness and website ranking is caused by some factors. The first factor is capability to manage website between companies in Java and outside Java, including capability to develop a web which is search engine friendly by applying Search Engine Optimization (SEO). Website needs to be optimized to attract visitors and to sell product and service (Gregurec & Petra, 2012). SEO allows website to be shown on top search results in search engine for some keywords (Yalcin & Kose, 2010). The second factor is capacity of web server which allows the limitation of storage, so the capacity to store digital content is lower. The third factor is content update which is still low for the website administrator outside Java. This condition can be related to availability of human resource which is lower than in Java.

Content update and completeness of web feature as well as content quality are determinants in increasing the number of visitors of a news website, which will increase website ranking. Usefulness and attraction of content with various news, contents, and documents based on the websites need to be considered by the administrators of news portal. The improvement in attraction for visitors is then considered by the advertisers as the main source of income from online news portal in Indonesia. News content can be accessed for free by public. Supported by the result of webometric analysis, the editors must have capability to predict an interesting topic and to determine main news and the popular news which can result in traffic to website or visitors (Tandoc, 2014). The visit to a news portal is also determined by the result in news index published in search engine. The search result based on keywords can be list of link from search engine to the address of news which is relevant to the search result. According to Grosej (2014), the concern of content seekers is formed through a pattern of link which affects ranking (or visibility) from a website in the search result. A higher ranking in search result makes content provider more superior and it affects the perception of users about the quality of content from that website. Another alternative to improve visibility from a website is to provide content in foreign language which can attract more international visitors (Didegah & Erfanmanesh, 2010).

## Summary And Further Research

Website of mass online media in Indonesia is still dominated by online newspaper or magazine, while online website for radio and television are complimentary from its main business in conventional broadcast of radio and television. The higher the content richness or content on a website, the better the website ranking is, measured from the number of visitors. Content richness in hybrid business model is higher than content richness in the company which only provides online service. Digital divide is proven in different content richness and website ranking, viewed from location, Java and outside Java. All attributes of website in Java is better than outside Java. The digital divide potentially affects the difference of income obtained from advertisement on

website. As stated by Graham & Greenhill (2013), the growth of media in digital environment encourages value of the asset from news media companies.

The research with webometrics approach in this research is limited to quantitative analysis from web attributes, based on measurement by two search engines. The quantity of content does not yet show quality or credibility of content on every website. Positive relation between content richness and content is not viewed from the perspective of website visitors, but it is based on WIF analysis which uses algorithm from every search engine which is relatively black-box for common people. Further research can be conducted by relating web traffic analysis and perception about quality or credibility of content based on the perspective of website visitors. Several variables which can be investigated are related to perception of website quality, credibility, and behaviour to revisit (Hasley & Gregg, 2010). Some research models which can be applied are behaviour models of content system adoption, such as technology acceptance model and Unified Theory of Acceptance and Use of Technology, measurement models for website quality, such as Webqual Model or e-servqual model, and the success model of content system, including website, from the perspective of users, such as IS Success Model or Diffusion of Innovation Theory.

## Reference

- Ahlers, D., 2006, News Consumption and the New Electronic Media. *Press/Politics*, 11(1), 29-52, doi: 10.1177/1081180X05284317.
- Almind, T. C., & Ingwersen, P., 1997, Informetric analyses on the world wide web: methodological approaches to 'webometrics'. *Journal of Documentation*, 53(4), 404-426, doi: <https://doi.org/10.1108/EUM0000000007205>.
- Alves, R. C., 2001, The future of online journalism: mediamorphosis or mediacide? *Info*, 3(1), 63-72, doi: <https://doi.org/10.1108/14636690110801815>.
- Arunachalam, S. M., Koumpis, A., & Handschuh, S., 2018, Webometrics: Some Critical Issues of WWW Size Estimation Methods. *Multimodal Technologies and Interact*, 2(12), 1-11, doi:10.3390/mti2020012.
- Berman, S. J., Battino, B., & Feldman, K., 2011, New business models for emerging media and entertainment revenue opportunities. *Strategy & Leadership*, 39(3), 44-53, doi: <https://doi.org/10.1108/10878571111128810>.
- Björneborn, L. (2004). Small-world link structures across an academic Web space: Alibrary and content science approach. Doctoral dissertation, Royal School of Library and Content Science, Copenhagen, Denmark.
- Björneborn, L., & Ingwersen, P., 2004, Toward a Basic Framework for Webometrics. *Journal of The American Society for Content Science and Technology*, 55(14), 1216-1227.
- Caruana, J., 2013, The effect of online news delivery platform on elements in the communication process: An experimental investigation. *Journal of Content, Communication and Ethics in Society*, 11(4), 233-244, doi: <https://doi.org/10.1108/JICES-09-2013-0036>.
- Chukwu, C. O., 2014, Online Journalism and The Changing Nature of Traditional Media in Nigeria. *International Journal of African Society Cultures and Traditions*, 2(3), 1-9.
- Chung, C. J., Kim, H., & Kim, J. H., 2010, An anatomy of the credibility of online newspapers. *Online Content Review*, 34(5), 669-685, doi: <https://doi.org/10.1108/14684521011084564>.
- Didegah, F., & Erfanmanesh, M. A., 2010, The study of Malaysian Public Universities' performance on the World Wide Web. *Library Hi Tech News*, 27(3), 7-11, doi: <https://doi.org/10.1108/07419051011070635>.
- Graham, G., & Greenhill, A., 2013, Exploring interaction: print and online news media synergies. *Internet Research*, 23(1), 89-108, doi: <https://doi.org/10.1108/10662241311295791>.
- Gregurec, I., & Petra Grd, 2012, Search Engine Optimization (SEO): Website analysis of selected faculties in Croatia. *Proceeding of Central European Conference on Content and Intelligent Systems*, September 19-21, 2012, Croatia.
- Groselj, D., 2014, A webometric analysis of online health content: sponsorship, platform type and link structures. *Online Content Review*, 38(2), 209-231, doi: <https://doi.org/10.1108/OIR-01-2013-0011>.
- Hasley, J. P., & Gregg, D. G., 2010, An Exploratory Study of Website Content Content. *Journal of Theoretical and Applied Electronic Commerce Research*, 5(3), 27-38, doi: 10.4067/S0718-18762010000300004.



- Hodakova, M., & Nemethova, Z., 2011, The Analysis of the Influence of Selected Factors on Website Traffic. *Ekonomická revue – Central European Review of Economic Issues*, 14(2011), 37-49, doi: doi:10.7327/cerei.2011.03.03.
- Ingwersen, P., 1998, The calculation of web impact factors. *Journal of Documentation*, 54(2), 236-43.
- Kumar, J., 2016, Webometric and Technology Analysis of Journal of Inequalities and Application. *International Journal of Library & Content Science*, 5(2), 79–88.
- Kumar, K., 2018, Web Impact Factor and Link Analysis of Indian Council of Agricultural Research (ICAR) Organizations. *International Journal of Knowledge Content Development & Technology*, 8(1), 5-23, doi: tp://dx.doi.org/10.5865/IJKCT.2018.8.1.005.
- Jalal, S. K., Sutradhar, B., & Mukhopadhyay, P., 2016, Webometric Analysis of Top Ten Asian and Indian Universities. *SRELS Journal of Content Management*, 53(2), 119-127, doi: 10.17821/srels/2016/v53i2/91271.
- Noci, J. D., 2013, A History of Journalism On The Internet: A state of the art and some methodological trends. *Revista internacional de Historia de la Comunicación*, 1(1), 253-272.
- Noruzi, A., 2006, Web presence and impact factors for Middle-Eastern countries. *Online*, 30(2), 22-28.
- Pandey, R. K., 2014, Empirical Validation of Webometrics based Ranking of World Universities. *International Journal of Computer Science and Content Technologies*, 5(1) , 580-584.
- Schroeder, R., 2018, *Social Theory after The Internet: Media, technology, and globalization*. UCL Press, University College London.
- Shari, S. Haddow, G., & Genoni, P., 2012, Bibliometric and webometric methods for assessing research collaboration. *Library Review*, 61(8/9), 592-607, doi: https:// doi.org/10.1108/00242531211292097.
- Sujithai, M., & Jeysankar, R., 2013, Web Page Analysis of Indian Institute of Technologies' (IITs) Websites: A Webometric Study. *International Journal of Digital Library Services*, 3(1), 55-65.
- Tandoc, E. C., 2014, Journalism is twerking? How web analytics is changing the process of gatekeeping. *New Media & Society*, 16(4), 559-575, doi: 10.1177/1461444814530541.
- Thanuskodi, S., 2012, A Webometric Analysis of selected Institutes of National Importance Websites in India. *International Journal of Library Science*, 1(1), 13-18, doi: 10.5923/j.library.20120101.03.
- Thelwall, M., 2011, A comparison of link and URL citation counting. *Aslib Proceedings*, 63(4), 419-425, doi: https://doi.org/10.1108/00012531111148985.
- Williams, P., & Nicholas, D., 1999, The migration of news to the web. *Aslib Proceedings*, 51(4), 122-134, doi: https://doi.org/10.1108/EUM0000000006971.
- Yalcin, N., & Kose, U., 2010, What is search engine optimization: SEO? *Procedia Social and Behavioral Sciences*, 9(2010), 487–493. doi:10.1016/j.sbspro.2010.12.185.